BREAKING CARBON-CARBON BONDS WITH CYTOCHROME P450s FROM PLANTS, ANIMALS, AND BACTERIA

BY

JUSTIN C. MILLER

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Chemistry
with a minor in College Teaching
in the Graduate College of the
University of Illinois Urbana-Champaign, 2022

Urbana, Illinois

Doctoral Committee:

Professor Paul J. Hergenrother, Chair
Professor Mary A. Schuler, Director of Research
Professor Stephen G. Sligar
Professor Yi Lu

ABSTRACT

The breaking of carbon-carbon (C-C) bonds facilitated by cytochrome P450s (CYPs, P450s) is remarkable for several reasons. Chemically, the general inertness of the C-C bond makes these reactions intriguing for the selectivity with which particular C-C bonds are oxidized and cleaved. Biochemically, the inclusion of such transformations in biosynthetic pathways enables the rearrangement of carbon scaffolds into different shapes without the requirement for developing a new set of C-C bond forming enzymes. Biologically, such P450s are included in pathways that process steroids and other hormones, synthesize essential cofactors, deactivate xenobiotics, and produce compounds to ward off microbial invaders or animal predators. This dissertation summarizes investigations into CYPs mediating C-C bond breaking reactions from the plant, animal, and bacterial world.

First, I identified two secologanic acid synthases (SLASs; CYP72A564 and CYP72A565) from the medicinal plant *Camptotheca acuminata* responsible for producing secoiridoids from iridoids. After coupling these secoiridoids with tryptamine, the resulting strictosidinic acid core is diversified into species-specific terpene indole alkaloids (TIAs), a group of specialized metabolites many of which are used as medicines. Cloning the full-length coding sequences from cDNA, heterologous protein expression in *E. coli*, and *in vitro* assays of the full-length P450s confirmed that these *Camptotheca* SLASs are capable of breaking the C7-C8 bond of the iridoid glucosides loganic acid *and* loganin to make the secoiridoids secologanic acid and secologanin, respectively.

Next, I investigated the structural bases of the broadened substrate scope of these *Camptotheca* SLASs compared to the secologanin synthases (SLSs; CYP72As) from other species that can only utilize loganin. A combination of ancestral sequence reconstruction,

homology modeling, site-directed mutagenesis, and *in vitro* assays of the purified proteins revealed that the amino acid identity at two adjacent sites toggles the selectivity of the SLASs. The His131, His132 wild type proteins can turnover both loganic acid and loganin; a His132Asp mutation decreases loganic acid turnover and eliminates loganin turnover; a His131Phe mutant increases loganin turnover and conversely eliminates loganic acid turnover. The combined identification of two SLASs and mutants that toggle their substrate selectivity furnishes an improved set of secoiridoid-producing enzymes to use in heterologous expression of TIAs, replacing the yield-limiting SLS from *Catharanthus roseus*.

Subsequently, I investigated a third *Camptotheca* SLAS candidate (CYP72A730) that, although closely related to the SLASs, did not consume either loganic acid or loganin. Observing a three amino-acid deletion in CYP72A730 relative to the *Camptotheca* SLASs, I hypothesized that this deletion was the reason for this P450's inactivity. Site-directed mutagenesis to add these three amino acids to CYP72A730 and remove them from CYP72A564 and subsequent *in vitro* assays revealed that the inactivity of CYP72A730 and activity of CYP72A564 were unaffected. Consequently, it appears the inability of CYP72A730 to produce secoiridoids from loganic acid and loganin originates in some combination of the other 30 differences between it and the *Camptotheca* SLASs, most likely among the 11 more drastic changes in side chain size, charge, and flexibility.

Discerning the mechanism of the C-C bond breaking catalyzed by these SLASs was another avenue of inquiry. My approach was the synthesis of an iridoid glucoside as a probe substrate interrogating the substrate-based requirements for C-C bond scission. Working from the readily available iridoid geniposide, troubles in the early portions of several synthetic routes slowed progress. After troubleshooting these early reactions and redesigning parts of the

synthesis, the 7-deoxy-7-methyl-loganin compound is nearly complete. Once synthesized, a combination of *in vitro* reconstitutions with CYP72A564 and high resolution LC-MS for analysis will give insights into how the substrate's composition influences C-C bond scission.

The final series of studies described herein involve modeling the C-C lyase reaction catalyzed by the human, steroid-metabolizing CYP17A1 by reverse engineering this reactivity in bacterial P450s. Seeking a clearer understanding of whether this C-C bond breaking reaction occurs using a peroxyanion-mediated or a Compound I-mediated mechanism, the soluble, well-studied bacterial P450s present access to experiments not possible with CYP17A1. Using commercially available as well as specifically designed and synthesized α-hydroxyketones, neither *Pseudomonas putida* CYP101A1 nor *Bacillus megaterium* CYP102A1 could be coaxed into performing the C-C bond scission of interest. Molecular dynamics simulations with CYP101A1 suggested that the P450 positioned the α-hydroxyketone moiety too infrequently for catalysis by the peroxyanion-mediated mechanism. Subsequent work with another bacterial P450 (*Rhodopseudomonas palustris* strain HaA2 CYP199A4) and its benzoic acid substrates presented a system tested this hypothesis (*i.e.*, that the C-C lyase reaction demands holding the α-hydroxyketone in a specific orientation over the heme for an extended period of time). For this, I synthesized an α-hydroxyketone onto a benzoic acid scaffold. This synthesis allowed the Bell and De Voss groups in Australia to provide early evidence of a P450-dependent C-C bond scission with this model substrate. With a functional system now in hand, future experiments will probe the reaction in CYP199A4 to ensure its comparability to that catalyzed by CYP17A1 and definitively determine the mechanism of its C-C lyase chemistry.

ACKNOWLEDGMENTS

To my parents and family, thank you for your consistent support of my inquisitiveness, your loving care for me, and your ceaseless willingness to smile and nod as I talk about my work. Although my education and now profession have drawn me away from you, your company is always a welcome respite amid a wearying world.

To my many outstanding teachers, thank you for likewise supporting by inquisitiveness and giving me the skills that I have needed to succeed. Special thanks are due to Mrs. Bobby Thompson, Mrs. Connie Bowman, Mrs. Diann Everitt, the late Ms. Donna Hollabaugh, and Mrs. Mindy Kitchen for taking extra efforts to feed my desire to learn.

To my undergraduate professors, thank you for training me in critical thinking—especially reading, writing, and experimental design. Your enthusiasm for seeing your students succeed within and without the classroom encouraged me to learn though frequently wearied and models how I desire to motivate my own future pupils.

To Prof. Laura Wysocki, thank you for welcoming me into the world of scientific research. Your eagerness to teach, willingness to listen, and general desire to see your student succeed helped me beyond measure. Your mentorship is in particular a model for how I hope to serve my future students.

To Prof. Mary Schuler, thank you for patiently training me these past five years. From the start, you have supplied me great freedom to pursue the questions and solutions of my choosing, developing me into the scientist I now am. Your trust in me to mentor undergraduates from the start stretched my abilities, exposed my shortcomings, enabled my growth, and confirmed my calling to be an educator.

To the members of All Souls Presbyterian Church, thank you for your consistent care for me, both in body and soul, these past five years. Leaving your fellowship on earth sorrows me, but we shall again be together on the Last Day in eternal fellowship!

"Bless the LORD, O my soul!
O LORD, my God, you are very great!
You are clothed with splendor and majesty…

O LORD, how manifold are your works!
In wisdom you have made them all;
the earth is full of your creatures…

I will sing to the LORD as long as I live;
I will sing praise to my God while I have being.
May my meditation be pleasing to him,
for I rejoice in the LORD.
Let sinners be consumed from the earth,
and let the wicked be no more!
Bless the Lord, O my soul!
Praise the LORD!"

Psalm 104: 1,24,33-35

TABLE OF CONTENTS

CHAPTER 1: CYTOCHROME P450-MEDIATED C-C BREAKING REACTIONS IN BIOSYTNTHETHEIC PATHWAYS THROUGHOUT THE BIOSPHERE

1.1. CYTOCHROME P450s, AN INTRODUCTION TO THESE VERSATILE ENZYMES

Cytochrome P450s (CYPs, P450s) are a major enzyme family found in all domains of life. These highly versatile proteins feature a thiolate-bound heme that provides the spectral feature for which they are named (a 450 nm Soret peak in CO-bound reduced spectra) and the enzymes' ability to use diatomic oxygen for oxidation reactions. The time since the first documentation of a P450 in 1955 (1) has revealed that these are not the only enzymes capable of using oxygen to functionalize molecules, but CYPs continue to stand out for their ability to oxidize the most inactivated of bonds.

P450 diversity extends beyond the voluminous number of species that encode such proteins (over 18,000 as of March 2022; Table 1.1). At this time, genomic sequencing data have identified no fewer than 444,000 P450s across all three domains of life (eubacteria, archaea, and eukaryota) and over 100 in viruses (2). Although there are sequencing data for nearly three-times more species of bacteria than eukaryotes that encode CYPs, eukaryotic P450s account for over 70% of all of P450s known to date (Fig. 1.1) with an average of 66 CYPs per species.

It is also now evident that "CYPomes" (the number of P450s encoded in a given species' genome) are generally large within the separate eukaryotic lineages. Metazoans lie on the lower end with an average of 57 CYPs per species whereas the average CYPomes of plants (73 CYPs) and fungi (80 CYPs) are larger (Table 1.2). For perspective, the human genome encodes 57 P450s (3, 4). One-third of sequenced plant orders have over 100 P450s per species (Table 1.4) with the water lily species *Nymphaea colorata* alone encoding 511 P450s in its genome (2, 5).

The large numbers of P450s align well with the many roles these enzymes perform and the importance of these roles in living organisms. No fewer than 20 kinds of chemical reactions (6) are catalyzed by P450s on an array of scaffolds and with a variety of biological functions: steroid processing (7), oxidation of other lipids (8), pheromone biosynthesis (9), and maturation of specialized metabolites for defense against pathogens and predators (9). Perhaps the best-known role of P450s is the oxidation of xenobiotics—*e.g.*, by hepatic members of the CYP1, CYP2, and CYP3 families in humans (4, 10).

1.2. OXYGEN ACTIVATION BY CYPs

The near ubiquity of P450s in organisms across the biosphere and the immense number of CYPs as a whole rely on a shared chemistry that undergirds the utility of these enzymes. Using diatomic oxygen, protons from solution, and electrons from a nicotinamide cofactor supplied *via* electron transfer partners, CYPs progress through a catalytic cycle (Fig. 1.2) featuring a ferryl−oxo π-cation porphyrin radical (Compound I) usually responsible for substrate hydroxylation (11, 12). Groves and McClusky (13) proposed Compound I and the related oxygen-rebound mechanism in the mid-1970s, but this species was not directly observed in a P450 until 2010 (14) using a CYP from a thermophilic archaeon.

Entering the catalytic cycle from water-bound, six-coordinate, low-spin, oxidized iron (Fig. 1.2), binding of a substrate frequently displaces the ligating water to form a five-coordinate high-spin iron (11, 12, 15). This change in the iron's spin-state has two important consequences: 1. An easily observable "Type I" blue shift in the Soret from ~417 nm to ~390 nm (16); 2. An increase in redox potential enabling quicker reduction by an electron transfer partner (11, 12). Continuing along the cycle, a reduced iron(II) is required for oxygen binding, the result of which is a ferrous dioxygen complex or ferric superoxide complex (11, 12, 17). Supplying an additional

electron yields the peroxoanion complex that (by sequential delivery of protons) is prepared for O-O bond cleavage (11, 12, 17). Heterolytic O-O bond cleavage yields the previously introduced Compound I (11, 12, 17). Using an alkane as an example substrate, Compound I abstracts a hydrogen to form a carbon-centered radical and an iron(IV) species. Oxygen rebound produces the hydroxylated product and ferric iron (17).

This catalytic cycle features Compound I as the species reacting with the substrate and explains C-H hydroxylation, C=C epoxidation, N- and S- oxidation, and aryl C-C coupling well. Other intermeidates have also been implicated in catalysis. Nitration of tryptophan by a bacterial CYP involves reacting the ferric superoxide intermediate with nitrous oxide to generate the nitrating species—a ferric peroxynitrite (18, 19). The peroxyanion and hydroperoxy intermediates have also been implicated in catalysis (11, 12)—especially in the C-C breaking reactions discussed below.

1.3. C-C BOND BREAKING AS AN UNUSUAL BUT IMPORTANT REACTIVITY

The scaffolds of compounds produced in biological systems are somewhat restricted by how the carbon scaffold is constructed. Whether from the successive C2 units of acetyl-CoA, the C5 units of isopropropyl pyrophosphate and dimethylallyl pyrophosphate, or variegated side chains of amino acids, biosynthetic pathways generate from these building blocks the scaffolds of all natural products. Breaking C-C bonds reshapes and rearranges the structures of biomolecules and greatly transforms their bioactivities. To conduct these difficult C-C breaking reactions, P450s utilize oxygen to activate these inert bonds by oxidation. The consequence is access to compounds used for hormone signaling, enzyme cofactors, defense against predation, and beyond.

1.3.1. Steroid and hormone production

The processing of steroids in humans from cholesterol involves four P450-mediated, C-C breaking reactions (7, 20) (Fig. 1.3): lanosterol 14α-demethylation by CYP51A1 (21), cholesterol sidechain cleavage by CYP11A1 (22), 17,20-lyase of 17α-hydroxy steroids by CYP17A1 (23), and the aromatization of androgens by CYP19A1 (24, 25). Each of these CYPs catalyzes hydroxylation by Compound I as part of its repertoire with different proposals for C-C bond cleavage. After forming a diol on adjacent carbons, the C-C breaking step catalyzed by CYP11A1, separate routes remain for how Compound I mediates C-C cleavage (20). Mechanisms explaining each of these C-C bond breaking reactions have implicated Compound I or the peroxyanion intermediate for CYP17A1 and CYP19A1 (11, 12, 20).

Steroid modification by P450-catalyzed C-C bond breaking is not unique to humans or animals (Fig. 1.3). CYP125A1 from *Mycobacterium tuberculosis* catalyzes the terminal demethylation of cholesterol yielding formate using the peroxyanion as the C-C breaking species (26). Comparable Baeyer-Villiger chemistry using this peroxyanion species is proposed for the production of brassinolide (a conserved brassinosteroid in plants) by *Arabidopsis thaliana* CYP85A2 (27, 28) and *Solanum lycopersicum* CYP85A3 (28).

Steroids are not the lone hormones to feature C-C breaking reactions in their biosynthesis. *En route* to the phytohormone gibberellin, CYP88A catalyzes the contraction of a cyclohexyl to a cyclopentyl ring, the formation an aldehyde, and the oxidation of this aldehyde product to make *ent*-kaurenoic acid (29, 30). Whether the series of rearrangements catalyzed by CYP88A occurs with homolytic (31) or heterolytic (20) bond forming and breaking remains unsettled.

Because vitamin D is derived from steroids, the use an additional P450-mediated C-C breaking reaction in its metabolism seems unsurprising (Fig. 1.4). Work to date supports a series of Compound I-mediated oxidations forming a C24-C25 diol (32, 33). Additional oxidations cleave the C24-C25 bond to form calcitroic acid and an as-yet-undetected carboxylic acid (20).

1.3.3. Long chain fatty acids and other lipids

Biosynthesis of coenzymes also relies upon P450-mediated C-C bond breaking chemistry (Fig. 1.4). CYP101H1 ($P450_{BioI}$) of *Bacillus subtilis* first dihydroxylates myristic acid to form a C6-C7 *cis*-glycol (34–36). This diol is then cleaved to form heptanal and 7-oxoheptanoic acid, components transformed into the essential vitamin biotin. Although the mechanism has not been extensively studied, the similarity of this C-C bond breaking reaction to that catalyzed by CYP11A1 suggests they likely share a mechanism.

Myristic acid is also a substrate for another bacterial P450 able to break C-C bonds (Fig. 1.5). CYP152L1 ($P450_{OleT}$) decarboxylates fatty acids to form terminal alkenes with a competing reaction forming β-hydroxyacids (37–39). The decarboxylation begins with hydrogen atom abstraction from the β-carbon (37), but CYP152L1 disfavors oxygen rebound to enable the decarboxylation to occur instead (38).

CYP4G subfamily members and CYP6A1 in *Musca domestica* (the common housefly) catalyze C-C bond breaking reactions to produce olefins (Fig. 1.5) used as water repellants and pheromones (40). Unlike CYP152L1 which uses carboxylic acids, these insect P450s utilize aldehydes to produce olefins by deformylation (41). Nucleophilic attack of the substrate's aldehyde by the peroxyanion intermediate decomposes to yield olefin and formate products with the latter originating as the aldehyde (41).

Production of a terminal olefin also factors into the decomposition of nerolidol in plants (Fig. 1.5). CYP82G1 from *Arabidopsis* catalyzes this C-C bond breaking reaction by hydrogen atom transfer from C5 of nerolidol (42). Whether by forming a carbocation from the resulting radical or elimination of the alcohol from oxygen rebound, the nearby C3 alcohol collapses to yield an α,β-unsaturated ketone. The homoterpene also produced as a result of this reaction is volatile with roles as a pollinator attractant and/or herbivore defense (42).

1.3.4. Elsewhere in specialized metabolism

Some of the more bizarre C-C bond breaking reactions catalyzed by P450s are the methyl migrations of PenM and PntM from pentalenolactone biosynthesis in two species of *Streptomyces* (Fig. 1.6). Here, a methyl group attached to a quaternary carbon undergoes a stereospecific 1,2-shift while an olefin forms on the former quaternary carbon (43). Although small-scale *in vitro* reactions with these purified CYPs did not yield detectable cyclopropyl and other olefin byproducts observed in culture extracts, these compounds suggest a Compound I-mediated hydrogen abstraction and subsequent electron transfer to form a carbocation that undergoes a 1,2-shift. The rearrangement of a [3.1.1]-bicycle in fumagillin biosynthesis likewise involves C-C breaking to reshape a scaffold (Fig. 1.6). Most likely, Compound I abstracts a hydrogen from C9; the consequent C9 radical rearranges by breaking the C5-C7 bond before oxygen rebound to C5 yields a *gem*-diol (44).

Other plant P450s in specialized metabolism feature in C-C bond breaking reactions (Fig. 1.6). Comparable to the mechanism reported for nerolidol decomposition, CYP71AJ1 (psoralen synthase) abstracts a hydrogen in close proximity to an alcohol (45). Collapse of the alcohol to a carbonyl breaks a C-C bond and (in this case) aromatizes to form a furanocoumarin. This class of

compounds are common herbivore deterrents in the *Apiaceae* plant family that includes carrot, celery, parsnip, and hemlock.

Tropaine alkaloids from species such as *Atropa belladonna* (deadly nightshade) and *Hyoscyamus niger* (black henbane) feature a C-C bond breaking reaction in their biosynthesis and have important uses in medicine related to sedation. Rearrangement and C-C bond breaking through carbocations were proposed for CYP80F1 (littorine mutase)-mediated reactions in tropaine alkaloid biosynthesis (46) (Fig. 1.6). Further work with mechanistic probes found stronger support for radical rearrangements yielding the aldehyde product (47).

Other classes of medicinally relevant alkaloids also feature C-C bond breaking in their biosynthesis. Terpene indole alkaloids (TIAs) are a major class of compounds produced by various plants within the Asterid clade. These compounds share a biosynthetic pathway from strictosidine—a product of secologanin (a terpene derivative) and tryptamine (an indole). Secologanin is the product of a C-C bond breaking reaction on the iridoid glucoside loganin (Fig. 1.6) mediated by the CYP72A also called secologanin synthase (48, 49). Whether a carbocation or radical rearrangement is at play in the production of secologanin from loganin (48), however, remains unknown.

1.4. INVESTIGATING TWO GENERA OF C-C BOND BREAKING REACTIONS

As this brief survey of P450-mediated, C-C bond breaking reactions demonstrates, this small subset of CYPs catalyzes reactions of interest for their biological implications, biotechnological applications, and mechanistic chemistries. This dissertation investigates two genera of these reactions: The production of secoiridoids from iridoids by CYP72As (Chapters 2-5), and the C-C lyase reaction of α-hydroxyketones by human CYP17A1 (Chapters 6-7).

Beginning with identifying the CYPs (CYP72A564, CYP72A565) responsible for secologanic acid production in the medicinal plant *Camptotheca acuminata* (Chapter 2), I subsequently determined a pair of residues whose mutagenesis toggled substrate specificity between the unmethylated iridoid loganic acid and the methylated iridoid loganin (Chapter 3). I also evaluated whether a three amino acid deletion in the closely related CYP72A730 explained its inactivity (Chapter 4). Lastly, I attempted to synthesize an iridoid-like compound to assay the substrate-dependent determinants of C-C scission (Chapter 5) and probe the mechanism of this reaction.

Tackling the C17-C20 lyase reaction in steroid processing facilitated by human CYP17A1, I attempted to coax well-studied bacterial P450s CYP101A1 ($P450_{cam}$) and CYP102A1 ($P450_{BM3}$) to perform this reaction on α-hydroxyketones that I synthesized (Chapter 6). The failure of these reactions alongside molecular dynamics simulations supported transitioning to another bacterial P450 (CYP199A4) predicted to better position the α-hydroxyketone for catalysis (Chapter 7). Early results with another α-hydroxyketone that I synthesized from the De Voss and Bell laboratories in Australia have indicated P450-dependent C-C lyase reactivity, opening the door for further investigations into this reaction.

1.5. TABLES AND FIGURES

Table 1.1. Cytochrome P450 sequences in each domain of life.

Domain	*Number CYPs*	*Number Species*	*Avg. per Species*
Eukaryota	319,877	4,828	66
Bacteria	123,494	12,992	9.5
Archaea	1,105	413	2.7
Viruses	102	56	1.8
	444,578	18,289	

All values were taken from the InterPro database (2) for superfamily "Cytochrome P450" (IPR001128), accessed 14 March 2022.

Table 1.2. Cytochrome P450 sequences in kingdoms within the eukaryota.

Kingdom	*Number CYPs*	*Number Species*	*Avg. per Species*
Metazoa	102,600	1,812	57
Viridiplantae	92,971	1,267	73
Fungi	119,750	1,498	80
	315,321	4,577	

All values were taken from the InterPro database (2) for superfamily "Cytochrome P450" (IPR001128), accessed 14 March 2022.

Table 1.3. Cytochrome P450 sequences in phyla within the metazoa.

Phylum	*Number CYPs*	*Number Species*	*Avg. per Species*
Chordata	59,460	1,115	53
Arthropoda	30,995	497	62
Mollusca	1,652	39	42
Nematoda	3,162	78	41
Rotifera	4,964	14	355
	100,233	1,743	

All values were taken from the InterPro database (2) for superfamily "Cytochrome P450" (IPR001128), accessed 14 March 2022.

Table 1.4. Cytochrome P450 sequences in families within the viridiplantae.

Family	*Number CYPs*	*Number Species*	*Avg. per Species*
Amborellales	234	1	234
Apiales	498	32	16
Arecales	516	18	29
Asparagales	965	45	21
Asterales	3,306	78	42
Brassicales	7,339	56	131
Caryophyllales	1,252	81	15
Celastrales	193	2	97
Cornales	664	3	221
Cucurbitales	1,450	11	132
Ericales	1,987	33	60
Fabales	7,875	82	96
Fagales	1,994	13	153
Funariales	196	1	196
Gentianales	1,406	29	48
Icacinales	96	1	96
Lamiales	2,691	111	24
Laurales	394	4	99
Malpighiales	4,174	68	61
Malvales	10,889	33	330
Marchantiales	296	7	42
Myrtales	1,778	21	85
Nymphaeales	511	1	511
Oxalidales	208	1	208
Pinidae	692	66	10
Poales	20,424	85	240
Proteales	332	2	166
Ranunculales	2,668	62	43
Rosales	5,462	52	105
Sapindales	1,623	21	77
Saxifragales	376	9	42
Solanales	4,938	107	46
Vitales	1,437	8	180
Zingiberales	1,081	11	98
	89,945	1,155	

All values were taken from the InterPro database (2) for superfamily “Cytochrome P450” (IPR001128), accessed 14 March 2022.

Figure 1.1. Cytochrome P450 sequences from across the biosphere.
Sunburst plot of over 444,500 cytochrome P450-enoding gene sequences from eukaryotes (green), archaea (yellow), bacteria (violet), and viruses (pink). The size of each wedge corresponds to the number of cytochrome P450 sequences encoded.

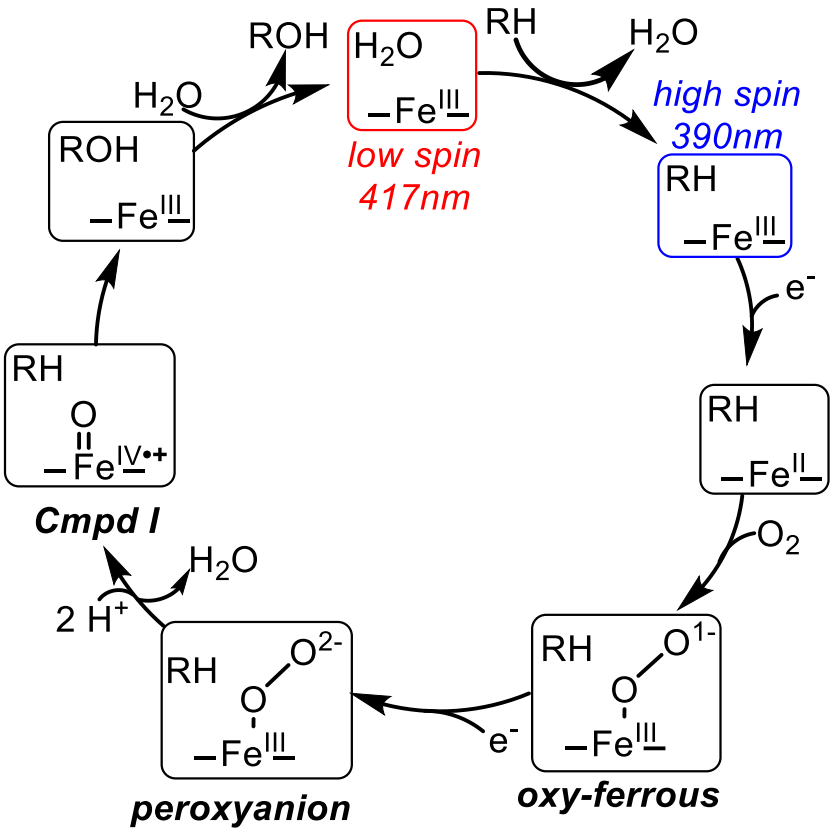


Figure 1.2. Cytochrome P450 catalytic cycle.

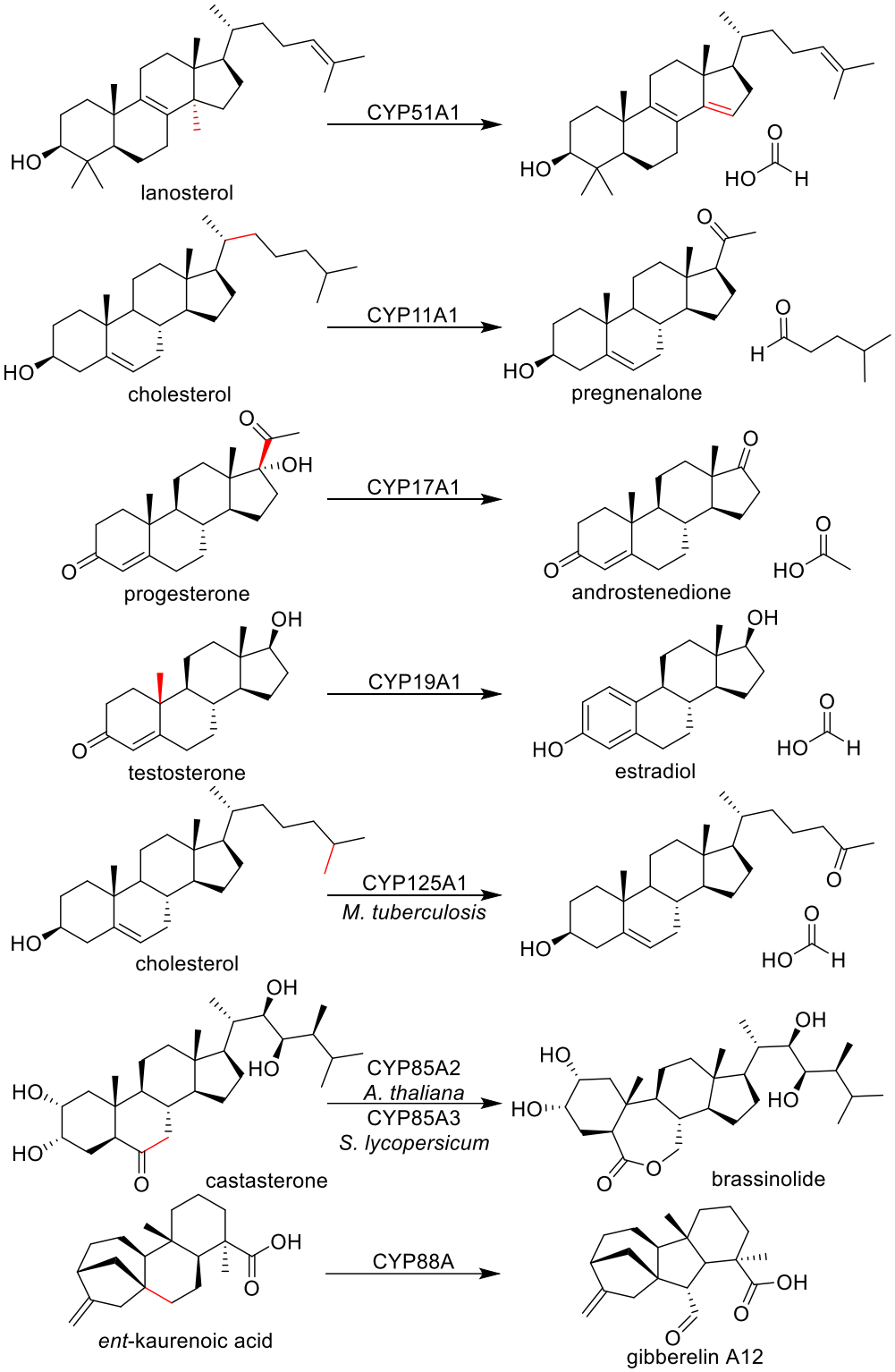


Figure 1.3. A selection of cytochrome P450-mediated C-C bond breaking reactions in steroid and hormone processing.

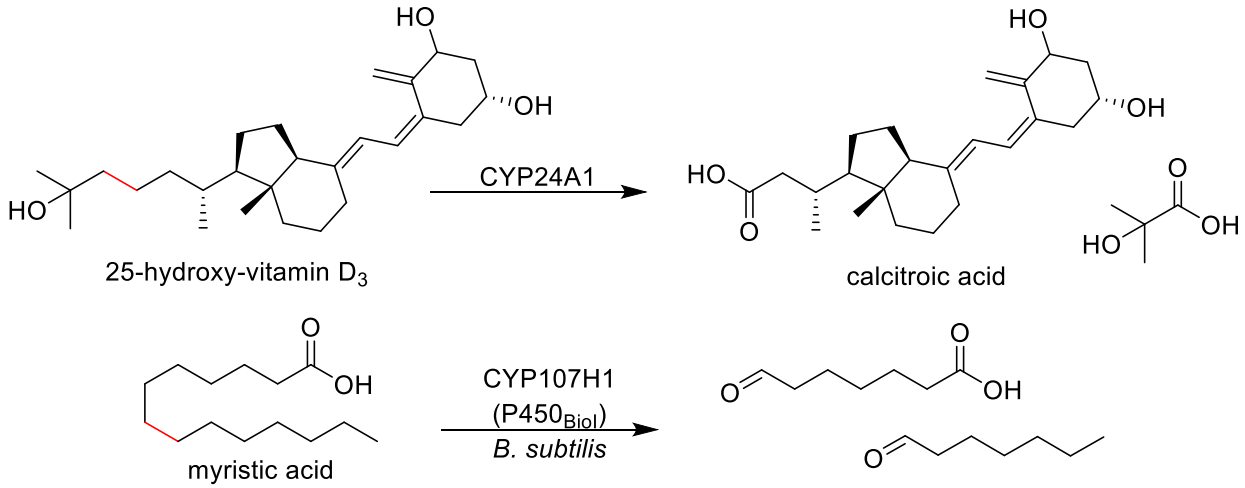


Figure 1.4. A selection of cytochrome P450-mediated C-C bond breaking reactions related to vitamin metabolism.

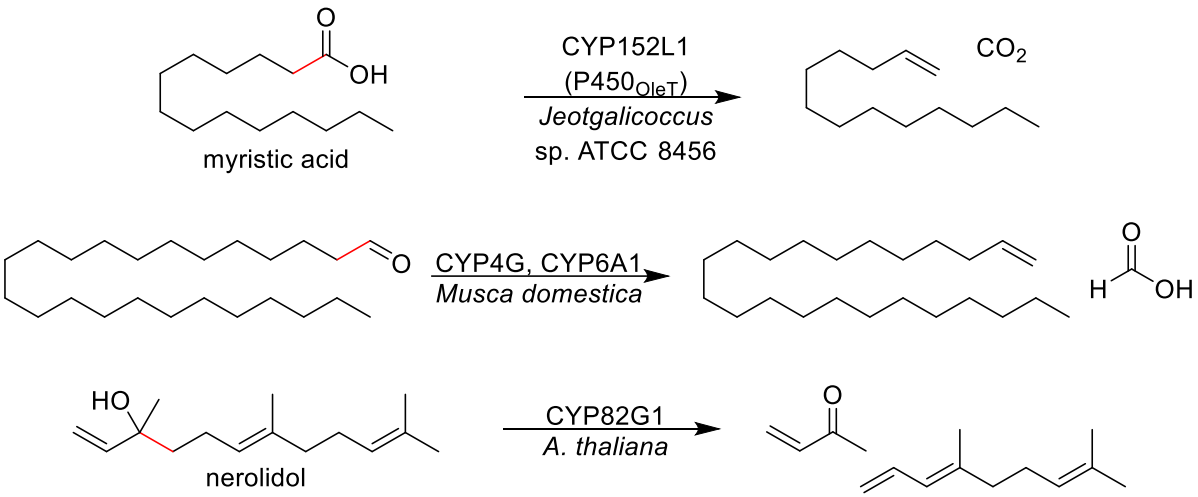


Figure 1.5. A selection of cytochrome P450-mediated C-C bond breaking reactions with long chain fatty acids and other lipids.

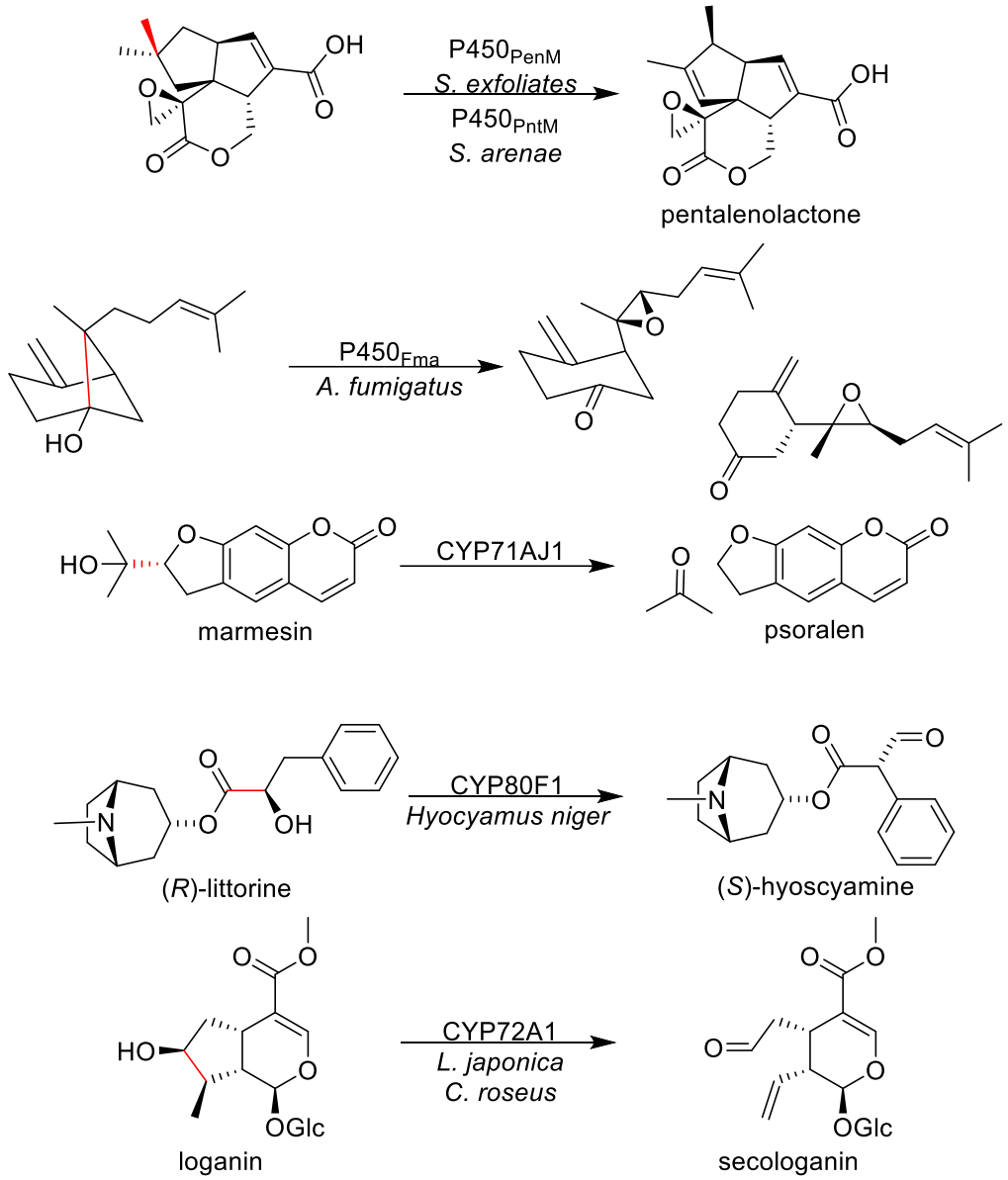


Figure 1.6. A selection of cytochrome P450-mediated C-C bond breaking reactions throughout specialized metabolism.

1.6. REFERENCES

1. Hayaishi, O., Katagiri, M., and Rothberg, S. (1955) Mechanism of the pyrocatechase reaction. *J. Am. Chem. Soc.* **77**, 5450–5451
2. Cytochrome P450 (IPR001128) - InterPro entry - InterPro *InterPro*. [online] https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001128/ (Accessed March 14, 2022)
3. Guengerich, F. P., Wu, Z.-L., and Bartleson, C. J. (2005) Function of human cytochrome P450s: Characterization of the orphans. *Biochem. Biophys. Res. Commun.* **338**, 465–469
4. Guengerich, F. P. (2015) Human cytochrome P450 enzymes. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed), pp. 523–785, Springer International Publishing, Cham, 10.1007/978-3-319-12108-6_9
5. Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., Chang, X., Dong, W., Ho, S. Y. W., Liu, X., Song, A., Chen, J., Guo, W., Wang, Z., Zhuang, Y., Wang, H., Chen, X., Hu, J., Liu, Y., Qin, Y., Wang, K., Dong, S., Liu, Y., Zhang, S., Yu, X., Wu, Q., Wang, L., Yan, X., Jiao, Y., Kong, H., Zhou, X., Yu, C., Chen, Y., Li, F., Wang, J., Chen, W., Chen, X., Jia, Q., Zhang, C., Jiang, Y., Zhang, W., Liu, G., Fu, J., Chen, F., Ma, H., Van de Peer, Y., and Tang, H. (2020) The water lily genome and the early evolution of flowering plants. *Nature*. **577**, 79–84
6. Sono, M., Roach, M. P., Coulter, E. D., and Dawson, J. H. (1996) Heme-containing oxygenases. *Chem. Rev.* **96**, 2841–2888
7. Auchus, R. J., and Miller, W. L. (2015) P450 enzymes in steroid processing. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed), pp. 851–879, Springer International Publishing, Cham, 10.1007/978-3-319-12108-6_12
8. Edin, M. L., Cheng, J., Gruzdev, A., Hoopes, S. L., and Zeldin, D. C. (2015) P450 enzymes in lipid oxidation. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed), pp. 881–905, Springer International Publishing, Cham, 10.1007/978-3-319-12108-6_13
9. Schuler, M. A. (2015) P450s in plants, insects, and their fungal pathogens. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed), pp. 409–449, Springer International Publishing, Cham, 10.1007/978-3-319-12108-6_7
10. Ingelman-Sundberg, M. (2004) Human drug metabolising cytochrome P450 enzymes: Properties and polymorphisms. *Naunyn. Schmiedebergs Arch. Pharmacol.* **369**, 89–104
11. Denisov, I. G., and Sligar, S. G. (2015) Activation of molecular oxygen in cytochromes P450. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed), pp. 69–109, Springer International Publishing, Cham, 10.1007/978-3-319-12108-6_3
12. Denisov, I. G., Makris, T. M., Sligar, S. G., and Schlichting, I. (2005) Structure and chemistry of cytochrome P450. *Chem. Rev.* **105**, 2253–2278
13. Groves, J. T., and McClusky, G. A. (1976) Aliphatic hydroxylation via oxygen rebound. Oxygen transfer catalyzed by iron. *J. Am. Chem. Soc.* **98**, 859–861
14. Rittle, J., and Green, M. T. (2010) Cytochrome P450 Compound I: Capture, characterization, and C-H bond activation kinetics. *Science*. **330**, 933–937
15. Sligar, S. G. (1976) Coupling of spin, substrate, and redox equilibriums in cytochrome P450. *Biochemistry*. **15**, 5399–5406

16. Jefcoate, C. R. (1978) Measurement of substrate and inhibitor binding to microsomal cytochrome P-450 by optical-difference spectroscopy. in *Methods in Enzymology* (Fleischer, S., and Packer, L. eds), pp. 258–279, Biomembranes - Part C: Biological Oxidations, Academic Press, **52**, 258–279
17. Huang, X., and Groves, J. T. (2018) Oxygen activation and radical transformations in heme proteins and metalloporphyrins. *Chem. Rev.* **118**, 2491–2553
18. Barry, S. M., Kers, J. A., Johnson, E. G., Song, L., Aston, P. R., Patel, B., Krasnoff, S. B., Crane, B. R., Gibson, D. M., Loria, R., and Challis, G. L. (2012) Cytochrome P450–catalyzed L-tryptophan nitration in thaxtomin phytotoxin biosynthesis. *Nat. Chem. Biol.* **8**, 814–816
19. Louka, S., Barry, S. M., Heyes, D. J., Mubarak, M. Q. E., Ali, H. S., Alkhalaf, L. M., Munro, A. W., Scrutton, N. S., Challis, G. L., and de Visser, S. P. (2020) Catalytic mechanism of aromatic nitration by cytochrome P450 TxtE: Involvement of a ferric-peroxynitrite intermediate. *J. Am. Chem. Soc.* **142**, 15764–15779
20. Guengerich, F. P., and Yoshimoto, F. K. (2018) Formation and cleavage of C–C bonds by enzymatic oxidation–reduction reactions. *Chem. Rev.* **118**, 6573–6655
21. Lepesheva, G. I., and Waterman, M. R. (2007) Sterol 14α-demethylase cytochrome P450 (CYP51), a P450 in all biological kingdoms. *Biochim. Biophys. Acta BBA - Gen. Subj.* **1770**, 467–477
22. Yoshimoto, F. K., Jung, I.-J., Goyal, S., Gonzalez, E., and Guengerich, F. P. (2016) Isotope-labeling studies support the electrophilic Compound I iron active species, FeO3+, for the carbon–carbon bond cleavage reaction of the cholesterol side-chain cleavage enzyme, cytochrome P450 11A1. *J. Am. Chem. Soc.* **138**, 12124–12141
23. Yoshimoto, F. K., and Auchus, R. J. (2015) The diverse chemistry of cytochrome P450 17A1 (P450c17, CYP17A1). *J. Steroid Biochem. Mol. Biol.* **151**, 52–65
24. Ryan, K. J. (1959) Biological aromatization of steroids. *J. Biol. Chem.* **234**, 268–272
25. Yoshimoto, F. K., and Guengerich, F. P. (2014) Mechanism of the third oxidative step in the conversion of androgens to estrogens by cytochrome P450 19A1 steroid aromatase. *J. Am. Chem. Soc.* **136**, 15016–15025
26. Sivaramakrishnan, S., Ouellet, H., Matsumura, H., Guan, S., Moënne-Loccoz, P., Burlingame, A. L., and Ortiz de Montellano, P. R. (2012) Proximal ligand electron donation and reactivity of the cytochrome P450 ferric–peroxo anion. *J. Am. Chem. Soc.* **134**, 6673–6684
27. Kim, T.-W., Hwang, J.-Y., Kim, Y.-S., Joo, S.-H., Chang, S. C., Lee, J. S., Takatsuto, S., and Kim, S.-K. (2005) *Arabidopsis* CYP85A2, a cytochrome P450, mediates the Baeyer-Villiger oxidation of castasterone to brassinolide in brassinosteroid biosynthesis. *Plant Cell*. **17**, 2397–2412
28. Nomura, T., Kushiro, T., Yokota, T., Kamiya, Y., Bishop, G. J., and Yamaguchi, S. (2005) The last reaction producing brassinolide is catalyzed by cytochrome P-450s, CYP85A3 in tomato and CYP85A2 in *Arabidopsis*. *J. Biol. Chem.* **280**, 17873–17879
29. Helliwell, C. A., Chandler, P. M., Poole, A., Dennis, E. S., and Peacock, W. J. (2001) The CYP88A cytochrome P450, *ent*-kaurenoic acid oxidase, catalyzes three steps of the gibberellin biosynthesis pathway. *Proc. Natl. Acad. Sci.* **98**, 2065–2070

30. Rojas, M. C., Hedden, P., Gaskin, P., and Tudzynski, B. (2001) The *P450–1* gene of *Gibberella fujikuroi* encodes a multifunctional enzyme in gibberellin biosynthesis. *Proc. Natl. Acad. Sci.* **98**, 5838–5843
31. Mizutani, M., and Sato, F. (2011) Unusual P450 reactions in plant secondary metabolism. *Arch. Biochem. Biophys.* **507**, 194–203
32. Jones, G., Prosser, D. E., and Kaufmann, M. (2012) 25-Hydroxyvitamin D-24-hydroxylase (CYP24A1): Its important role in the degradation of vitamin D. *Arch. Biochem. Biophys.* **523**, 9–18
33. Beckman, M. J., Tadikonda, P., Werner, E., Prahl, J., Yamada, S., and DeLuca, H. F. (1996) Human 25-hydroxyvitamin D3-24-hydroxylase, a multicatalytic enzyme. *Biochemistry*. **35**, 8465–8472
34. Stok, J. E., and De Voss, J. J. (2000) Expression, purification, and characterization of BioI: A carbon–carbon bond cleaving cytochrome P450 involved in biotin biosynthesis in *Bacillus subtilis*. *Arch. Biochem. Biophys.* **384**, 351–360
35. J. Cryle, M., and Voss, J. J. D. (2004) Carbon–carbon bond cleavage by cytochrome $P450_{BioI}$ (CYP107H1). *Chem. Commun.* **40**, 86–87
36. Cryle, M. J., and Schlichting, I. (2008) Structural insights from a P450 carrier protein complex reveal how specificity is achieved in the $P450_{BioI}$ ACP complex. *Proc. Natl. Acad. Sci.* **105**, 15696–15701
37. Grant, J. L., Hsieh, C. H., and Makris, T. M. (2015) Decarboxylation of fatty acids to terminal alkenes by cytochrome P450 Compound I. *J. Am. Chem. Soc.* **137**, 4940–4943
38. Hsieh, C. H., Huang, X., Amaya, J. A., Rutland, C. D., Keys, C. L., Groves, J. T., Austin, R. N., and Makris, T. M. (2017) The enigmatic P450 decarboxylase OleT is capable of, but evolved to frustrate, oxygen rebound chemistry. *Biochemistry*. **56**, 3347–3357
39. Amaya, J. A., Rutland, C. D., Leschinsky, N., and Makris, T. M. (2018) A distal loop controls product release and chemo- and regioselectivity in cytochrome P450 decarboxylases. *Biochemistry*. **57**, 344–353
40. Qiu, Y., Tittiger, C., Wicker-Thomas, C., Le Goff, G., Young, S., Wajnberg, E., Fricaux, T., Taquet, N., Blomquist, G. J., and Feyereisen, R. (2012) An insect-specific P450 oxidative decarbonylase for cuticular hydrocarbon biosynthesis. *Proc. Natl. Acad. Sci.* **109**, 14858–14863
41. Reed, J. R., Quilici, D. R., Blomquist, G. J., and Reitz, R. C. (1995) Proposed mechanism for the cytochrome P 450-catalyzed conversion of aldehydes to hydrocarbons in the house fly, *Musca domestica*. *Biochemistry*. **34**, 16221–16227
42. Lee, S., Badieyan, S., Bevan, D. R., Herde, M., Gatz, C., and Tholl, D. (2010) Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in *Arabidopsis*. *Proc. Natl. Acad. Sci.* **107**, 21205–21210
43. Zhu, D., Seo, M.-J., Ikeda, H., and Cane, D. E. (2011) Genome mining in streptomyces. Discovery of an unprecedented P450-catalyzed oxidative rearrangement that is the final step in the biosynthesis of pentalenolactone. *J. Am. Chem. Soc.* **133**, 2128–2131
44. Lin, H.-C., Tsunematsu, Y., Dhingra, S., Xu, W., Fukutomi, M., Chooi, Y.-H., Cane, D. E., Calvo, A. M., Watanabe, K., and Tang, Y. (2014) Generation of complexity in fungal terpene biosynthesis: Discovery of a multifunctional cytochrome P450 in the fumagillin pathway. *J. Am. Chem. Soc.* **136**, 4426–4436

45. Stanjek, V., Miksch, M., Lueer, P., Matern, U., and Boland, W. (1999) Biosynthesis of psoralen: Mechanism of a cytochrome P450 catalyzed oxidative bond cleavage. *Angew. Chem. Int. Ed.* **38**, 400–402
46. Li, R., Reed, D. W., Liu, E., Nowak, J., Pelcher, L. E., Page, J. E., and Covello, P. S. (2006) Functional genomic analysis of alkaloid biosynthesis in *Hyoscyamus niger* reveals a cytochrome P450 involved in littorine rearrangement. *Chem. Biol.* **13**, 513–520
47. Nasomjai, P., Reed, D. W., Tozer, D. J., Peach, M. J. G., Slawin, A. M. Z., Covello, P. S., and O'Hagan, D. (2009) Mechanistic insights into the cytochrome P450-mediated oxidation and rearrangement of littorine in tropane alkaloid biosynthesis. *ChemBioChem*. **10**, 2382–2393
48. Yamamoto, H., Katano, N., Ooi, A., and Inoue, K. (2000) Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. *Phytochemistry*. **53**, 7–12
49. Irmler, S., Schröder, G., St-Pierre, B., Crouch, N. P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J. (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: New enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797–804

CHAPTER 2: IDENTIFICATION OF MULTIFUNCTIONAL P450s FROM *CAMPTOTHECA ACUMINATA* INVOLVED IN TERPENE INDOLE ALKALOID BIOSYNTHESIS[1]

Divergent terpene indole alkaloid (TIA) pathways in *Catharanthus roseus* and *Camptotheca acuminata* generate vinblastine and vincristine, and camptothecin, respectively. In contrast to *Catharanthus*, which feeds secologanin (from methylated loganin) into its species-specific late pathway, *Camptotheca* feeds secologanic acid (from unmethylated loganic acid) into its late pathway. Having identified putative *Camptotheca* secologanic acid synthases (SLASs) and cytochrome P450 reductases (CPRs) in transcriptome databases, *in vitro* analyses of proteins expressed in and purified from *E. coli* have confirmed that two cytochrome P450s, CYP72A564 and CYP72A565, are capable of utilizing both loganic acid and loganin to generate secologanic acid and secologanin. Kinetic analyses of purified full-length *Camptotheca* SLASs have indicated that both process loganic acid with nearly identical catalytic rates and efficiencies as measured by their k_{cat} and k_{cat}/K_M. In contrast, CYP72A564 processes loganin with two-fold greater efficiency than CYP72A565 correlating with the former's three-fold greater affinity for loganin. The closely-related CYP72A730 does not bind or process either compound. Molecular modeling of these three proteins and comparisons with *Catharanthus* secologanin synthase (SLS) has identified key differences that likely determine their SLAS versus SLS selectivity.

2.1. INTRODUCTION

Throughout history, a wide variety of medical plant compounds have been used as pharmaceuticals to treat disease (1). Prominent among these are the terpene indole alkaloids

[1] Chapter adapted from Miller, J. C., Hollatz, A. J., and Schuler, M. A. (2021) P450 variations bifurcate the early terpene indole alkaloid pathway in *Catharanthus roseus* and *Camptotheca acuminata*. *Phytochemistry*. **183**, 112626.

(TIAs) vinblastine and vincristine from *Catharanthus roseus* L. (Apocynaceae)—anti-neoplastic drugs capable of inhibiting tubulin polymerization in actively growing cells (2, 3)—and camptothecin from *Camptotheca acuminata* Decne. (Nyssaceae)—an anti-viral and anti-cancer drug capable of blocking DNA replication (4, 5). The availabilities of these compounds are becoming ever more limited by the dwindling supplies of biologic materials for chemical extraction, the low and unpredictable levels of these compounds in plants grown under different conditions and the dearth of reproducible extraction technologies for the most valuable compounds.

For *Catharanthus* and its vinblastine and vincristine metabolites, these limitations in the biological and chemical supply chains have been somewhat alleviated in recent years. The last of a full series of enzymes capable of transforming geraniol into strictosidine (the common TIA from which species-specific metabolites are formed) were reported in 2014 (6). Further work identifying enzymes capable of transforming strictosidine to vincristine and vinblastine (7, 8) has made possible heterologous expression of these TIAs to facilitate their more robust production.

The general lack of molecular and biochemical information about *Camptotheca* and its specialized metabolism does not presently permit heterologous expression and synthetic biology as a solution to camptothecin production. The general supposition has been that most of the chemical steps and enzymes that conduct them are related to those in the better-studied *Catharanthus* and *Rauvolfia* species of Apocynaceae. In these and other alkaloid-producing species, there are identifiable homologues for all enzymes in the early TIA pathway extending from geraniol to secologanin and strictosidine. Among the eight enzymes in *Catharanthus*, some of which are shown in Fig. 2.1, there are four membrane-bound cytochrome P450s (CYPs, P450s) (G10H, IO, 7DLH, SLS) and five soluble proteins (8HGO, IS, 7DLGT, LAMT, STR).

Those involved in strictosidine production have been most extensively characterized at a biochemical level in *Catharanthus* (6, 7, 9–13) and to a lesser extent in *Rauvolfia* (14–16).

With significant evolutionary distances existing between *Catharanthus* and some other medicinal plant species targeted for TIA pathway analyses, duplications and divergences in some of the relevant enzyme families (P450s, reductases, cyclases, *etc*.) have made it difficult to discern which specific genes within a family are involved in TIA synthesis. Combining transcriptomic and metabolomic data for *Camptotheca* and homology searches to genes of known function, we have examined *Camptotheca* transcriptome datasets available in the Medicinal Plant Genomics Resource (MPGR, http://medicinalplantgenomics.msu.edu) (17) and identified a number of candidates likely involved in the production of early TIA pathway components.

Taking into account recent metabolomics data indicating that strictosidinic acid (a conjugated product formed from secologanic acid and tryptamine, Fig. 2.1) is the central TIA intermediate in *Camptotheca* (18) as opposed to strictosidine (a conjugated product formed from secologanin and tryptamine) in *Catharanthus*, we[2] have cloned the *Camptotheca* secologanic acid synthase (SLAS) candidates most closely related to *Catharanthus* secologanin synthase (SLS) (9, 13) as well as *Camptotheca* cytochrome P450 reductase (CPR) candidates.

Biochemical analyses of these SLAS candidates purified and reconstituted *in vitro* with *Camptotheca* CPRs (*Caa* CPR1, *Caa* CPR2) from *E. coli* have identified two of these as functional in the conversion of loganic acid and loganin into secologanic acid and secologanin, respectively. Corroborating the parallel identification of these P450s as SLASs mediating

[2] The use of first-person plural nouns indicates that the work discussed in that sentence included contributions from other individuals. Their contributions are indicated in corresponding footnotes.

carbon-carbon bond scission on loganic acid and loganin as well as hydroxylation on 7-deoxyloganic acid (19), these results have demonstrated that purified SLAS proteins can be reconstituted with purified *Camptotheca* CPR1 and provided insights on their substrate affinities, steady state kinetics and predicted binding modes. Additional modeling of 7-deoxyloganic acid, the preceding compound in the *Camptotheca* TIA pathway, yielded docking poses consistent with its hydroxylation by these P450s. Analyses of another closely-related member of the CYP72A subfamily from *Camptotheca* have demonstrated that it does not mediate either of these conversions despite its 82-88% identity with the SLAS proteins.

2.2. RESULTS

2.2.1. Omics data used to identify SLS/SLAS, CPR candidates[3]

Searches of the *Camptotheca* transcriptome in MPGR using the previously reported *Camptotheca* SLS (Genbank ADU19849) entry (20) identified one SLAS candidate whose three transcript fragments (caa_locus_2958, caa_locus_2126, caa_locus_6275) were 98%-100% identical to this cDNA and one longer but truncated transcriptome locus (caa_locus_364) that was 82% identical to this cDNA. Other members of the *Camptotheca* CYP72A subfamily in the transcriptome and *Catharanthus* SLS shared significantly lower identities (<55%) with the *Camptotheca* SLS Genbank entry.

Similar BLAST analyses using *Arabidopsis* cytochrome P450 reductase 1 (ATR1, CPR1) (Genbank NM_118585) identified a full-length *Camptotheca* transcriptome locus (caa_locus_4225) (*Caa* CPR1) that was 74% identical to *Arabidopsis* ATR1. BLAST analyses using *Camptotheca* cytochrome P450 reductase 2 (*Caa* CPR2) (Genbank KP162177) (21)

[3] Dr. Bernarda Calla-Zalles performed the cluster analysis on transcriptomics data used for this work. Allison J. Hollatz, Layna Henry, and Dr. Mary A. Schuler all contributed to the gene identification. Allison J. Hollatz, Yaning (Nina) Han, Iyoncy Rodrigo, Eryk Radziszewski, and Pha Thaprawat participated in cloning genes from cDNA.

identified two transcriptome loci (caa_locus_6894, caa_locus_12198) that include 371 of 708 amino acids and are 96%-100% identical to corresponding sections of the full-length CPR2 coding region and 66-78% identical to corresponding sections of *Arabidopsis* ATR2. Subsequent searches of the *Camptotheca* genome located the *Caa* CPR1 gene at Cac_g006149.t1 and a pair of closely-related *Caa* CPR2 genes at Cac_g008486.t1 and Cac_g008487.t1 (99.2% identical).

Analyses of the *Camptotheca* transcript expression data associated with individual genes (22) indicated that the CYP72A565 gene is highly expressed in immature bark, roots and stems; the CYP72A564 gene is highly expressed in immature fruit and the CYP72A730 gene is highly expressed in immature fruit and mature fruit (Fig. 2.2). These expression patterns largely matched those of other genes identified that together produce the iridoid core conserved amongst TIA-producers. In comparison, CPR1 was modestly expressed in most tissues and the adjacent CPR2 genes were expressed at substantially higher levels in nearly all tissues analyzed (with highest concentrations in immature bark, roots and lower stems).

RT-PCR cloning of the SLS/SLAS candidates into the β-galactosidase screenable pGEMT-easy vector using the primers listed in Table 2.1 resulted in a full-length CYP72A565 ORF that had one amino acid change (E82Q) compared to the *Camptotheca* SLS Genbank entry. Cloning of a second of these loci resulted in an N-terminally truncated CYP72A564 lacking its first 64 codons and that had six amino acid changes (N135S, D316E, V324L, S326Y, V328A) compared to the caa_locus_364_iso_10_len_1705_ver_4 sequence in the transcriptome. These changes lie outside the undefined region of the transcriptome sequence. Because of their close identities, engineering a fusion of the N-terminus of CYP72A565 with the body of the truncated CYP72A564 resulted in a full-length CYP72A564 fusion protein (Fig. 2.3). This fusion protein differed from the subsequently identified CYP72A564 gene (Cac_g012666.t1; equivalent to

CYP72A630 (19)) in 18 N-terminal amino acids (contributed by the N-terminus of CYP72A565) and a N135S change. All differences in the N-terminal region precede the proline hinge region that separates the signal anchor domain (SAD) from the P450 catalytic domain.

Our subsequent searches of the *Camptotheca* genome (22) identified genes for these putative SLS/SLAS sequences in two clusters each containing three CYP72A genes (Cac_g012663.t1-Cac_g012666.t1 and Cac_g017135.t1-Cac_g017137.t1). RT-PCR cloning of the CYP72A sequence most closely related to CYP72A564 (Cac_g012666.t1) and CYP72A565 (Cac_g017137.t1) in these three-gene clusters resulted in a full-length CYP72A730 that was an exact amino acid match to the Cac_g012664.t1 gene and a short caa_locus_364_iso_2_len_496_ver_4 transcript.

Alignments of the *Camptotheca* SLAS candidates with one another showed nearly identical sequences in the substrate recognition regions (SRS) of CYP72A564 and CYP72A565 with only one change in SRS1 and three in SRS3 (Fig. 2.3). Compared with the previously characterized *Catharanthus* SLS, designated CYP72A1v3 (Genbank L10081), these two SLAS candidates showed six changes in SRS1, four in SRS3, one each in SRS4 and SRS5 and two in SRS6. *Camptotheca* CYP72A730 showed substantially more divergence from *Catharanthus* CYP72A1v3 with seven changes in SRS1, one in SRS2, six in SRS3, five in SRS4, one in SRS5, three in SRS6 and a deletion of three amino acids between SRS3 and SRS4.

RT-PCR cloning of *Caa* CPR1 into the β-galactosidase screenable pGEMT-easy vector using the primers listed in Table 2.1 resulted in a full-length *Caa* CPR1 that differed in only one amino acid position (Glu165Asp) from caa_locus_4225_iso_5_len_2922_ver_4 in the MPGR transcriptome database and with no amino acid changes compared to the Cac_g006149.t1 gene. In contrast, the full-length CPR2 differed from the previously deposited CPR2 (GenBank

KP162177) (21) at 14 amino acid positions (Ser15Thr, Arg19Lys, Arg21Lys, Met85Val, Lys87Glu, Val101Ala, Val125Ala, Lys152Arg, Arg210Lys, Ala474Val, Asn528Asp, Ile593Met, Gly597Arg, Thr617Ala). Many of these changes occur in isoforms of caa_locus_6894 and caa_locus_12198 in the MPGR transcriptome database as well as the Cac_g008486.t1 and Cac_g008487.t1 genes. Compared to the full-length CPR2 genes, the full-length CPR2 ORF that was cloned and used in subsequent analyses has 10 amino acid changes compared to Cac_g008486.t1 and three amino acid changes (Ile461Val, Asn528Asp, Gly597Arg) compared to Cac_g008487.t1.

2.2.2. Expression and Purification of CYPs, CPRs in E. coli

For *in vitro* assays, His_6-tagged *Camptotheca* CYP72A564, CYP72A565 and CYP72A730 and *Catharanthus* CYP72A1v3 ORFs were cloned into the IPTG-inducible pCW bacterial expression vector, and their full-length proteins were purified from *E. coli* as detailed in 2.4.4. Reduced carbon monoxide difference spectra (23) indicated that all three *Camptotheca* His_6-tagged CYP72A proteins were mixtures of properly folded protein (indicated by a peak around 450nm) with some free heme from misfolded protein (corresponding to a peak around 420nm) (Fig. 2.4). The calculated P450 concentrations for the *Camptotheca* protein preparations used for the majority of these analyses were in the micromolar range (15.9 µM for CYP72A564, 31.9 µM for CYP72A565, 3.34 µM for CYP72A730) and thus amenable to a variety of *in vitro* assays.

His_6-tagged full-length membrane-bound forms of *Camptotheca* CPR1 (His_6 CPR1 full) and CPR2 (His_6 CPR2 full) as well as truncated soluble forms of these electron transfer proteins (His_6 CPR1Δ48, His_6 CPR2Δ68) were cloned into the IPTG-inducible pET28 bacterial expression vector and purified from *E. coli* as detailed in 2.4.4. Cytochrome *c* reduction assays (24) used to quantify CPR activities indicated that the full-length and truncated His_6-tagged CPR1 proteins

efficiently reduced cytochrome *c* whereas the full-length and truncated His_6-tagged CPR2 proteins inefficiently reduced cytochrome *c* at equivalent protein levels (Fig. 2.5).

2.2.3. Two of three candidate SLASs bind, turnover loganic acid and loganin

Loganic acid and loganin binding affinities to each of these three P450s were determined by Type I substrate-induced spectral shifts (see 2.4.6). Both substrates induced blue shifts in the spectra of CYP72A564 and CYP72A565 (Fig. 2.6, A and B) consistent with Type I binding of compounds in the active sites of the P450s (25). In contrast, neither of these compounds induced observable shifts in the spectra of CYP72A730 at concentrations of up to 1.90 mM (Fig. 2.7). For the functional CYP72A564 and CYP72A565 enzymes, plots of the differences between the 389 nm maxima and the 419 nm minima in the difference spectra against the concentration of either substrate (Fig. 2.6C) were fit via nonlinear regression to determine the substrate binding constants (K_S; Table 2.2). Comparing these values indicated that loganin bound with roughly three-fold greater affinity than loganic acid in both CYP72A proteins. Comparisons between CYP72A564 and CYP72A565 indicated that CYP72A564 bound both substrates with approximately three-fold greater affinity than did CYP72A565.

Reconstitution of the purified His_6-tagged CYP72A564 and CYP72A565 proteins with His_6 CPR1 full and DLPC (see 2.4.5) showed that both of these *Camptotheca* P450s metabolize loganic acid and loganin in the presence of NADPH (Fig. 2.8, A and B). For both proteins, products from loganic acid include significant amounts of secologanic acid and smaller amounts of secoxyloganic acid—an overoxidation of the aldehyde to a carboxylic acid that cannot be converted to later products in the TIA pathway. Both proteins also produced secologanin from loganin. Similar reconstitutions of these two *Camptotheca* P450s with His_6 CPR2 full showed no detectable conversion of loganic acid or loganin (Fig. 2.8, C and D). Reconstitutions with His_6

CPR1Δ48 and His_6 CPR2Δ68 showed conversion of neither loganin nor loganic acid (data not shown). Reconstitutions of the more divergent CYP72A730 with His_6 CPR1 full likewise showed no conversion of either potential substrate. Reconstitutions with full-length *Catharanthus* CYP72A1v3 showed conversion of loganin to secologanin but not conversion of loganic acid to secologanic acid (Fig. 2.8, A and B).

The success of this *in vitro* reconstitution system enabled steady-state kinetics analyses. For these kinetic experiments of CYP72A564 and CYP72A565, increases in the rate of NADPH consumption were used to estimate the rates of product formation at substrate concentrations ranging from 50 μM to 5 mM, as detailed in 2.4.7. The initial rates from three technical replicates were combined for nonlinear curve fitting to the Michaelis-Menten equation. CYP72A730 was not assayed because it did not turn over loganic acid or loganin nor did it demonstrate Type I binding spectra with either of these compounds. Recorded in Table 2.3 are the kinetic parameters for both enzymes and both substrates. As evident in these values, the maximum rates and corollary rate constants did not differ greatly between loganic acid and loganin for either CYP72A564 or CYP72A565. However, CYP72A564 performed the reaction with greater turnover than CYP72A565 did for both substrates. The catalytic efficiencies, as represented by k_{cat}/K_M, of these enzymes did not differ for loganic acid, but CYP72A564 has twice the rate as CYP72A565 for loganin for which it has greater affinity.

2.2.4. Molecular modeling suggests several amino acids as important in multifunctional activity

Molecular models for *Catharanthus* CYP72A1 and *Camptotheca* CYP72A564 and CYP72A565, constructed according to published procedures (26), indicated that the backbones of these three proteins overlay well (RMSD less than 2.0 Å, green, Fig. 2.9A) except for the following regions: the C-terminus of the A'-A loop, β1-2, the N-terminus of the C helix, the D-

E' loop, the N-terminus of the G-helix, the N-terminal half of the H-I loop, the middle of the loop between β1-3 and β1-4, and in a loop region immediately C-terminal of the PERF domain. None of these regions have RMSD values greater than 3.5 Å (orange), and all are outside the predicted SRSs.

Dockings of loganin in the *Catharanthus* CYP72A1 catalytic site and loganic acid in the *Camptotheca* CYP72A564 and CYP72A565 catalytic sites predicted that both substrates position C10 within 4.5 Å of the oxygen bound to the heme with comparable interaction energies (Table 2.4). As both proposed mechanisms (27, 28) for the C-C bond scission demand abstraction of a hydrogen from C10 before cleaving the C7-C8 bond in loganin or loganic acid, all of these models appeared to present poses consistent with facilitating catalysis. Even so, the orientations of the substrates were different with loganic acid oriented more vertically in the two *Camptotheca* CYP72A proteins than loganin is in the *Catharanthus* CYP72A1 model.

Overlays of the predicted contact residues highlighting amino acids identical in each (Fig. 2.9, C and E) and different in each (Fig. 2.9, D and F) showed many identical residues within substrate contact distance and relatively few different residues in this area. Among the variant residues, several differences occurred in SRS1 at positions 132, 133, 134, 135 and 138[4] (Fig. 2.3; Fig. 2.9, B, C, and E). The *Catharanthus* protein differed from the two *Camptotheca* proteins at all of these positions, and the two *Camptotheca* proteins differed only at position 135 (Asn/Ser). One additional difference among these three proteins occurred in SRS3 at position 270 with all three proteins having different residues (Thr/Lys/Arg). Similar overlays of identical and different

[4] *N.B.*, CYP72A1, CYP72A564, and CYP72A565 are all of equal length (524 amino acids without the His_6 tag). They also aligned to each other without any gaps (Fig. 2.3).

residues in these two *Camptotheca* proteins (Fig. 2.9, B-E) highlighted their substantial identity in all SRSs comprising the catalytic site.

Molecular modeling and docking of *Camptotheca* CYP72A730 with loganic acid, which it neither binds nor metabolizes, suggested that this non-substrate can be positioned within the catalytic site (Fig. 2.10). The loganic acid in the CYP72A730 model had a similar distance to the oxygen positioned above the heme and interaction energy compared to the *Camptotheca* CYP72A564 and CYP72A565 models (Table 2.4). Overlays of the backbones in these three *Camptotheca* models showed the most substantial (>4.5 Å) RMSD variance extending from the F-helix through the G-helix (red in Fig. 2.10A) and little in most other regions ($\leq$2.0 Å, green in Fig. 2.10A). Overlays of these three predicted catalytic sites highlighting amino acids identical in each (Fig. 2.10, C and E) and different in each (Fig. 2.10, D and F) showed multiple residues that differed within loganic acid contact distance and within the upper regions of its predicted catalytic site. Among the variant residues, several differences occurred in SRS1: the *Camptotheca* CYP72A730 protein differed from the CYP72A564 and CYP72A565 proteins at positions 132 and 133 (numbered according to CYP72A730)[5] (Fig. 2.3; Fig. 2.10B). Differences between these three protein models also occurred in SRS2 at position 234, SRS3 at position 268 and SRS4 at positions 323 and 327 (for more about these differences, see Chapter 4).

Other differences between CYP72A730 the *Camptotheca* SLASs CYP72A564 and CYP72A565, while occurring in SRS regions were more than 4.5 Å away from the docked substrate: SRS4, CYP72A730=Asp315 *versus* Glu320; SRS6 at positions CYP72A730=Leu499 *versus* Phe504 and CYP72A730=Ser501 *versus* Val506). Notable among the sequence variations

[5] Compared to the other three CYP72As studied, CYP72A730 is two amino acids shorter at its N-terminus and has a three amino acid deletion between the predicted SRS3 and SRS4.

within substrate contact distance were those in SRS2 at Ile234 (Met236 for CYP72A564 and CYP72A565) and SRS3 at Gln268 (Arg/Lys270 for CYP72A564 and CYP72A565, respectively). Because of the substantial RMSD variance in the placement of the F-to G-helices shown in Fig. 2.10A, the first of these differences (Ile234 in CYP72A730), was substantially closer to the docked loganic acid than Met236 in CYP72A564 and CYP72A565. The second of these (Gln268 in CYP72A730) was displaced so significantly compared to Arg270 in CYP72A564 and Lys270 in CYP72A565 that it was not predicted to lie within substrate contact distance.

2.3. DISCUSSION

2.3.1. Two of three candidates are secologanic acid synthases

These characterizations of the *Camptotheca* CYP72A564 and CYP72A565 proteins indicate that both are capable of cleaving C-C bonds in loganic acid and loganin to form secologanic acid and secologanin, respectively. In contrast to these two *Camptotheca* CYP72A proteins, *Catharanthus* CYP72A1v3 metabolizes loganin but not loganic acid, and *Camptotheca* CYP72A730 metabolizes neither compound. Coupled with the transcriptome data available for these *Camptotheca* CYP72A genes (Fig. 2.2), it appears that CYP72A565 is available for metabolizing these substrates in a wide array of tissues including immature bark, immature and mature fruit, upper and lower stem, and roots whereas CYP72A564 is available in immature fruit. Despite its inability to metabolize these alkaloids, CYP72A730 is available in immature and mature fruit, suggesting an alternative activity.

These characterizations also indicate that His_6-tagged CPR1 full can serve as an effective electron transfer partner for *Camptotheca* and *Catharanthus* SLS/SLAS proteins in *in vitro* reconstitutions with DLPC. Isolated under the same conditions, the *Camptotheca* His_6-tagged

CPR2 full variant that we have cloned does not serve as an electron transfer partner for these P450s. This last finding contrasts directly with the reported abilities of *Arabidopsis* ATR1 (CPR1) and ATR2 (CPR2) to serve as electron transfer partners for the *Catharanthus* SLS (13) and *Camptotheca* SLAS proteins (19) in yeast extracts and microsomes, respectively. *Camptotheca* CPR1 and another CPR2 variant were also previously reported as electron transfer partners for the *Catharanthus* SLS in yeast microsomal assays (29). Given that P450 reductases undergo rapid equilibria between open and closed states that are both pH- and salt-dependent (30), it may be that these purified His_6-tagged CPR2 full variant exists in a conformationally inactive state under the reported reaction conditions. Transcriptomic data (Fig 2.2) substantiates that both CPRs are broadly available in all tissues in which CYP72A564 and CYP72A565 show heightened expression presenting CPR1 and/or CPR2 as the electron transfer partner for these *Camptotheca* SLASs.

The elimination of the closely-related CYP72A730 as a SLAS candidate implicates it in another current or evolving TIA function in immature and mature fruit where it is strongly expressed. The positioning of this gene in a three-gene cluster between the more-related CYP72A564 gene and the less-related CYP72A728 gene (a 7DLH candidate encoded by Cac_g012163.t1) indicates that it duplicated from the CYP72A564 gene after the larger scale genomic duplication that produced the two current clusters of CYP72A genes.

2.3.2 These Camptotheca SLASs also demonstrate (stronger) SLS activity

Despite the ability of CYP72A564 and CYP72A565 to modify both unmethylated and methylated derivatives of loganic acid, metabolomic data have indicated that multiple *Camptotheca* tissues contain loganic acid, secologanic acid and strictosidinic acid but not loganin, secologanin or strictosidine (18). Additional metabolomic data from Jin *et al*. (29) have

verified the presence of loganic acid, secologanic acid and strictosidinic acid in shoot apices and leaves but not in roots. The use of these unmethylated precursors for the formation of subsequent products represents an important bifurcation in the *Camptotheca* pathway from the use of methylated precursors in *Catharanthus, Lonicera*, and many other medicinal plant species. Presuming that loganin does not exist in *Camptotheca* even at low levels, the ability of CYP72A564 and CYP72A565 to metabolize loganin is surprising and indicates that it has retained an ancestral activity (toward loganin) while obtaining another activity (toward loganic acid) or *vice versa*—matters further explored in Chapter 3.

The yields from heterologous expression of these full-length *Camptotheca* P450s in *E. coli* were sufficiently large to enable examination of substrate-enzyme interactions not possible with many other types of proteins. Strikingly, the binding isotherms reveal that both CYP72A564 and CYP72A565 bind methylated loganin, a compound not detected in *Camptotheca,* with three-fold greater affinity than unmethylated loganic acid, the compound detected in *Camptotheca*.

Using an *in vitro* reconstitution system that includes species-matched P450 and CPR proteins, although loganin binds to CYP72A564 with a three-fold greater binding affinity than loganic acid, the enzyme turns over loganic acid and loganin at equivalent rates (Table 2.3). Turnover of both substrates is likewise the same for CYP72A565. The rate constants for loganic acid and loganin are larger for CYP72A565 than for CYP2A564.

The differences in binding affinity and catalytic rate for loganic acid combine to produce equivalent catalytic efficiencies (as measured in k_{cat}/K_M) for both SLAS proteins. The enzymes do not process loganin with the same k_{cat}/K_M: CYP72A564 is twice as efficient at turning over loganin as CYP72A565 in spite of having the slower rate constant. This effect is well explained by the three-fold smaller substrate binding constant of CYP72A564 because the overall catalytic

efficiency is a combination of the rate of turnover (given by k_{cat}) and the affinity for the substrate (a factor included in K_M).

This study's steady-state kinetic parameters for the *Camptotheca* SLASs differ substantially from a previous report of their activities in yeast microsomes (19). Only the Michaelis concentration of loganic acid for CYP72A564 (equivalent to CYP72A610 in Yang *et al*. (19)) and CYP72A565 are of equal magnitude with all other parameters differing by orders of magnitude. Several experimental factors can individually accentuate the differences in these studies and together compound into the drastic differences observed in the kinetic parameters.

First, Yang and colleagues coexpressed their SLAS proteins with *Arabidopsis* ATR1 (CPR1) whereas I employed *Camptotheca* CPR1 in our assays. Clearly, this *Arabidopsis* CPR is able to deliver the electrons that the *Camptotheca* P450s require for turnover, but the exogenous electron donors may influence SLAS activity by having different coupling efficiencies and/or by inducing different P450 conformations. Variations in CPR identities and activities may produce the observed difference in kinetic parameters if the activity of the CPR in the previous report was less than reported herein. This condition is likely given that, in my hands, the activities of CPRs in yeast microsomes, as measured by cytochrome *c* reduction, are much lower than for those purified from *E. coli* (data not shown).

Second and probably most relevant, limitations in the yeast expression system employed by Yang *et al*. for assessing P450 activities prevent robust control of the ratios of active CPR and P450. Knowing this ratio is important for at least two reasons: (1) The activity of CPR, when sufficiently small, affects the observed rate of reaction by shifting the rate-determining step to include electron delivery; (2) Not knowing the concentration of P450 prevents the comparison of k_{cat} where the maximum rate is reported per amount of enzyme. Granting the difficulties with

directly controlling CPR-P450 ratios in microsomes, neither the CPR activity (from cytochrome *c* reduction assays) nor the P450 concentration (as specifically determinable via CO difference spectra) were reported and taken into account by Yang and coworkers (19), though both were possible. As noted earlier, the CPR activity in microsomes is very likely to be much smaller than that of those purified from *E. coli*, and this phenomenon likely explains the slower catalytic rates reported for the microsomal assays. Estimating the concentration of P450 via a total protein method guarantees an overestimation of the P450 present as the *Arabidopsis* CPR and endogenous yeast proteins are certainly present and not corrected for. This overestimation of P450 concentrations guarantees an underestimation of catalytic rates and efficiencies (*i.e.*, k_{cat} and k_{cat}/K_M) because the V_{max} is divided by the enzyme concentration to determine these values.

Third, the lipid environments in the assays differ greatly. Yeast microsomal isolations embed the CPRs and P450s in the endoplasmic reticulum of this eukaryotic host. *In vitro* reconstitutions with purified proteins mix the P450 and CPR samples containing lipids from *E. coli* membranes and residual detergents (CHAPS and Triton X-100, respectively) with DLPC to facilitate their association in a lipid environment. Some CPR-P450 interactions are known to be affected by local lipid compositions (31). Thus, variations in the lipid composition of the endoplasmic reticular versus DLPC reconstituted systems may have some as-yet-undefined role on these SLAS activities.

Fourth and lastly, the methods employed to determine the rates of SLAS activity differed substantially between these studies. Monitoring product formation as done by Yang *et al*. is advantageous in that it provides a direct measure of the conversion rate of substrate into product. Monitoring NADPH consumption, as in this report, monitors CPR activity directly and is subject to complications from the uncoupling of electron delivery from the CPR to the P450. Having

controlled for potential uncoupling by subtracting the rate of NADPH consumption in the absence of substrate from the rate in the presence of substrate, my measurements should accurately reflect the catalytic activities of these SLAS enzymes. If perchance this correction insufficiently controls for uncoupling, the rates determined herein might represent overestimates of their respective enzymatic activities. However, for this to be a plausible explanation, uncoupling would have to significantly and differentially increase based on the concentration of loganic acid or loganin. Because this disruption of CPR-P450 coupling by the substrates is highly unlikely, these differences in assay methods are not likely to have affected the observed rates of reaction.

2.3.3. Putative structural explanations

The numerous differences between *Catharanthus* CYP72A1v3 and the *Camptotheca* SLASs in SRS1 likely explain the latter's ability to metabolize loganic acid as well as loganin. Of these differences, the His132 of *Camptotheca* adjacent to the carboxylic acid of docked loganic acid (Fig. 2.9) is particularly striking since its replacement with Asp132 in *Catharanthus* CYP72A1 is predicted to prevent loganic acid binding.

The homology models of these two SLASs present several differences in contact and noncontact residues in SRS1, SRS3, SRS4 and SRS5 between the *Camptotheca* SLASs. Among these variations are changes in contact residue Arg270Lys (with amino acid in CYP72A564 listed first) in SRS3 and in noncontact residues Gln128Met and Ser135Asn in SRS1 and noncontact residue Arg267Gly in SRS3. The homology models also predict variations in the contacts of residues conserved in the sequences of both SLASs (CYP72A564: Leu141 in SRS1, Ser147 just beyond SRS1, Val265 in SRS3, and Leu399 in SRS5; CYP72A565: Ile141 just beyond SRS1, Thr337 in SRS4, and Val396 in SRS5). Interestingly, the differences in predicted

contacts at positions 141 and 270 align with the hydrogen-bonding Val106 and Lys241 pair in human CYP2C5 that are predicted to dissociate in the course of product exit (32). Further experiments (see Chapter 3) explore whether these variations correspond to significant differences in enzyme activity.

The more numerous differences of *Camptotheca* CYP72A730, including those in the SRS regions, likely account for its inability to bind and metabolize loganic acid and loganin. Most notable among these differences is the three amino acid deletion in the F-G loop (Fig. 2.10A). This deletion may constrain the flexibility of this loop region. Based on studies by Tietz *et al*. (33, 34), the interactions of residues in this region of the P450 fold vary as substrates bind to bacterial P450cam (CYP101A1) and MycG. These changes in interacting residues substantiate significant movements and rearrangements around the F-G loop as compounds enter the active site through the substrate access channel. It is, therefore, possible that the three amino acid deletion limits these movements in CYP72A730 to such an extent that the somewhat large iridoid glycosides cannot enter the active site—a hypothesis explored in Chapter 4. This explanation is consistent with the homology modeling and docking where loganic acid is docked in the active site with parameters comparable to the *Camptotheca* SLASs. Although the *in vivo* activity of CYP72A730 remains ambiguous, it is very likely that this gene with the heightened amount of the mRNA encoding this protein in TIA-producing tissues has some role in camptothecin biosynthesis.

2.3.4. Conclusions

Having identified three putative SLASs from *Camptotheca*, confirmed SLAS *and* SLS activity for two of them, and developed homology models to theorize the structural basis of their broadened substrate scope, the work outlined in this chapter presented several avenues for further

investigation. My pursuit of several of these avenues is the subject of the following three chapters: Which residues are important for discriminating loganic acid from loganin and thus SLAS from SLS activity (Chapter 3); whether the small (three amino acid) deletion theorized as eliminating SLAS/SLS activity in CYP72A730 does so (Chapter 4); which of two mechanisms proposed (27, 28) for the oxidative C-C bond scission catalyzed by these SLASs/SLSs is operative (Chapter 5).

2.4. EXPERIMENTAL

PCR reagents, plasmid purification, restriction enzymes, and T4 DNA ligase were used as recommended by New England Biolabs (NEB) unless otherwise noted. Reverse transcriptase was used as recommended by Invitrogen. Vectors used in this study included the cloning and sequencing vector pGEM-T Easy (Promega), bacterial expression vectors pCWori (35), and pET28a (Novagen). *E. coli* cell strains used in this study included Top10 (Invitrogen), DH5α, and BL21(DE3) (NEB).

For high resolution LC-MS, samples were analyzed using the Dionex Ultimate 3000 series HPLC system (Thermo) with Q-Exactive MS system (Thermo) in the Metabolomics Laboratory of Roy J. Carver Biotechnology Center, University of Illinois Urbana-Champaign. Spectroscopy was recorded in a Cary UV-Vis Bio100 dual beam spectrophotometer or Molecular Devices SpectraMax M-series plate reader as appropriate.

Loganic acid was synthesized by Arctom Chemical (Cambridge, MA); loganin was sourced from Cayman Chemicals (Ann Arbor, MI). Horse heart cytochrome *c*, NADPH and DLPC (dilauroyl phosphatidyl choline) were from Millipore-Sigma (St. Louis, MO).

Nucleotide sequence data are available in the GenBank database under the following accession numbers: CYP72A564, MN815881; CYP72A565, MN815882; CYP72A730, MN815883; *Caa* CPR1, MN815884; *Caa* CPR2, MN815885

2.4.1. Plant materials and RNA isolation

Camptotheca acuminata, Decne. (Nyssaceae) mRNA were the generous gift of Dr. Dean DellaPenna (Michigan State University) and correspond to collections reported previously (17). *Catharanthus roseus* (L.) (Apocynaceae) Pacifica Polka Dot Hybrid were grown outdoors in Urbana, Illinois, USA (40.1, -88.2), and young leaves were collected in October 2010. Leaves were washed with autoclaved water, flash frozen, and crushed to a powder by bead beating. RNAs were extracted and isolated using a TRIzol reagent (Invitrogen) according to the manufacturer's recommended procedure, and further purified by LiCl extraction before storage at -80°C.

2.4.2. Transcriptomics cluster analysis and genomics database searches

Cluster analyses of the final *Camptotheca* transcriptome data obtained from 14 different *Camptotheca* tissues (MPGR, http://medicinalplantgenomics.msu.edu), performed by Dr. Bernarda Calla-Zalles (University of Illinois), filtered out transcripts with low expression (<15 FPKMs) in at least one tissue. The remaining dataset was clustered by using a hierarchical cluster algorithm in Rv3.2.3 (hclust) on a distance matrix created with all the data after filtering. Separately, the final *Camptotheca* transcriptome dataset was sorted for high expression (>300 FPKMs) in immature bark, young bark and immature fruit, all of which are camptothecin-producing tissues.

Searches of the *Camptotheca* transcriptome Medicinal Plant Genomics Resource (MPGR, http://medicinalplantgenomics.msu.edu) were conducted using protein BLAST functions with

Catharanthus SLS (GenBank CAC80833) (9), a putative *Camptotheca* SLS (Genbank ADU19849) (20) and P450 annotation procedures (courtesy of Dr. David Nelson, University of Tennessee). These BLAST analyses identified fragments of *Camptotheca* Genbank ADU19849 in three transcriptome loci (caa_locus_2958_iso_3_len_413_ver_4, caa_locus_2126_iso_2_len_595_ver_4, caa_locus_6275_iso_1_len_256_ver_4) that were 98%-100% identical to the previous cDNA clone obtained from young leaves (20) as well as a N-terminally truncated partially ambiguous transcriptome locus (caa_locus_364_iso_10_len_1705_ver_4) that was 82% identical to the *Camptotheca* SLS (Genbank ADU19849). Other members of the *Camptotheca* CYP72A subfamily present in the transcriptome and *Catharanthus* SLS shared significantly lower identities (<55%) with the putative *Camptotheca* SLS ORF. Further analysis of the caa_locus_364 isoforms indicated that they included transcripts from at least three closely related CYP72A genes.

Subsequent searches of the *Camptotheca* genome (22) identified genes for these putative SLS/SLAS sequences in two clusters each containing three CYP72A genes (Cac_g012663.t1-Cac_g012666.t1 and Cac_g017135.t1-Cac_g017137.t1). Several of the caa_locus_364 isoforms appeared to represent allelic variations (<3% a.a. differences) of the Cac_g012664.t1 and Cac_g012666.t1 sequences.

Similar BLAST searches using *Arabidopsis* cytochrome P450 reductase 1 (ATR1, CPR1) (Genbank NM_118585) identified a full length *Camptotheca* transcriptome locus (caa_locus_4225_iso_5_len_2922_ver_4) (*Caa* CPR1) that was 74% identical to *Arabidopsis* ATR1. BLAST analyses using *Camptotheca* cytochrome P450 reductase 2 (*Caa* CPR2; Genbank KP162177) (21) identified fragments in three transcripts covering two transcriptome loci (caa_locus_6894_iso_6_len_766_ver_4, caa_locus_6894_iso_7_len_417_ver_4,

caa_locus_12198_iso_1_len_286_ver_4) that include 371 of 708 amino acids and are 96%-100% identical to various sections of the full-length CPR2 coding region and 66-78% identical to sections of *Arabidopsis* ATR2. Searches of the *Camptotheca* genome located the *Caa* CPR1 gene at Cac_g006149.t1 and a pair of closely related *Caa* CPR2 genes at Cac_g008486.t1 and Cac_g008487.t1 (99.2% identical).

2.4.3. Cloning of full-length cDNAs

First strand cDNAs were generated by reverse transcription of total mRNA samples from young *Catharanthus* leaves or mixed *Camptotheca* tissues with an oligo$(dT)_{18}$ primer directed against the poly(A) tail on mRNAs. For the first-step reverse transcription, 2.5-10 µl total mRNA, 2-4 µL 20 mM oligo$(dT)_{18}$ EcoRI primer, and molecular biology grade water for a final volume of 18 µL were incubated at 65°C for five min and then on ice for 10 min. After this, 8 µL 5x First-Strand buffer (Invitrogen), 4 µL 50 mM $MgCl_2$, 4 µL 0.1 M DTT, 1 µL 5U μL^{-1} RNasin (Promega), 2-4 µL 10 mM dNTP mix, and 1.5 µL 200 U μL^{-1} Superscript III Reverse Transcriptase (Invitrogen) and molecular biology grade water were added for a total reaction volume of 40 µl. The reactions were incubated at 50°C for 2 h and stored at -20°C.

Individual ORFs were PCR amplified in 50 µL reactions containing 2.5 µL cDNA, 2 µL of each 20 µM gene-specific primer (Table 2.1), 5 µL 10X Taq buffer (NEB), 2 µL 10 mM dNTP mix, 1.5 µL 50 mM $MgCl_2$, 1 µL Taq polymerase (NEB), and molecular biology grade water. PCR amplifications were run with a 5 min denaturation step at 95°C, 30 cycles each of 1 min at 95°C, 1.5 min at a primer-specific temperature, and 2 min at 72°C, followed by a final extension step for 10 min at 72°C.

For our original clonings into pGEM-T Easy vector (Promega), PCR products were cleaned up using the Qiagen QIAquick PCR Purification Kit according to the manufacturer's protocol

with samples eluted in a total volume of 80 µL molecular biology grade water. Ligation reactions were set up with 1-6 µL PCR clean-up product, 0.5 µL pGEMT-easy vector, 7.5 µL 2x T4 DNA ligase buffer (Promega) and 1 µL T4 DNA ligase (Promega), and molecular biology grade water was added for a total reaction volume of 15 µL. Reactions were incubated overnight at 4°C and electroporated into competent Top10 cells.

For subsequent clonings into the pCW and pET28a expression vectors, PCR products generated from pGEM-T clones with gene-specific primers were prepared as above, digested with appropriate restriction enzymes (NEB: NdeI-XbaI for pCW; NcoI-XhoI for pET28a), purified from agarose using β-agarase (NEB) and isopropyl alcohol precipitation, ligated into the corresponding vector with T4 DNA ligase (NEB), and electroporated into Top10 cells. Primers for the pCW expression were designed to include His_6-tags on the C-terminus of P450s; primers for the pET28a expression were designed to include His_6-tags on the N-terminus of CPRs.

Insert-containing plasmids from positive colonies were sequenced using M13For-21 and M13Rev-24 primers (pGEM-T), pCW 5' and pCW 3' primers (pCW), pET28 5' and pET28 3' primers (pET28a), and as needed, gene-specific internal (mid) primers (Table 2.1). Sequence variations in the final expressed proteins compared to reference sequences are as follows: *Caa* CYP72A564 (contains first 51 a.a. of CYP72A565 and a N135S change compared to the Cac_g012666.t1 gene), *Caa* CYP72A565 (no changes compared to the Cac_g017137.t1 gene, a E82Q change compared to ADU19849), *Caa* CYP72A730 (no changes compared to the Cac_g012664.t1 gene); *Cra* CYP72A1v3, *Caa* CYP72A564, CYP72A565 and CYP72A730 have C-terminal HHHHHH extensions. *Caa* CPR1 full-length (E165D change compared to caa_locus_4225 transcript, R657H change compared to the Cac_g006149.t1 gene and N-terminal MGHHHHHH extension), *Caa* CPR2 full-length (S15T, R19K, R21K, M85V, K87E, V101A,

V125A, K152R, R210K, A474V, N528D, I593M, G597R, T617A changes compared to CPR2 Genbank KP162177 ORF and K87E, N528D, G597R changes compared to the Cac_g008487.t1 gene and N-terminal MGHHHHHH extension).

2.4.4. Expression and purification of full-length His_6-tagged CPRs and P450s in E. coli

For His_6-tagged P450 protein expression, IPTG-inducible CYP72A/pCW constructs were cotransformed with pGRO7 into *E. coli* DH5a cells and inoculated into two 1.0 L TB cultures in Fernbach flasks containing 30 μg mL^{-1} chloramphenicol and 100 μg mL^{-1} ampicillin. Cultures were incubated at 37°C for 4-6 h with shaking (160 rpm) until their OD_{600} reached 1.5-1.8,[6] induced by adding 0.143 g L^{-1} IPTG, 4 g L^{-1} L-arabinose, and 0.168 g L^{-1} δ-aminolaevulinic acid, incubated at 25°C for an additional 24-29 h with shaking (200 rpm), and harvested by centrifugation at 8,000 *xg* for 10 min. Cell pellets were stored at -80°C.

For P450 purification, cell pellets from the 2.0 L cultures were resuspended in 125 mL sterile water and 150 mL 2x cell lysis buffer [0.2 M KPO_4 (pH 7.4), 0.1 M NaCl, 40% glycerol, 20 mM β–mercaptoethanol (βME), 1 mM phenylmethylsulfonylfluoride (PMSF)], treated for at least 1 h at 25°C with 120 mg lysozyme (0.4 mg mL^{-1} final concentration), and sonicated on ice six times for 15 s with rests of at least 30 s in between. Membranes were collected by centrifugation at 34,000 *xg* for 20 min, resuspended in 100 mL membrane wash buffer [0.1 M KPO_4 (pH 7.4), 0.1 M NaCl, 20% glycerol, 10 mM βME, 0.5 mM PMSF] using a Dounce homogenizer, recollected by centrifugation, and directly solubilized with CHAPS or frozen at -80°C. To resuspend the membrane-bound P450s, membranes were first homogenized in 110 mL membrane resuspension buffer [0.1 M KPO_4 (pH 7.4), 0.5 M NaCl, 20% glycerol, 1% (w/v)

[6] Later work in expressing these P450s has shown that ~1.0 is a better OD_{600} for inducing protein production.

CHAPS, 10 mM βME, 0.5 mM PMSF] using a Dounce homogenizer. To ensure the chaperones were released, $MgCl_2$ (to 20 mM) and ATP (to 5 mM) were added to the homogenized membrane solution.[7] Stirring at 4°C for at least 3 h (and up to overnight) ensured widespread resuspension of the P450s before centrifugation at 34,000 *xg* for 30 min removed insoluble debris. The cleared supernatant was loaded onto a 30 mL Ni-NTA Superflow (Qiagen) agarose affinity column pre-equilibrated with two volumes membrane resuspension buffer and washed with thee column volumes of wash buffer [0.05 M Tris-HCl (pH 7.4), 0.3 M NaCl, 20% glycerol, 1% CHAPS, 10 mM βME, 5 mM L-histidine (not containing chloride)], and eluted with thee column volumes of elution buffer [0.05 M Tris-HCl (pH 7.4), 0.3 M NaCl, 20% glycerol, 0.5% CHAPS, 100 mM L-histidine, 10 mM βME]. Protein fractions containing P450 as determined by CO-difference spectra (23, 24) were combined and dialyzed in 12-14,000 kDA cutoff dialysis tubing at 4°C against 500 mL dialysis buffer [100 mM Tris-HCl (pH 7.4), 50 mM NaCl, 20% glycerol, 10 mM βME] before concentrating in a 10,000 MWCO centrifugal filter device to approximately 2 mL at 4,100 x*g*. The concentrated sample was then exchanged into P450 dialysis buffer [100 mM Tris-HCl (pH 7.4), 50 mM NaCl, 20% glycerol, 1 mM DTT] over a G-25 Sephadex resin and, if necessary, concentrated as before. The dialyzed and concentrated protein samples were aliquoted, flash frozen, and stored at -80°C.

P450 contents in fractions eluted from the Ni-NTA column and the final concentrated samples were monitored by reduced CO difference analyses according to standard methods (23, 24). P450 concentrations were estimated using an extinction coefficient at 450 nm of 91 mM^{-1}

[7] Later, the addition of Mg^{2+} and ATP was omitted as not doing so did not diminish the yields or activities of these P450s.

cm^{-1} when using CO-reduced minus reduced spectra or 106,000 mM^{-1} cm^{-1} when using CO-reduced minus CO-oxidized spectra.

For His_6-tagged CPR protein expressions, IPTG-inducible CPR/pET28 constructs were transformed into competent BL21(DE3) cells and inoculated into two 1.0 L cultures of Terrific Broth (TB) in Fernbach flasks containing 50 μg mL^{-1} kanamycin. Cultures were incubated at 37°C for 4-6 h with shaking (180 rpm) until their OD600 readings reached 0.6-0.8, induced with 0.143 g L^{-1} IPTG and 10 mg L^{-1} riboflavin, incubated at 25°C for an additional 24 h with shaking (200 rpm) and harvested by centrifugation at 8,000 *xg* for 10 min. Cell pellets were stored at -80°C.

For full-length CPR protein purifications, the combined cell pellets were resuspended in 60 mL no imidazole buffer [100 mM Tris-HCl (pH 8.0), 10% glycerol, 0.25mM EDTA],[8] treated for 1 h at 25°C with 1 mg mL^{-1} lysozyme, and then sonicated at least three times for 15 s each. Membranes were harvested by centrifugation at 34,000 *xg* for 30 min, resuspended in 60 mL no imidazole buffer containing 10% (v/v) Triton X-100, incubated at 4°C overnight with stirring, and recentrifuged at 34,000 *xg* for 30 min. The cleared supernatant was loaded onto a 10 mL Ni-NTA agarose affinity column (Qiagen) pre-equilibrated with no imidazole buffer and washed with two column volumes of the same buffer containing 25 mM imidazole. The His_6-tagged CPRs were eluted with two column volumes of the same buffer containing 250 mM imidazole and collected in 20 aliquots of ~1 mL. By a combination of SDS-PAGE and/or cytochrome *c* reduction assays (24), those fractions containing His_6-tagged CPRs were combined and dialyzed in a 10,000 MWCO dialysis cassette at 4°C against four volumes of 375 mL [50 mM Tris-HCl

[8] Later preparations replaced this and all subsequent Tris-HCl buffers with a sodium MOPS buffer [100 mM Na·MOPS (pH 7.3), 10% glycerol] (21).

(pH 7.5), 10% glycerol] (1,500 mL total). The dialyzed protein samples were concentrated using a 10,000 MWCO centrifugal filter device until the volume was approximately 1.5 mL, aliquoted, and frozen at -80°C. For N-terminally truncated CPR protein purifications, cell pellets were incubated with lysozyme for 1 h at 4°C and not solubilized with Triton X-100.

2.4.5. In vitro reconstitution and LC-MS analysis

100 µL *in vitro* reactions with purified P450s and P450 reductases contained 1-2 µL purified and concentrated CYP72A (2-32 pmoles P450), 1 µL purified His_6-tagged *Caa* CPR1 full (62 nmol cytochrome *c* reduced/min/µL) or 3 µL purified His_6-tagged *Caa* CPR2 full (0.074 nmol cytochrome *c* reduced/min/µL), 0.25 mM loganic acid or loganin, 6 µL 1 mM DLPC (dilauroyl phosphatidyl choline), 0.5 mM NADPH in 100 mM $NaPO_4$ (pH 7.5) (or an equivalent volume of $NaPO_4$ in control reactions). Reactions were then incubated for 60-120 min at 30°C, stopped with 150 µL [0.67% formic acid, 33% acetonitrile in water], and centrifuged at 15,000 x*g* for 8 min.

High resolution LC-MS of samples was performed using the Q-Exactive MS system in the Metabolomics Laboratory of Roy J. Carver Biotechnology Center, University of Illinois Urbana-Champaign. The Software Xcalibur 3.0.63 was used for data acquisition and analysis. The Dionex Ultimate 3000 series HPLC system used includes a degasser, an autosampler, a diode array detector (DAD), and a binary pump. The LC separation was performed on an XTerra RP18 5µm 4.6x150mm column with an in-line XTerra RP18 5µm 3.9x20mm guard column with mobile phase A (water with 0.1% formic acid) and mobile phase B (acetonitrile with 0.1% formic acid). The DAD was set at 240 nm. The flow rate was 1 mL/min. The isocratic separation was performed with 10 % B over a running time of 25 min. The autosampler was set to 10°C with the injection volume of 20 µL. Mass spectra were acquired under both positive electrospray

ionization (spray voltage: 3.5 kV) and negative electrospray ionization (spray voltage: - 2.5kV): auxiliary gas flow rate, 65; aux gas flow rate: 20; sweep gas flow rate, 4; capillary temp, 300°C; auxiliary gas heater temp, 500°C. The resolution was set to 70,000. The AGC target was $1x10^6$ with a maximum injection time of 200 ms. The scan range was *m/z* 50 to *m/z* 750.

2.4.6. Substrate-induced spectral shift assays

Concentrated CYPs were diluted to 0.60 μM in 100 mM $NaPO_4$ (pH 7.5) and their spectra obtained from 300-700 nm against a reference cuvette containing only buffer. Successive small aliquots of loganic acid, loganin or buffer were added to the sample and reference cuvettes to account for absorbances not associated with P450s. To correct for variations occurring from the removal and repositioning of cuvettes, the difference spectra were calculated by internally correcting each spectrum by subtracting the Abs 700 nm value from all other wavelengths. Then, the spectrum obtained from adding buffer was subtracted from each substrate-containing sample of the same wavelength. Binding curves from these spectra plotted the ΔAbs from peak (at 389 nm) to trough (at 419 nm) against substrate concentration. The binding constants are reported with standard error from the curve fitting regression conducted in OriginPro2019 to a one-site binding model with the same equation as the Michaelis-Menten equation.

2.4.7 Steady state kinetics

To estimate the rate of product formation, the rate of NADPH consumption was monitored spectroscopically for 100 μL reactions performed at 30°C in a 96-well plate as technical triplicates. Individual wells contained 50 mU full length *Caa* CPR1, 25 pmol CYP, 60 μM DLPC (final concentration of 0.6% v/v methanol), no substrate or 50.0 μM to 5.00 mM substrate in 100 mM $NaPO_4$ (pH 7.5). Reactions were initiated by the addition of NADPH (0.8 mM final

concentration), shaking for 15 s and monitoring at 340 nm every 15 s with shaking for 3 s between reads.

Initial rates were estimated via a linear regression of the absorbance versus time data, converted from a.u. min^{-1} to μM NADPH min^{-1} using an extinction coefficient of 6,220 M^{-1} cm^{-1}, and the no substrate rate subtracted from substrate containing assays to correct for the background rate of NADPH turnover by CPR. These corrected rates were plotted against substrate concentration in OriginPro2019 and fit to the Michaelis-Menten equation to estimate kinetic parameters using nonlinear curve fitting.

2.4.8. Homology modeling

Three-dimensional protein models based on single-template homology were constructed utilizing MOE (Chemical Computing Group Inc.) according to published procedures (26). The template structure was determined via a multiple sequence alignment in MOE with all substrate-free CYPs in the PDB as of January 2018. CYP72A1, CYP72A564 and CYP72A565 were modeled on substrate-free CYP2D6 (PDB 2F9Q) (36); CYP72A730 was modeled on substrate-free human CYP3A4 structure as a template (PDB 1TQN) (37). Loganic acid and loganin were docked into each CYP72A before a final minimization. The final models are appended (Appendices A, CYP72A1; B, CYP72A564; C, CYP72A565; and D, CYP72A730).

2.5. TABLES AND FIGURES

Table 2.1. Primers used for cloning and sequencing.

Gene-specific identities are underlined, restriction sites named in primers are gray-filled, His_6-tags are double underlined, ATG start codons are bolded.

Gene	Primer name	Sequence (5' to 3')	Tm
72A565	Caa SLS 37 KpnI 5'	ATTAGGCCGGTACCA**ATG**GAAATTCAAATGGACGTTCT	62°C
72A565	Caa SLS 37 SacII 3'	ATTAGGCCCCGCGGTCAACACTTGATCTTGCGGAAG	64°C
72A1	Cra SLS 4 KpnI 5'	CGGGGTACC**ATG**GAAATGGATATGGATACT	54°C
72A1	Cra SLS 4 SacI 3'	AAGCGAGCTCCTAGCTCTCAAGCTTCTTGTAG	42°C
72A564	Caa SLS 364 KpnI 5'	GGCCGGTACC**ATG**CTCAAAGAAGCCATGTCCAA	66ºC
72A564	Caa SLS 364 SacII 3'	GGCCCCGCGGTCAACATTTGAGCCTGCGGAA	62ºC
72A730	Caa SLS 364iso2 Kpn 5'	ATTAGGTACC**ATG**GAAATGGAGCTAATGTACAAGTCC	76ºC
72A730	Caa SLS 364iso2 SacII 3'	ATTACCGCGGTCAATATTTGAGTTTGCGGAAGATCAC	74ºC
CPR1	Caa CPR1full NcoI 5'	ATCC**ATG**GGCCACCACCACCACCACCACAGTTCAGATT CGGAGTTGATTAG	64°C
CPR1	Caa CPR1Δ48 NcoI 5'	ATCC**ATG**GGCCACCACCACCACCACCACAGATCGTCTG ATCGAAGTAAAGA	62°C
CPR2	Caa CPR2full NcoI 5'	ATCC**ATG**GGCCACCACCACCACCACCACCAATCGAGTT CGGTTAAGGTG	64°C
CPR2	Caa CPR2Δ68 NcoI 5'	ATCC**ATG**GGCCACCACCACCACCACCACAGATCGTCAG GGAAGTCGGG	64°C
CPR1	Caa CPR1 BamHI 3'	ATGGATCCTCACCAGACATCCCTAAGATAC	64°C
CPR2	Caa CPR2 SacI 3'	ATTAGAGCTCTCACCACACATCACGCAAATACC	68°C
72A565	Caa SLS 37 pCW NdeI 5'	ATTACAT**ATG**GAAATTCAAATGGACGTTCTGTAC	74ºC
72A565	Caa SLS 37 pCW His6 XbaI 3'	TAATTCTAGATCAGTGGTGGTGGTGGTGGTGACACTTG ATCTTGCGGAAGACCAC	72ºC
72A564	Caa SLS 364 pCW His6 XbaI 3'	TAATTCTAGATCAGTGGTGGTGGTGGTGGTGACATTTG AGCCTGCGGAAGATC	66ºC
72A730	Caa SLS 364iso2 pCW NdeI 5'	ATTACAT**ATG**GAAATGGAGCTAATGTACAAGTC	76ºC
72A730	Caa SLS 364iso2 pCW His6 XbaI 3'	TAATTCTAGATCAGTGGTGGTGGTGGTGGTGATATTTG AGTTTGCGGAAGATCACTTG	74ºC
	M13For	GTAAAACGACGGCCAGT	58°C
	M13Rev	AACAGCTATGACCATG	45°C
	pET28a 5'	GAATTGTGAGCGGATAACAATTCC	55°C
	pET28a 3'	TTATGCTAGTTATTGCTCAGCGG	55°C
	CaaCPR2 mid	GTCTCGGCAACAGGCAGTATGA	68°C
	72A565 mid1	CTTCACCGTGACCAAGAGATGTGG	74°C
	72A565 mid2	TTGAAGACTATGCTGCCTGC	60°C
	72A564 mid1	CTTCACCGTGACAAGGAGATGTGG	74°C

Table 2.2. Binding parameters derived from Type I binding spectra.

Parameters ± S.E. estimated from non-linear curve fit of spectral response against varying substrate concentrations (*N*=17)

	loganic acid		**loganin**	
CYP	K_s	ΔAbs_{max}	K_s	ΔAbs_{max}
	μM	*mAU*	*μM*	*mAU*
72A564	874 ± 57	39.5 ± 1.3	256 ± 19	46.6 ± 1.2
72A565	2,410 ± 370	43.9 ± 4.6	841 ± 70	51.0 ± 2.2

Table 2.3. Kinetic parameters estimated from steady-state kinetics.

Parameters ± S.E. estimated from a non-linear curve fit of corrected NADPH consumption rates against varying substrate concentrations (*N*=7) from combined technical triplicate assays

CYP	**Substrate**	K_M	V_{max}	k_{cat}	k_{cat}/K_M
		mM	$\mu M\ min^{-1}$	min^{-1}	$min^{-1}\ mM^{-1}$
72A564	loganic acid	2.67 ± 0.80	3.49 ± 0.53	13.9 ± 2.1	5.2 ± 1.7
	loganin	0.24 ± 0.02	2.997 ± 0.075	12.0 ± 0.3	49.5 ± 4.8
72A565	loganic acid	3.57 ± 0.42	4.86 ± 0.32	19.4 ± 1.3	5.5 ± 0.7
	loganin	0.72 ± 0.07	1.92 ± 0.13	15.3 ± 0.5	21.2 ± 2.1

Table 2.4. Docking parameters from homology models.

Results computed using MOE software of models docked with substrates and minimized to root mean squared ±0.01 kcal mol^{-1} $(Å^2)^{-1}$

	loganic acid		**loganin**	
CYP	ΔE^a	C-O Distance[b]	ΔE^a	C-O Distance[b]
	$kJ\ mol^{-1}$	*Å*	$kJ\ mol^{-1}$	*Å*
72A1v3	-140.	3.79	-102	3.69
72A564	-157	4.20	-90.3	5.55
72A565	-183	3.56	-100.	4.32
72A730	-164	3.64	-126	3.10

[a] Ligand-receptor interaction energy calculated in MOE software

[b] Distance from C10 of substrate to oxygen ligand of heme

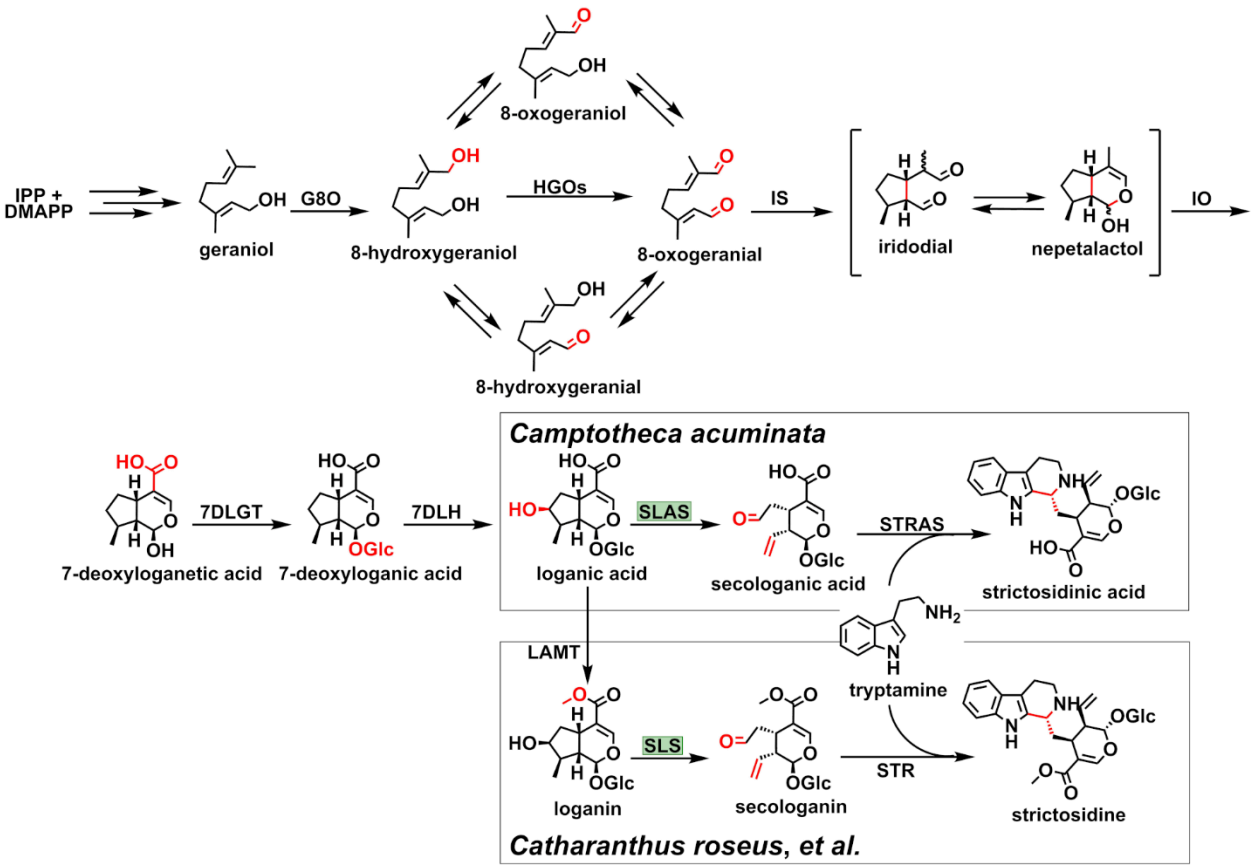


Figure 2.1. Proposed divergence of the TIA pathway between *Camptotheca* and *Catharanthus*.

Starting from geraniol by a series of transformations highlighted in red, the TIAs strictosidinic acid and strictosidine are synthesized in *Camptotheca* and *Catharanthus*, respectively. After 7DLH converts 7-deoxyloganic acid to loganic acid, the pathways in these two species diverge. The *Catharanthus* pathway uses LAMT to convert loganic acid into loganin and SLS to convert loganin into secologanin. The *Camptotheca* pathway bypasses LAMT and uses SLAS to metabolize loganic acid directly to secologanic acid.

IPP, isopentyl pyrophosphate; DMAPP, dimethylallyl pyrophosphate; G8O, geraniol 8-oxidase; HGOs, hydroxygeraniol oxidoreductases; IS, iridoid synthase; IO, iridoid oxidase; 7DLGT, 7-deoxyloganetic acid glucosyl transferase; Glc, glucose; 7DLH, 7-deoxyloganic acid hydroxylase; LAMT, loganic acid *O*-methyltransferase; SLAS, secologanic acid synthase; SLS, secologanin synthase; STRAS, strictosidinic acid synthase; STR, strictosidine synthase.

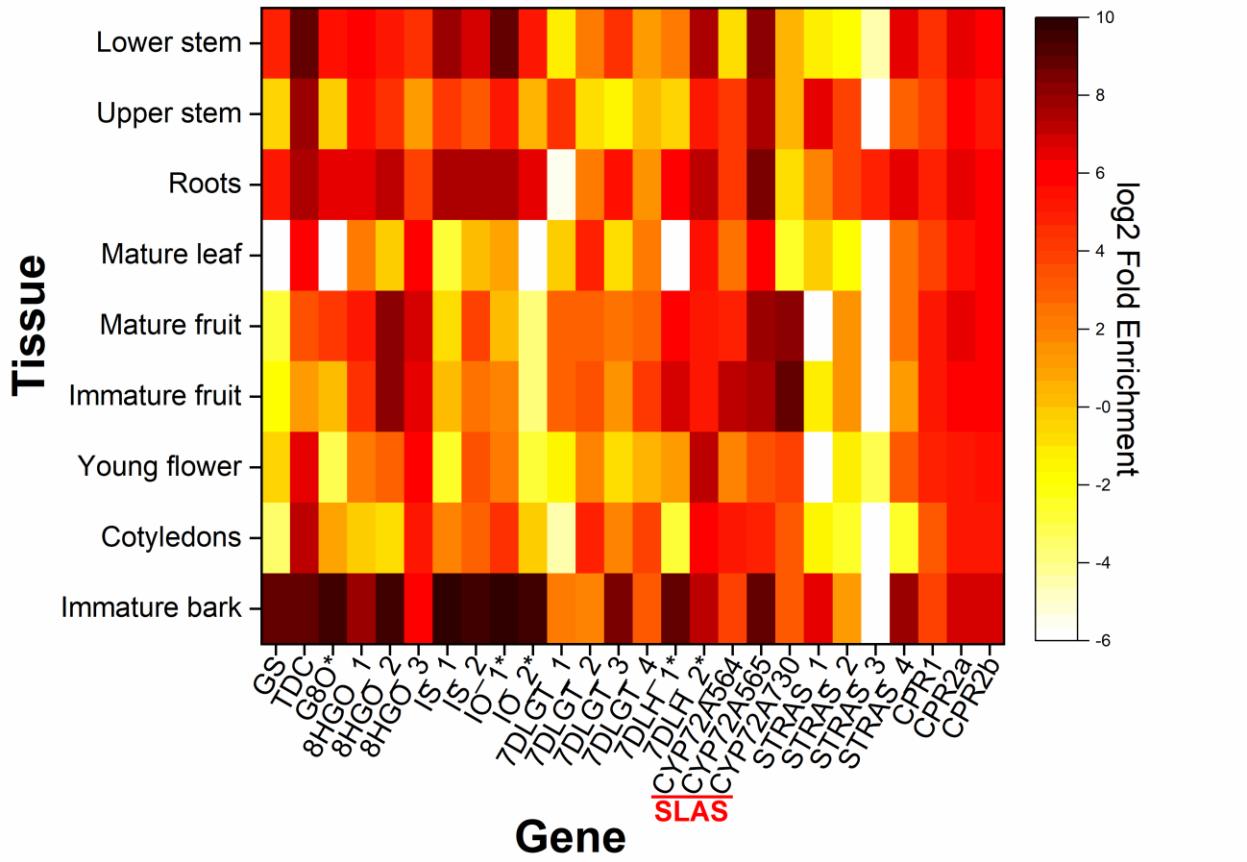


Figure 2.2. mRNA expression patterns of gene candidates for secoridoid biosynthesis from *Camptotheca acuminata*.

After hierarchical clustering[9] of RNA-seq data cross-referenced to the *Camptotheca acuminata* genome (22), gene candidates[10] were determined by identifying loci with increased expression in immature bark and immature fruit as well as comparison to the biosynthetic genes from *Catharanthus roseus*. The $\log_2$ fold enrichment is shown for these candidate genes. An asterisk (*) denotes a cytochrome P450.

GS, geraniol synthase; TDC, tryptophan decarboxylase; G8O, geraniol 8-oxidase; 8HGO, 8-hydroxygeraniol oxidoreductase; IS, iridoid synthase; IO, iridoid oxidase; 7DLGT, 7-deoxyloganetic acid glucosyl transferase; 7DLH, 7-deoxyloganic acid hydroxylase; SLAS, secologanic acid synthase; STRAS, strictosidinic acid synthase; CPR, cytochrome P450 reductase.

[9] Dr. Bernarda Calla-Zalles performed RNA-seq data processing.

[10] Allison J. Hollatz, Layna Henry, and Dr. Mary A. Schuler performed candidate gene identification.

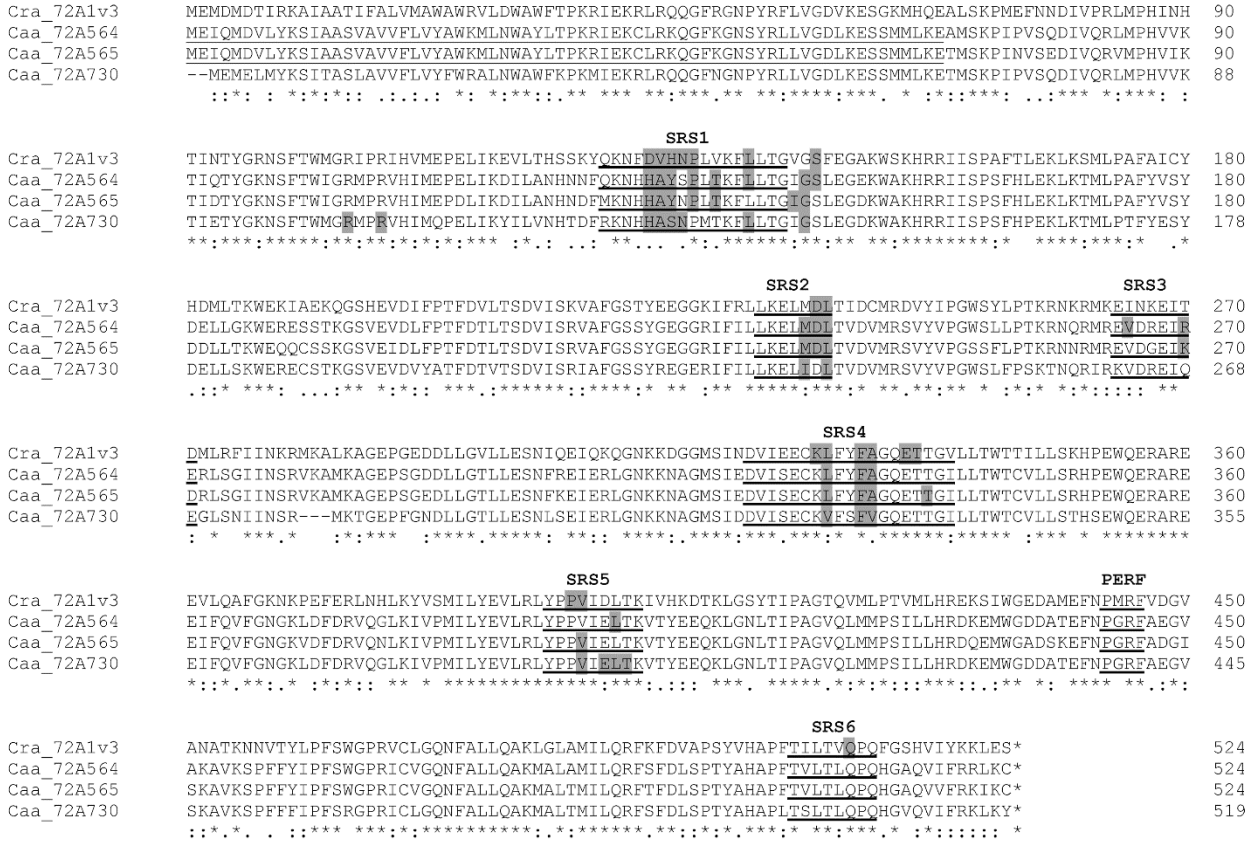


Figure 2.3. CYP72A multiple sequence alignment.
The signal anchor domain (SAD) fusion of CYP72A565 into CYP72A564 is underlined; SRS regions are underlined in bold; predicted substrate contacts within 4.5 Å of loganic acid/loganin are grey-filled.

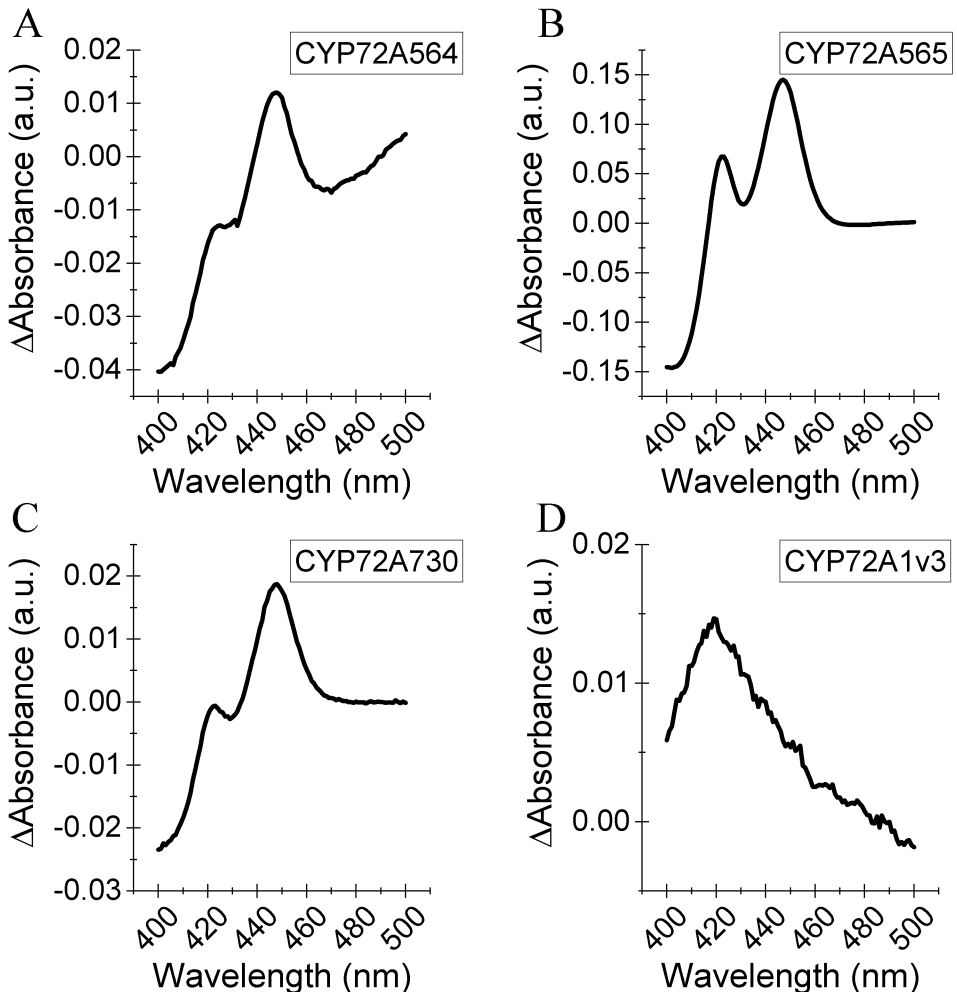
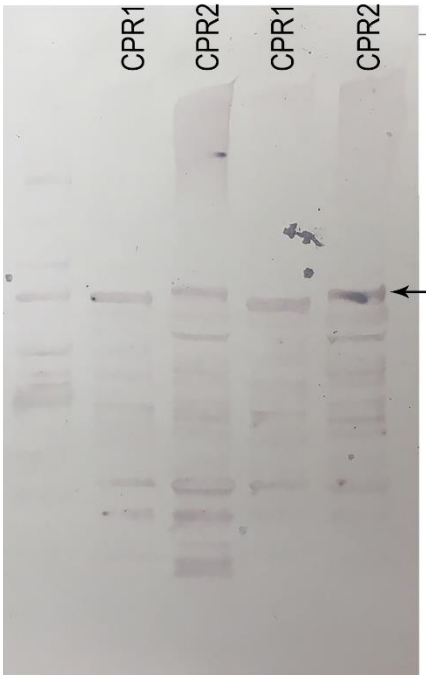


Figure 2.4. Reduced carbon monoxide difference analyses of purified His_6-tagged CYP72A564, CYP72A565, CYP72A730 and CYP72A1v3.

A



B

CPR Construct	Activity ($U\ \mu L^{-1}$)	[CPR] (μM)	Spec. Activity ($U\ mol^{-1}$)
CPR1 full	62,030	2.97	20,900
CPR2 full	74.11	2.50	29.6
CPR1 Δ2-48	102,600	3.77	27,200
CPR2 Δ2-68	63.69	1.84	34.6

Figure 2.5. Western blot analysis of purified His_6-tagged CPR1full and His_6-tagged CPR2full.

(A) Western analysis of 2 µL (lane 2) and 6 µL (lane 4) His_6-tagged CPR1full or 3 µL (lane 3) and 3.57 µL (lane 5) His_6-tagged CPR2full. Lanes 2 and 3 are loaded for volumes of purified CPR used in reconstitution assays, lanes 4 and 5 are loaded for equimolar concentrations of CPR. (B) Activities of full-length and truncated CPR proteins based on cytochome *c* reduction assays and extinction coefficients of oxidized CPR of 21.4 $mM^{-1}\ cm^{-1}$ at 454 nm (38).

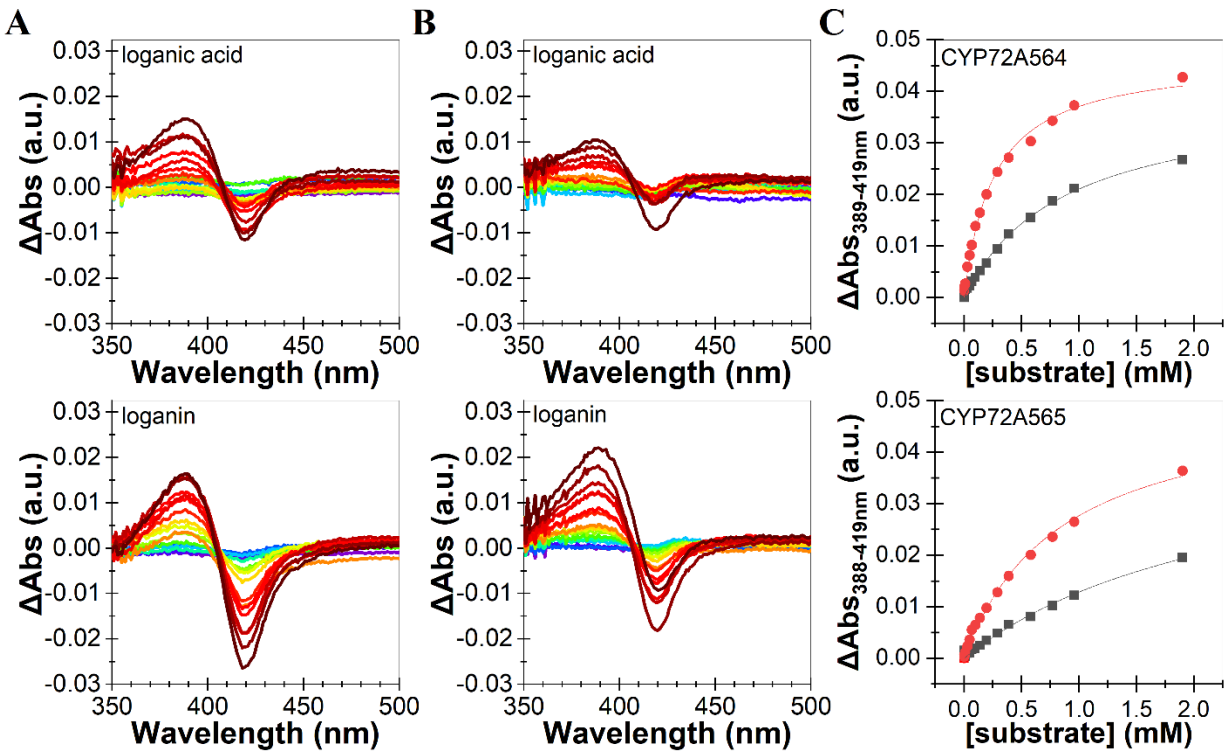


Figure 2.6. Type I binding spectra of SLAS candidates.
Substrate-induced Type I binding spectra for (A) CYP72A564, (B) CYP72A565 and from 998 nM (violet) to 1.90 mM (crimson) using loganic acid (top inset) or loganin (bottom inset). (C) Binding isotherms calculated as the difference of the valley (~419 nm) from the peak (~388 nm) are shown for loganic acid (■) and loganin (●).

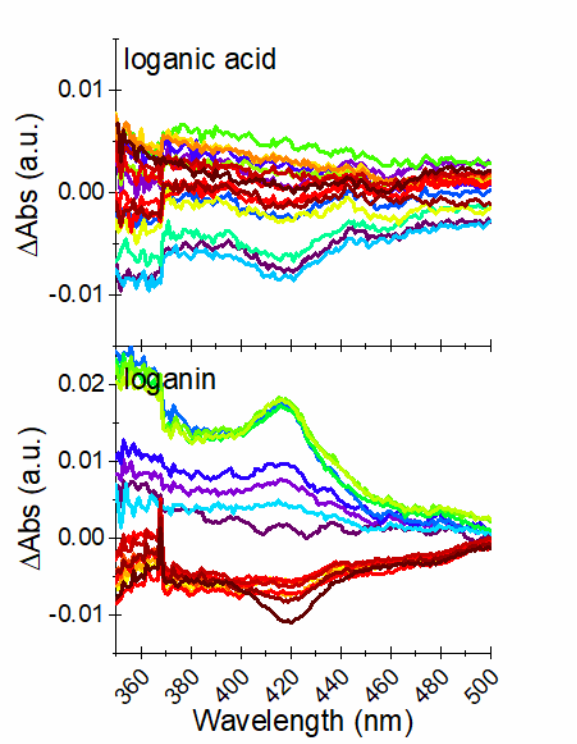


Figure 2.7. *Camptotheca* CYP72A730 loganic acid and loganin spectral titrations. Difference spectra are shown for loganic acid (top) and loganin (bottom) at concentrations ranging from 998 nM (violet) to 1.90 mM (crimson).

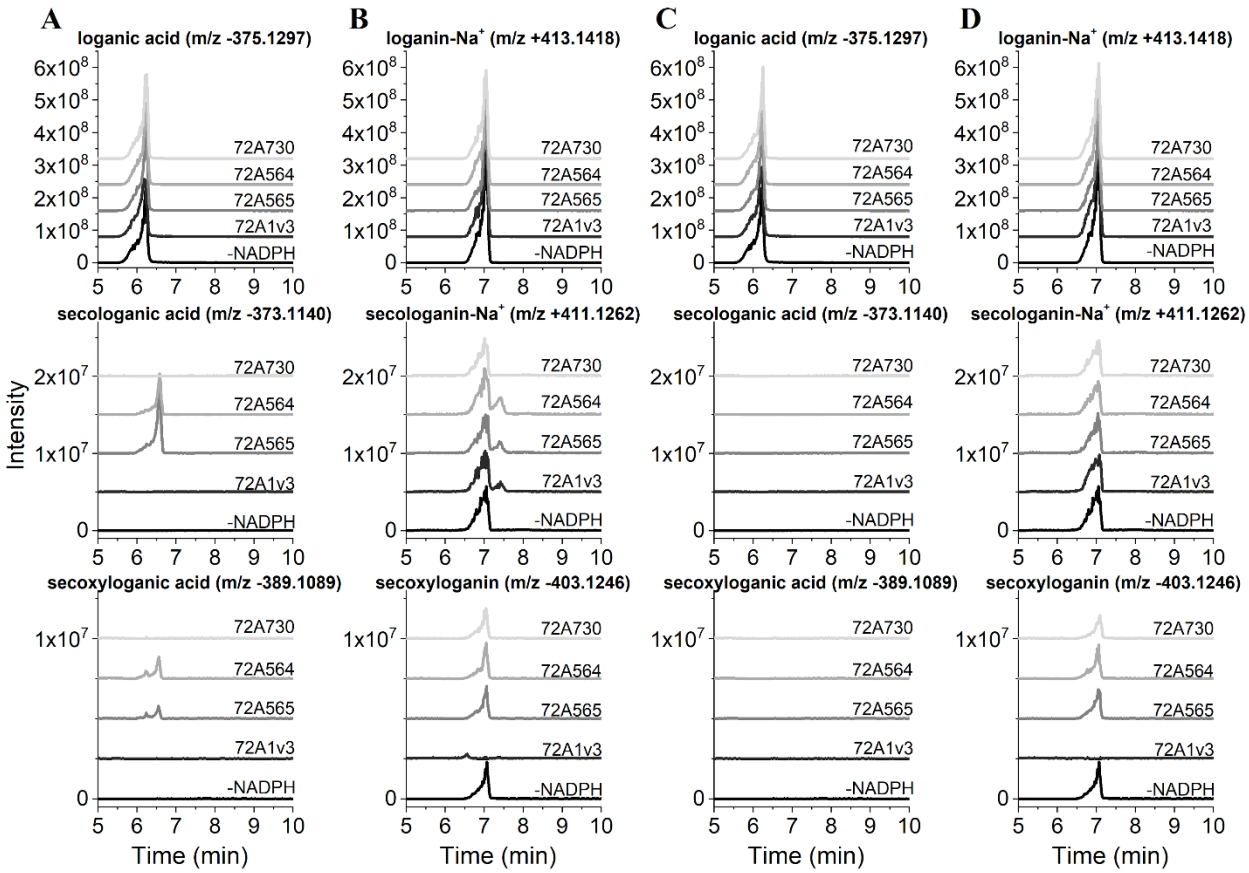


Figure 2.8. LC-MS analyses of *in vitro* assays with purified His$_6$-tagged CYP72A proteins reconstituted with His$_6$-tagged CPR proteins.

Reactions containing purified His$_6$-tagged CYP72A protein, full-length His$_6$-tagged *Caa* CPR1 protein (A, B) or full-length His$_6$-tagged *Caa* CPR2 protein (C, D), and 250 μM loganic acid (A,C) or loganin (B,D), were incubated at 30°C and analyzed by LC-MS as described in experimental procedures. Extracted ion chromatograms for loganic acid (*m/z* -375.1297), secologanic acid (*m/z* -373.1140), secoxyloganic acid (*m/z* -389.1089); loganin sodium salt (*m/z* +413.1418), secologanin sodium salt (*m/z* +411.1262), secoxyloganin (*m/z* -403.1246) are given with stacked chromatograms as marked.

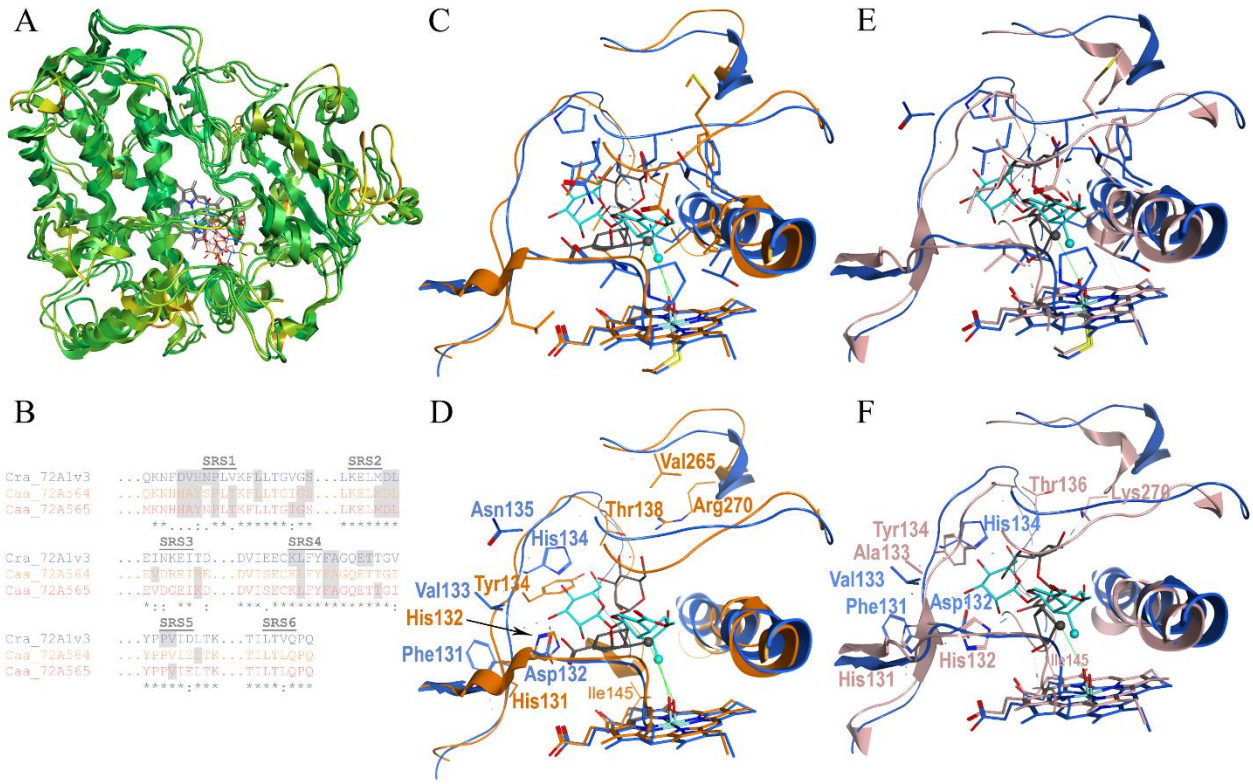


Figure 2.9. Molecular models of *Catharanthus* CYP72A1 and *Camptotheca* CYP72A564 and CYP72A565.

(A) Backbone overlays of *Catharanthus* CYP72A1 and *Camptotheca* CYP72A564 and CYP72A565 models are shown with the alpha-carbon RMSD amongst CYP72A1, CYP72A564 and CYP72A565 depicted from green (0.0 Å) to yellow (3.0 Å) to red (4.5 Å). (B) SRS regions in CYP72A proteins shown with predicted substrate contact residues (gray fill). (C) Identical versus (D) different side chain residues predicted within 4.5 Å of loganin (aqua) docked in *Catharanthus* CYP72A1 (blue) and loganic acid (gray) docked in *Camptotheca* CYP72A564 (orange). (E) Identical versus (F) different side chain residues predicted within 4.5 Å of loganin (aqua) docked in *Catharanthus* CYP72A1 (blue) and loganic acid (gray) docked in *Camptotheca* CYP72A565 (rose).

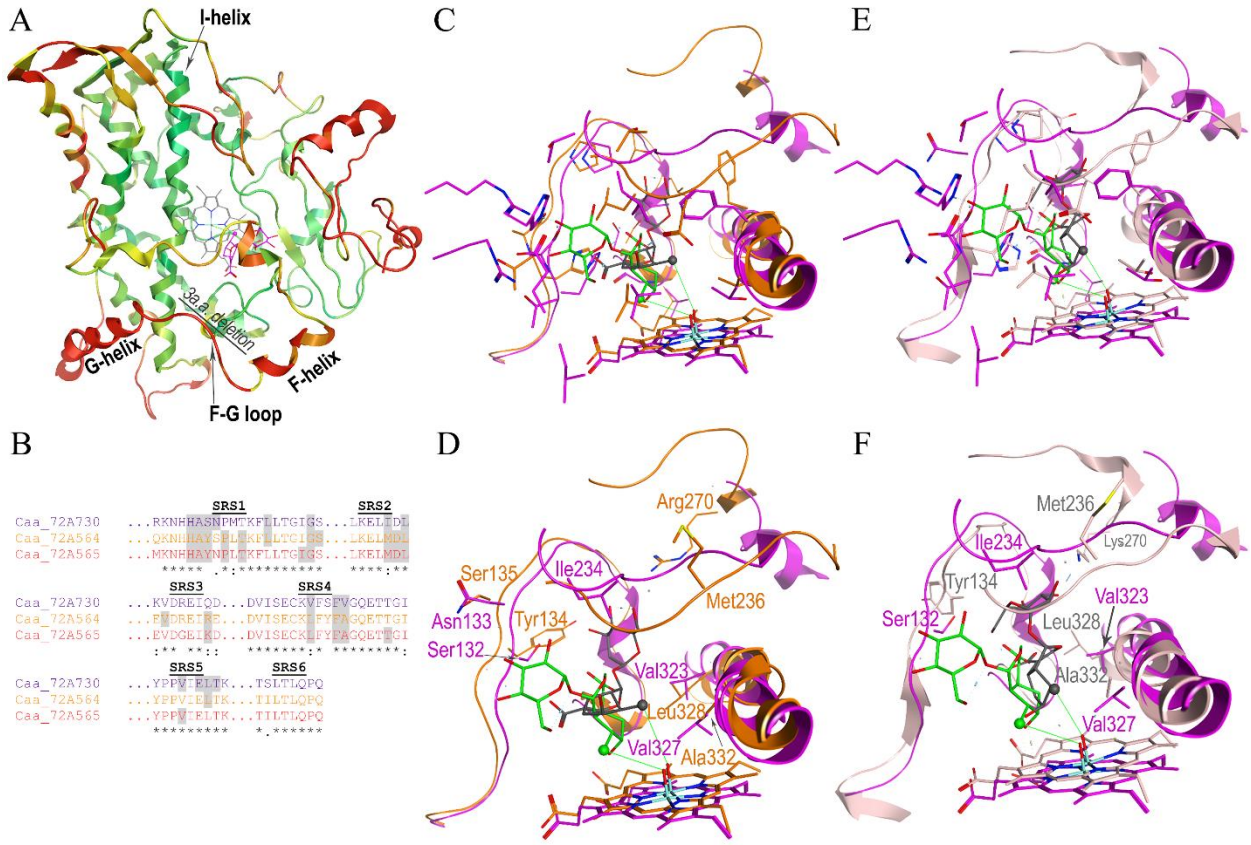


Figure 2.10. Molecular models of *Camptotheca* CYP72A564, CYP72A565 and CYP72A730. (A) Backbone overlays of *Camptotheca* CYP72A730, CYP72A564 and CYP72A565 models are shown with the RMSD variance of CYP72A564 and CYP72A565 from the CYP72A730 backbone depicted in green (0.0 Å), yellow (3.0 Å) and red (4.5 Å). (B) SRS regions in CYP72A proteins shown with predicted substrate contact residues (gray fill). (C) Identical versus (D) different side chain residues predicted within 4.5 Å of loganic acid (gray) docked in *Camptotheca* CYP72A564 (orange) versus loganic acid (aqua) docked in CYP72A730 (magenta). (E) Identical versus (F) different side chain residues predicted within 4.5 Å of loganic acid (gray) docked in *Camptotheca* CYP72A565 (rose) versus loganic acid (aqua) docked in CYP72A730 (magenta).

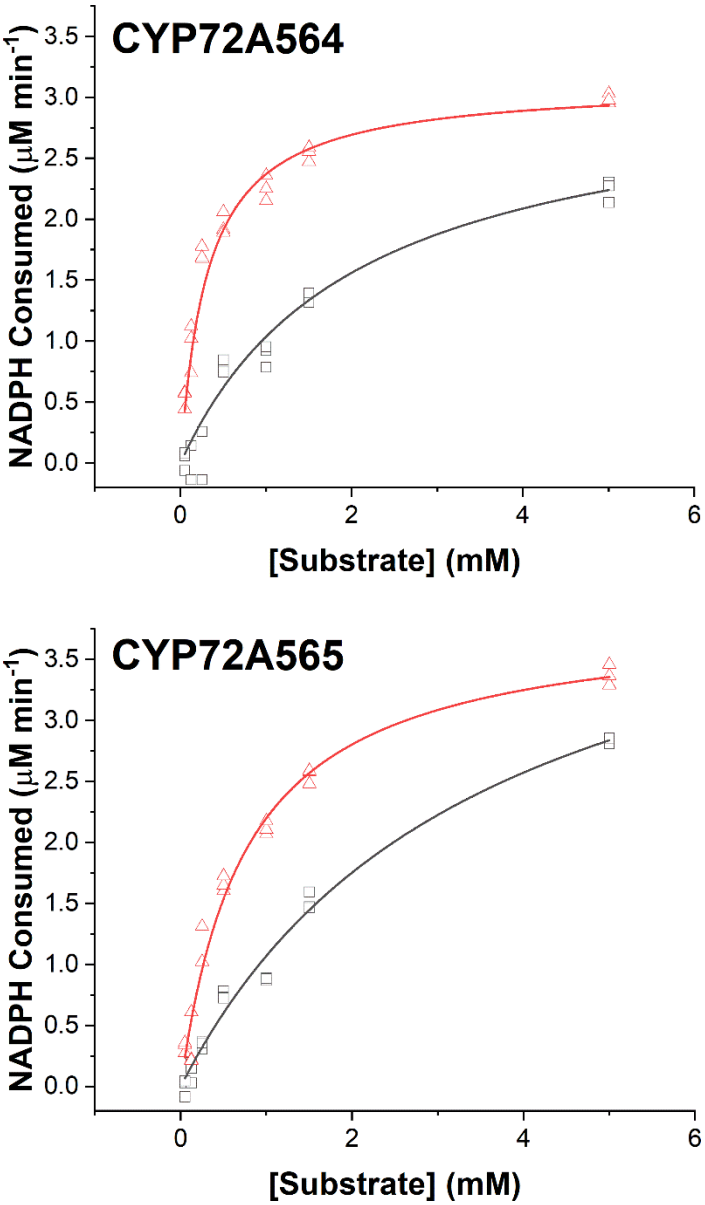


Figure 2.11. Nonlinear curve fitting of steady-state kinetics data to the Michaelis-Menten equation.

Corrected rates of NADPH consumption for loganic acid (□) and loganin (△) were fit to the Michaelis-Menten equation to estimate K_M and V_{max} for Camptotheca CYP72A564 (top) and CYP72A565 (bottom).

2.6. REFERENCES

1. Newman, D. J., and Cragg, G. M. (2016) Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* **79**, 629–661
2. Heijden, R. van der, Jacobs, D. I., Snoeijer, W., and Verpoorte, D. H. and R. (2004) The *Catharanthus* alkaloids: Pharmacognosy and biotechnology. *Curr. Med. Chem.* **11**, 607–628
3. Zhu, X., Zeng, X., Sun, C., and Chen, S. (2014) Biosynthetic pathway of terpenoid indole alkaloids in *Catharanthus roseus*. *Front. Med.* **8**, 285–293
4. Sirikantaramas, S., Asano, T., Sudo, H., and Saito, M. Y. and K. (2007) Camptothecin: Therapeutic potential and biotechnology. *Curr. Pharm. Biotechnol.* **8**, 196–202
5. Martino, E., Della Volpe, S., Terribile, E., Benetti, E., Sakaj, M., Centamore, A., Sala, A., and Collina, S. (2017) The long story of camptothecin: From traditional medicine to drugs. *Bioorg. Med. Chem. Lett.* **27**, 701–707
6. Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., Verpoorte, R., Oksman-Caldentey, K.-M., Martinoia, E., Bouwmeester, H., Goossens, A., Memelink, J., and Werck-Reichhart, D. (2014) The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606
7. Caputi, L., Franke, J., Farrow, S. C., Chung, K., Payne, R. M. E., Nguyen, T.-D., Dang, T.-T. T., Carqueijeiro, I. S. T., Koudounas, K., Bernonville, T. D. de, Ameyaw, B., Jones, D. M., Vieira, I. J. C., Courdavault, V., and O'Connor, S. E. (2018) Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science*. **360**, 1235–1239
8. Qu, Y., Safonova, O., and De Luca, V. (2019) Completion of the canonical pathway for assembly of anticancer drugs vincristine/vinblastine in *Catharanthus roseus*. *Plant J.* **97**, 257–266
9. Irmler, S., Schröder, G., St-Pierre, B., Crouch, N. P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J. (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: New enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797–804
10. Collu, G., Unver, N., Peltenburg-Looman, A. M. G., van der Heijden, R., Verpoorte, R., and Memelink, J. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* **508**, 215–220
11. Murata, J., Roepke, J., Gordon, H., and De Luca, V. (2008) The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell*. **20**, 524–542
12. Geu-Flores, F., Sherden, N. H., Courdavault, V., Burlat, V., Glenn, W. S., Wu, C., Nims, E., Cui, Y., and O'Connor, S. E. (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*. **492**, 138–142
13. Dugé de Bernonville, T., Foureau, E., Parage, C., Lanoue, A., Clastre, M., Londono, M. A., Oudin, A., Houillé, B., Papon, N., Besseau, S., Glévarec, G., Atehortùa, L., Giglioli-Guivarc'h, N., St-Pierre, B., De Luca, V., O'Connor, S. E., and Courdavault, V. (2015) Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced *de novo* assembly of a *Catharanthus roseus* transcriptome. *BMC Genomics*. **16**, 619

14. Kutchan, T. M., Bock, A., and Ditrrich, H. (1994) Heterologous expression of the plant proteins strictosidine synthase and berberine bridge enzyme in insect cell culture. *Phytochemistry*. **35**, 353–360
15. Gerasimenko, I., Sheludko, Y., Ma, X., and Stöckigt, J. (2002) Heterologous expression of a *Rauvolfia* cDNA encoding strictosidine glucosidase, a biosynthetic key to over 2000 monoterpenoid indole alkaloids. *Eur J Biochem*. **269**, 2204–2213
16. Ma, X., Koepke, J., Fritzsch, G., Diem, R., Kutchan, T. M., Michel, H., and Stöckigt, J. (2004) Crystallization and preliminary X-ray crystallographic analysis of strictosidine synthase from *Rauvolfia*: The first member of a novel enzyme family. *Biochim. Biophys. Acta BBA - Proteins Proteomics*. **1702**, 121–124
17. Góngora-Castillo, E., Childs, K. L., Fedewa, G., Hamilton, J. P., Liscombe, D. K., Magallanes-Lundback, M., Mandadi, K. K., Nims, E., Runguphan, W., Vaillancourt, B., Varbanova-Herde, M., DellaPenna, D., McKnight, T. D., O'Connor, S., and Buell, C. R. (2012) Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. *PLoS ONE*. **7**, e52506
18. Sadre, R., Magallanes-Lundback, M., Pradhan, S., Salim, V., Mesberg, A., Jones, A. D., and DellaPenna, D. (2016) Metabolite diversity in alkaloid biosynthesis: A multilane (diastereomer) highway for camptothecin synthesis in *Camptotheca acuminata*. *Plant Cell*. **28**, 1926–1944
19. Yang, Y., Li, W., Pang, J., Jiang, L., Qu, X., Pu, X., Zhang, G., and Luo, Y. (2019) Bifunctional cytochrome P450 enzymes involved in camptothecin biosynthesis. *ACS Chem. Biol.* **14**, 1091–1096
20. Sun, Y., Luo, H., Li, Y., Sun, C., Song, J., Niu, Y., Zhu, Y., Dong, L., Lv, A., Tramontano, E., and Chen, S. (2011) Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. *BMC Genomics*. **12**, 533
21. Qu, X., Pu, X., Chen, F., Yang, Y., Yang, L., Zhang, G., and Luo, Y. (2015) Molecular cloning, heterologous expression, and functional characterization of an NADPH-cytochrome P450 reductase gene from *Camptotheca acuminata*, a camptothecin-producing plant. *PLOS ONE*. **10**, e0135397
22. Zhao, D., Hamilton, J. P., Pham, G. M., Crisovan, E., Wiegert-Rininger, K., Vaillancourt, B., DellaPenna, D., and Buell, C. R. (2017) De novo genome assembly of *Camptotheca acuminata*, a natural source of the anti-cancer compound camptothecin. *GigaScience*. 10.1093/gigascience/gix065
23. Omura, T., and Sato, R. (1967) Isolation of cytochromes P-450 and P-420. in *Methods in Enzymology*, pp. 556–561, Oxidation and Phosphorylation, Academic Press, **10**, 556–561
24. Guengerich, F. P., Martin, M. V., Sohl, C. D., and Cheng, Q. (2009) Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. *Nat. Protoc.* **4**, 1245–1251
25. Jefcoate, C. R. (1978) Measurement of substrate and inhibitor binding to microsomal cytochrome P-450 by optical-difference spectroscopy. in *Methods in Enzymology* (Fleischer, S., and Packer, L. eds), pp. 258–279, Biomembranes - Part C: Biological Oxidations, Academic Press, **52**, 258–279
26. Rupasinghe, S., and Schuler, M. A. (2006) Homology modeling of plant cytochrome P450s. *Phytochem. Rev.* **5**, 473–505

27. Yamamoto, H., Katano, N., Ooi, A., and Inoue, K. (2000) Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. *Phytochemistry*. **53**, 7–12
28. Guengerich, F. P., and Yoshimoto, F. K. (2018) Formation and cleavage of C–C bonds by enzymatic oxidation–reduction reactions. *Chem. Rev.* **118**, 6573–6655
29. Jin, Z., Cong, Y., Zhu, S., Xing, R., Zhang, D., Yao, X., Wan, R., Wang, Y., and Yu, F. (2019) Two classes of cytochrome P450 reductase genes and their divergent functions in *Camptotheca acuminata* Decne. *Int. J. Biol. Macromol.* **138**, 1098–1108
30. Frances, O., Fatemi, F., Pompon, D., Guittet, E., Sizun, C., Pérez, J., Lescop, E., and Truan, G. (2015) A well-balanced preexisting equilibrium governs electron flux efficiency of a multidomain diflavin reductase. *Biophys. J.* **108**, 1527–1536
31. Brignac-Huber, L. M., Park, J. W., Reed, J. R., and Backes, W. L. (2016) Cytochrome P450 organization and function are modulated by endoplasmic reticulum phospholipid heterogeneity. *Drug Metab. Dispos.* **44**, 1859–1866
32. Wester, M. R., Johnson, E. F., Marques-Soares, C., Dijols, S., Dansette, P. M., Mansuy, D., and Stout, C. D. (2003) Structure of mammalian cytochrome P450 2C5 complexed with diclofenac at 2.1 Å resolution: Evidence for an induced fit model of substrate binding. *Biochemistry*. **42**, 9335–9345
33. Tietz, D. R., Podust, L. M., Sherman, D. H., and Pochapsky, T. C. (2017) Solution conformations and dynamics of substrate-bound cytochrome P450 MycG. *Biochemistry*. **56**, 2701–2714
34. Tietz, D. R., Colthart, A. M., Pochapsky, S. S., and Pochapsky, T. C. (2017) Substrate recognition by two different P450s: Evidence for conserved roles in a common fold. *Sci. Rep.* **7**, 1–9
35. Gegner, J. A., and Dahlquist, F. W. (1991) Signal transduction in bacteria: CheW forms a reversible complex with the protein kinase CheA. *Proc. Natl. Acad. Sci.* **88**, 750–754
36. Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Lewis, C. J., Tennant, M. G., Modi, S., Eggleston, D. S., Chenery, R. J., and Bridges, A. M. (2006) Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **281**, 7614–7622
37. Yano, J. K., Wester, M. R., Schoch, G. A., Griffin, K. J., Stout, C. D., and Johnson, E. F. (2004) The structure of human microsomal cytochrome P450 3A4 determined by X-ray crystallography to 2.05-Å resolution. *J. Biol. Chem.* **279**, 38091–38094
38. Manoj, K. M., Gade, S. K., and Mathew, L. (2010) Cytochrome P450 reductase: A harbinger of diffusible reduced oxygen species. *PLOS ONE*. **5**, e13272

CHAPTER 3: SINGLE MUTATIONS TOGGLE THE SELECTIVITY OF MULTIFUNCTIONAL *CAMPTOTHECA ACUMINATA* SECOLOGANIC ACID SYNTHASES (CYP72As)[1]

Terpene indole alkaloids (TIAs) are plant-derived specialized metabolites with widespread use in medicine. Species-specific pathways derive TIAs from a common intermediate, strictosidine, produced by coupling secologanin to tryptamine and, sometimes, from strictosidinic acid produced from secologanic acid and tryptamine. The penultimate reaction in this pathway is catalyzed by secologanin synthase (SLS) or secologanic acid synthase (SLAS) according to whether plants produce secologanin from loganin or secologanic acid from loganic acid. Combining molecular modeling, ancestral sequence reconstruction (ASR), and biochemical methodologies, I have identified key residues that toggle SLS and SLAS selectivity in two CYP72A (cytochrome P450) subfamily enzymes from *Camptotheca acuminata*. The positions of foremost importance are in substrate recognition sequence 1 (SRS1) where mutations to either of two adjacent histidine residues switched selectivity: His131Phe selects for and increases secologanin production; His132Asp selects for secologanic acid production. A third change in SRS3 in the predicted substrate entry channel (Arg/Lys270Thr) decreased enzyme activity towards either substrate; a fourth change in SRS4 at the start of the I-helix (Ser324Glu) severely decreased both activities. Surveying the ancestral and extant sequences, I propose that the *Camptotheca* SLASs have maintained a broader activity likely found in a common Asterid ancestor even as the *Camptotheca* lineage lost its ability to produce loganin. The Campanulid and Lamiid lineages, however, specialized to produce secologanin by acquiring mutations in

[1] Chapter adapted from Miller, J.C., and Schuler, M.A. (2022) Single mutations toggle selectivity of multifunctional *Camptotheca* secologanic acid synthases (CYP72As). *In review*.

SRS1. The identification of the residues essential for the broad substrate scope of the SLASs presents opportunities for more tailored heterologous production of TIAs.

3.1. INTRODUCTION

Terpene indole alkaloids (TIAs) encompass thousands of specialized metabolites produced by a wide array of Asterids. These compounds, acting as defensive agents against various biotic stresses (herbivory, fungal or bacterial infection, *etc.*), enable these species to adapt to changing environments. The diversity of TIAs produced, the many species that produce them, the geographic distribution of these plants, and the biological activity of the compounds underlie their use as medicines across cultures and throughout the world.

Often complex in their fused ring systems, position of oxidations, and stereochemistry (Fig. 3.1), TIAs are not readily synthesized at scales conducive to industrial production (1, 2). The molecules are instead isolated from producing species with variable success owing to their low concentrations in plant materials, environmentally influenced production, and difficulties in purifying the desired compound from the complex matrices in which they accumulate (bark, leaves, fruits, *etc*.).

In spite of the immense structural diversity of TIAs among medicinal plants, a relatively common biosynthetic pathway exists wherein a secoiridoid is coupled to tryptamine to produce a common strictosidinic acid core (Fig. 3.1). This shared biosynthetic system provides an avenue for heterologous production of TIAs to alleviate the limited productivity of harvesting from native producers. The elucidation of the biosynthetic genes from *Catharanthus roseus* (3–14), for example, inspired the production of strictosidine in *Saccharomyces cerevisiae* (15). Though able to isolate strictosidine, analyses of these cultures found large amounts of loganin—an indication that the cytochrome P450 (CYP) secologanin synthase (SLS) was limiting TIA production.

Though SLSs have been characterized from multiple species (*Lonicera japonica* (16), *Catharanthus roseus* (4, 5), *Nothapodytes nimmoniana* (17), and *Ophiorrhiza pumila* (18)) with multiple isoforms (5, 15, 18), the difficulties of working with membrane-bound CYPs have complicated biochemical characterizations aimed at optimizing the enzymatic activities of any SLS. Moreover, the lack of studies reporting SLS activity outside of microbial cultures (4, 15), lysates (4, 5), or microsomal fractions (4, 16, 17) suggests that the best-studied *Catharanthus* SLS isoforms suffer from poor expression and/or protein instability in heterologous hosts. The recent identification (19, 20) of two secologanic acid synthases (SLASs) from *Camptotheca acuminata* that are also active as SLSs provides a set of secologanin-producing enzymes that may serve as alternatives to the problematic *Catharanthus* SLSs. That these *Camptotheca* CYPs are amenable to heterologous expression, purification, and biochemical assessment (20) suggests that they are more stable in heterologous systems than any of the *Catharanthus* SLS variants.

The ability of these versatile *Camptotheca* CYPs to produce secologanic acid from loganic acid as well as secologanin from loganin, however, presents a key hurdle for the efficient production of strictosidine in heterologous species. I therefore undertook the identification of residues enabling the differentiation of loganic acid and loganin and, thereby, SLAS *versus* SLS activity. For this venture, homology modeling and ancestral sequence reconstruction (ASR) were combined to identify sites likely involved in substrate recognition by the various SLSs/SLASs. Site-directed mutagenesis of these sites in the *Camptotheca* SLASs and subsequent *in vitro* assays confirmed that mutation at any of four sites perturbed substrate binding and/or turnover for loganic acid and loganin. Importantly for future attempts aimed at improving secologanin production, one key mutation in either of these *Camptotheca* SLASs improved loganin binding and turnover while abolishing loganic acid binding and turnover.

3.2. RESULTS

3.2.1. Ancestral sequence reconstruction identifies four loci in substrate recognition sequences

Our previous work with homology models (20) identified the conserved His132 of *Camptotheca* CYP72A564 and CYP72A565 as likely important for these enzymes' ability to turnover loganic acid. Cognizant that these models are but snapshots of dynamic proteins, I determined to apply another methodology to identify residues effecting substrate differentiation between loganic acid and loganin.

The publication of multiple medicinal plant genomes over the past decade presented sufficient numbers of sequences for ASR—a method with several advantages for these endeavors: First, by including putative SLS proteins from species besides *Catharanthus,* ASR had potential to remove the focus from variations particular to just SLS sequences from *Catharanthus*. Second, the relative speed and computational expenses of ASR allowed the evaluation of more sequences than possible with molecular dynamics simulations. Third, unlike my previously developed static models (20), ASR had potential to identify residues important for substrate differentiation not immediately adjacent to the heme cofactor.

Sequences annotated as SLS, 7-deoxyloganic acid hydroxylase (7DLH), SLS-like, or 7DLH-like within GenBank from species known to produce secoiridoids (Table 3.1) were used to construct the phylogenetic tree (Fig. 3.2A). I used only sequences from species known to produce secoiridoids knowing that SLSs and 7DLHs are only a small subset of the enzymes within the CYP72A subfamily (21). For several reasons, the inclusion of 7DLHs was important to determine a common SLS/SLAS ancestor. First, these SLSs, SLASs, and 7DLHs are all members of the CYP72A subfamily, have upwards of 80% homology, and utilize highly similar substrates. Second, the *Camptotheca* SLASs have demonstrated all three activities—7DLH, SLS,

and SLAS (19)—presenting the possibility that they are more similar to extant 7DLHs than to SLSs. Third, only six of the sequences annotated as SLS (4, 5, 16–18) or 7DLH (7) in GenBank have been characterized biochemically.

After a multiple-sequence alignment that also included *Arabidopsis* CYP734A1 (a member of the closest subfamily to CYP72A (21)) to root the tree, I constructed a maximum likelihood bootstrap consensus tree (Fig. 3.2A) that was predominantly in agreement with the established phylogenetic relationships amongst the included species (22). The SLS/SLAS lineage separates Cornales (*Camptotheca*, Caa) from the Campanulids (Dipsacales: *Lonicera*, Lja; Aquifoliales II: *Nothapodytes*, Nni) and the Lamiids (Gentianales: *Cinchona*, Cca; *Swertia*, Sja; *Ophiorrhiza*, Opu; *Amsonia*, Ahu; *Tabernaemontana*, Tel; *Catharanthus*, Cra; *Rauvolfia*, Rsa; *Vinca*, Vmi). *Swertia* (Gentianaceae), however, does intercalate between two constituents of Rubiaceae (*Cinchona* and *Ophiorrhiza*)—albeit without a strong bootstrap value (61, Fig. 3.2A). The putative 7DLH from *Olea* (Oeu 7DLH) did not group with either the SLS or 7DLH lineage in spite the large number of Lamiids represented. Several other CYP72As from *Olea* with considerable homology to *Catharanthus* SLS are known to metabolize secoiridoids (23), supporting the possibility that this 7DLH-like gene may be among the many CYP72As that mediate other reactions. Notwithstanding these caveats, I retained this phylogenetic tree for the prediction of a common SLS/SLAS ancestor because they did not affect the common ancestor of *Camptotheca* and the core Asterids.

Focusing on the SLS/SLAS ancestor (node 34, Fig. 3.3, 3.4), I began by comparing the predicted substrate recognition sequences (SRSs) of this ancestral protein to those of the *Camptotheca* SLASs and *Catharanthus* SLS (Fig. 3.2B, 3.5; Appendix E). In addition to the previously implicated His132 (20) (numbered according to the *Camptotheca* SLASs), a

neighboring His131 in SRS1, Arg/Lys270 in SRS3, and Ser324 in SRS4 appeared as sites of change between the SLS/SLAS ancestor and extant sequences. I then constructed a homology model (Fig. 3.2, C and D, 3.6) of the putative SLS/SLAS ancestor with loganic acid docked into the active site in order to visualize these sites and infer their importance.

With my previous predictions (20) suggesting that His132 in the *Camptotheca* SLASs interacts with the carboxylic acid moiety of loganic acid and that Asp132 in nearly all SLSs (Appendix E) prevents favorable interactions, I predicted that a His132Asp mutation would logically prevent loganic acid binding and turnover completely. Facing outward from the predicted active site in the extant *Camptotheca* SLASs' and *Catharanthus* SLS's homology models but a contact in the SLS/SLAS ancestor's homology model, His131 (Phe in all the SLSs) arose as another residue of probable importance. I predicted that disrupting a possible hydrogen bond donor to and electrostatic contact with loganic acid would negatively affect binding and turnover of loganic acid without affecting these enzymes' use of loganin.

The reasons why the models for the extant SLASs and the predicted ancestral sequence differ so drastically in their placement of these adjacent His residues remains unclear. The SLS/SLAS ancestor has two fewer amino acids, but a multiple sequence alignment (Fig. 3.5, Appendix E) places the subsequent gap in the signal anchor domain—a portion of the protein not included in the homology model. The template crystal structure, sequence alignment parameters, force fields, and docking methodologies used to generate the three homology models were all identical. With the exception of SRS3 along the G-helix (Fig. 3.2C), the extant *Camptotheca* SLASs' and the SLS/SLAS ancestor's homology models have little backbone variation as measured by RMSD. Lacking a definitive model for differentiating whether His131 or His132 is of greater importance, I constructed single and double mutants at both of these sites. I anticipated

that mutations at one of the sites would produce a strong effect, mutations at the other would have a weak effect (if any), and mutation at both sites would produce the same effect as that of the stronger of the two single mutants.

ASR identified two other sites not highlighted in my previous homology models. The first of these (Arg/Lys270) in the G-helix, though distant from the site of oxidation within the enzyme, points into the substrate entry channel where a basic, cationic residue might facilitate the entry of loganic acid into the active site. I predicted that mutations here might produce only modest effects on loganic acid binding and turnover as favorable interactions lying in the active site would still permit binding.

The second of these (Ser324) is the lone residue within the SRSs at which the SLS/SLAS ancestor differed from both extant *Camptotheca* SLASs; the ancestral sequence has Glu at this position like *Catharanthus* and all other included SLSs (Appendix E). Relevant to this position, solution NMR experiments on two bacterial CYPs (CYP101A1 (24) and MycG (25)) have demonstrated that the end of the I-helix at which Ser324 sits in the *Camptotheca* SLASs serves as a hinge allowing the protein to open and close for substrate binding. The homology model of the SLS/SLAS ancestor includes more H-bonding interactions around its Glu than the Ser of CYP72A564 or CYP72565 in their models (Fig. 3.7), perhaps hindering the hinging required to open and close the active site. Though I did not rule out effects on binding, I anticipated that mutating Ser324 to the much larger Glu would slow substrate entry and product exit thereby slowing the turnover rate.

3.2.2. Mutation at these loci perturbs Type I binding, steady-state kinetics

As indicated by Type I binding spectra (Fig. 3.8, Table 3.3), each of these five mutations affected the binding of loganic acid and loganin to CYP72A564 and CYP72A565. Consistent

with predictions from the ASR and homology modeling, the His132Asp mutant eliminated loganic acid-induced spectral shifts suggesting that the substrate does not bind near the heme iron as required for catalysis. This same mutation also reduced the affinity of these SLASs for loganin as witnessed by the 3.8- to 4.7-fold increase in K_s for CYP72A565 and CYP72A564, respectively (Table 3.3). These perturbations in substrate binding were reflected in a lack of loganic acid turnover and increased K_M for loganin for both enzymes (Tables 3.6 and 3.7). The His132Asp mutant did not greatly affect the k_{cat} of either CYP resulting in a greater than 5-fold reduction in k_{cat}/K_M for both proteins.

Mutating the nearby His131 residue to Phe also eliminated loganic acid-induced spectral shifts (Fig. 3.8). Unlike the His132 mutants, the His131Phe mutation increased the enzymes' affinity for loganin: a 4.7-fold smaller K_s for CYP72A564 and 3.0-fold smaller K_s for CYP72A565 (Table 3.3). Once more, the perturbations in Type I binding corresponded with changes in steady state kinetic parameters. Although CYP72A565 His131Phe demonstrated a slightly lower K_M for loganin, the K_M of the same mutant in CYP72A564 was 4.6-fold lower (Tables 3.6 and 3.7). Small but appreciably slower rate constants yielded mutant enzymes with larger specificity constants compared to the wild type with a 3-fold increase for CYP72A564 and a 25% increase for CYP72A565. Neither of the mutants appeared to metabolize loganic acid.

Neither the CYP72A564 nor the CYP72A565 double mutants (His131Phe, His132Asp) demonstrated loganic acid-induced Type I spectra (Fig. 3.8, Table 3.3), nor was there evidence of turnover (Tables 3.6 and 3.7). When combining the increased binding and turnover of loganin in the His131Phe mutants with the immense loss of these for the His132Asp mutants, the binding and kinetic parameters of the double mutants were roughly the average of those observed with the individual mutations and comparable to those of the wild type enzymes. In CYP72A564 for

which the effect of His131Phe was larger, neither the K_s nor the K_M in the double mutant differed greatly from the wild type (Tables 3.3 and 3.6). In CYP72A565 for which the effect of His132Asp was larger, the double mutant's K_M was more than twice as large as that of the wild type (Table 3.7). The specificity constant was lower in the double mutants for both enzymes than for their respective wild types.

The Arg/Lys270Thr mutations did not significantly perturb the loganic acid-induced spectral changes for either CYP (Fig. 3.8, Table 3.3). Although CYP72A564 Arg270Thr demonstrated comparable steady state kinetic parameters for loganic acid compared to its wild type (Table 3.6), the corresponding CYP72A565 mutant did not show evidence of loganic acid turnover (Table 3.7). The Arg/Lys270Thr mutation slightly decreased affinity for loganin (a 1.4-fold increase in K_s, Table 3.3) for both enzymes corresponding with larger Michaelis constants and smaller specificity constants (Tables 3.6 and 3.7).

Ser324Glu either reduced affinity for loganic acid (a 2.9-fold increase in K_s for CYP72A564) or eliminated Type I binding (for CYP72A565). The effects of this mutation on loganin binding also differed between the SLASs: CYP72A564 saw a 5.1-fold increase in K_s whereas that of CYP72A565 was unaffected. The effects on steady state kinetics parameters were more uniform—no evidence of loganic acid turnover, larger Michaelis constants, rate constants comparable to the wild type, and smaller specificity constants (Tables 3.6 and 3.7).

3.2.3. In vitro reconstitution confirms the effects of these mutations on secoiridoid production

To evaluate how these mutations affect the production of secologanic acid and secologanin, I reconstituted the CYPs with full-length *Camptotheca* CPR1 and compared the amount of each secoiridoid produced by LC-MS analysis of these *in vitro* reactions (Fig. 3.9 and 3.10). Using the specificity constants from the steady state experiments, I predicted that Arg/Lys270Thr mutants

would produce secologanic acid at lower levels, that the Ser324Glu mutants would produce lesser amounts still, and that all other mutants (His131Phe; His132Asp; and His131Phe, His132Asp) would not produce secologanic acid. For assays using loganin, I predicted both His131Phe mutants as the greatest producers of secologanin followed by the wild type enzymes, His132Asp and Ser324Glu as the lowest producers of secologanin, and the double mutants as well as Arg/Lys270Thr as lying somewhere in between.

I surprisingly observed that the His132Asp mutant from CYP72A564 produced about the same amount of secologanic acid as its wild type protein and that the same mutant from CYP72A565 produced just above half the amount of secologanic acid compared to its wild type enzyme (Fig. 3.9A). Each mutant containing His131Phe failed to produce secologanic acid above the limit of detection in accordance with their lack of loganic acid-induced Type I spectra and no increase in NADPH consumption in the presence of this iridoid. For CYP72A564, Arg270Thr only slightly affected secologanic acid production whereas, for CYP72A565, Lys270Thr substantially decreased secologanic acid production to only one quarter that of the wild type. For both enzymes, Ser324Glu lowered activity (as predicted) to approximately 40% the activity of the wild type.

In contrast, the pattern of secologanin production was the same for both CYP72A564 and CYP72A565 and roughly followed the pattern predicted by the Type I spectra and specificity constants: His131Phe > wild type $\geq$ His131Phe, His132Asp = Arg/Lys270Thr > Ser324Glu $\geq$ His132Asp = no P450 (Fig. 3.9B). For CYP72A564, the His131Phe mutant produced almost twice as much secologanin as the wild type with the double mutant and Arg270Thr showing no significant change in activity. The His132Asp and Ser324Glu mutations drastically reduced secologanin production. The His131Phe mutation showed a weaker increase in secologanin

production with CYP72A565, but the double mutant and Lys270Thr mutation both reduced secologanin production by over half. As with CYP72A564, mutating His132Asp or Ser324Glu in CYP72A565 effectively eliminated secologanin production.

To summarize the effects of mutagenesis on both these P450s, none of the mutations investigated improved secologanic acid production whereas the His131Phe mutation improved secologanin production for both CYP72A564 and CYP72A565. The adjacent His131 and His132, when mutated independently, toggled selectivity for secologanic acid production from loganic acid (His132Asp) and secologanin production from loganin (His131Phe). Arg/Lys270Thr changes reduced secoiridoid production, though not as severely as other mutations. For both CYP72A564 and CYP72A565, Ser324Glu was consistently deleterious to secoiridoid production. Between the His131Phe and His132Asp mutations that toggled selectivity when introduced singularly, double mutants of these *Camptotheca* SLASs made to resemble all CYP72As annotated as SLS or SLS-like converted loganin to secologanin but not loganic acid to secologanic acid like the His131Phe mutants.

3.3. DISCUSSION

The ability of *Camptotheca* CYP72A564 and CYP72A565 to both produce and metabolize loganic acid as well as transform loganin into secologanin is unique amongst known 7DLHs (3, 7, 26) and SLSs (4, 16–18, 20). By combining ASR with homology modeling, I first implicated four sites as contributing to the multifunctionality of these *Camptotheca* SLASs and then assayed the importance of the identified residues by conducting *in vitro* assays with site-directed mutants. I found that the adjacent residues His131 and His132 in SRS1 are imperative for setting selectivity in these enzymes: His131Phe selects for secologanin production from loganin; His132Asp selects for secologanic acid production from loganic acid. I also found that the

effects of modifying Arg/Lys270 along the substrate entry channel in SRS3 moderately reduce the SLAS and SLS activities of these P450s. The effects of modifying Ser324 at the start of the I-helix in SRS4 severely reduces both activities.

That the His132Asp mutant did not undermine secologanic acid production is puzzling given the lack of substrate-induced Type I spectra or increased NADPH consumption. A lack of spin-shift and the resulting Type I spectra are not unknown for compounds that bind tightly to P450s (27). The proximity of His132 to the docked loganic acid in the CYP72A564 and CYP72A565 homology models (20) (Appendices B and C) suggests an adjacent positioning of Asp132 and the carboxylic acid moiety of the substrate. Coordination of a cation (*e.g.*, sodium or potassium) by the protein and substrate carboxylic acids would transform unfavorable Coulombic interactions into a stable binding interaction. Touching the rates of NADPH consumption near background, the lack of spin shift indicated by Type I spectra are rather explanatory. The shift from low-spin (S = 1/2) to high-spin (S = 5/2) heme induced by removal of water as the sixth ligand to iron increases the iron's reduction potential (28). This increase in redox potential brings it nearer that of NADPH and CPR thus improving electron transfer to the heme (29, 30). Without this spin shift (*i.e.*, as was the case for the His132Asp mutants in the presence of loganic acid), reduction of the iron is greatly slowed. In this case, the rate of electron transfer to the P450 for turnover was sufficiently slow to be masked by the background rate of NADPH consumption by CPR due to uncoupling.

The inability of these same His132Asp mutants to produce much secologanin further suggests that charges are not the only important biochemical characteristic important to loganic acid and loganin binding. Besides the transition in charge represented in the His132Asp mutation, it causes the loss of a H-bond donor, likely the cause of disrupting loganin binding and

turnover. The reverse selectivity for loganin observed for His131Phe likely involves perturbed interactions with the carboxylate of loganic acid, too. Replacing a likely cationic, H-bond donor (His) with a similarly sized hydrophobic residue (Phe) disfavors the binding of anions (loganic acid) while improving affinity for more hydrophobic functional groups like esters (loganin).

Perturbing interactions with the carboxylic acid moiety of loganic acid and loganin were not the lone determinants of modifying enzymatic activities. The inactivity of both Ser324Glu mutants corroborates the importance of this hinging region in the I-helix for active site opening and closing in as described for bacterial CYPs (24, 25). Like bacterial CYP MycG (25), the homology model of CYP72A565 has a H-bonding network including the I-helix (Ser324) and G-helix (Ser274); contrastingly, nearly all members of the SLS lineage have Arg274. This conserved Arg274 residue may work in concert with the Glu324 as Ser274 and Ser324 do in the SLASs countering observations of reduced activities for the Ser324Glu SLAS mutants. The less severe effects observed when modifying Arg/Lys270 support this region as one worth exploring for modifying these enzymes' activities.

Based on these results, any attempts at further improving secologanic acid production by these *Camptotheca* SLASs should target His132 building from the Asp mutant. Most mutations one might suggest, however, present possible drawbacks: Increasing the chain length (Glu) might push loganic acid out of position for catalysis within the enzyme active site; replacement with another H-bond donor might be too long (Arg, Gln, Lys) or too short (Ser, Thr). The modest changes in size while retaining a H-bond donor presented by mutation to Asn or Tyr render them the best options for rational design at this site. These results discourage attempts to mutate His132 to hydrophobic residues where secologanic acid production from loganic acid is desired because His132Phe eliminated this activity. Where increased secologanin production and/or

selectivity for loganin are desired, the increase in SLS activity related to the His131Phe mutation supports attempting comparable shifts to nonpolar sidechains at His132.

The comparatively efficient SLS activity of *Camptotheca* CYP72A564 and CYP72A565 render them reasonable candidates for heterologous production of TIAs (*e.g.*, ibogamine, ajmaline, reserpine, vinblastine; Fig. 1). The multifunctionality of these two *Camptotheca* SLASs, however, presents the likely accumulation of another metabolite—secologanic acid—disruptive of strictosidine production. The production of secologanic acid is especially problematic because there are mixed reports (12, 18, 31) as to whether loganic acid *O*-methyltransferases (LAMTs) can efficiently produce secologanin from secologanic acid. In the event that it is not quickly converted to secologanin, once secologanic acid becomes coupled to tryptamine (enzymatically or non-enzymatically) *via* a Pictet-Spangler condensation (32), the strictosidinic acid produced is spontaneously lactamized to strictosamide *in lieu* of strictosidine. In the camptothecin-producing organisms *Camptotheca* (33), *Ophiorrhiza* (18), strictosamide feeds the species-specific pathway producing camptothecin; in other TIA pathways such as from *Catharanthus* and *Rauvolfia* (10, 34–36), producing strictosamide problematically diverts carbon and nitrogen from the production of desired TIAs like vinblastine and ajmaline.

The simultaneous increase in secologanin production and loss of activity towards loganic acid present the His131Phe mutants of *Camptotheca* CYP72A564 and CYP72A565 as excellent SLS candidates for heterologous production of strictosidine and, thereby, several frequently targeted TIAs. Indeed, these *Camptotheca* enzymes are quality alternatives to the *Catharanthus* isoforms employed to produce strictosidine in *S. cerevisiae* (15) where their use led to an undesirable accumulation of loganin. In my hands, the *Catharanthus* SLS has poorer expression in *E. coli* and stability than these *Camptotheca* CYPs (20). This poorer *ex vivo* stability suggests

that *Catharanthus* CYP72A1 and its multiple isoforms are generally less stable, explaining the intractability of loganin accumulation when several of these isoforms were coexpressed or overexpressed in yeast (15). The amenability of *Camptotheca* CYP72A564 and CYP72A565 to heterologous expression in *E. coli* and purification to considerable concentrations (up to ~50 μM) mark these enzymes as preferable to *Catharanthus* CYP72A1 when employing heterologous production systems.

Secologanic acid production as facilitated the SLASs of *Camptotheca* appears unique amongst TIA- and camptothecin-producing plants (17, 18, 33). The nearly identical SRSs of *Camptotheca* CYP72A564, CYP72A565, and the SLS/SLAS common ancestor support the multifunctionality of the *Camptotheca* enzymes as the ancestral state of the last common ancestor of TIA-producing plants. The presence of a nonfunctional LAMT homologue in *Camptotheca* (37) suggests that this ancestral TIA-producer employed loganin, secologanin, and strictosidine as intermediates (perhaps along with secologanic acid, strictosidinic acid, and strictosamide also present). Continuing with this hypothesis, the lineage leading to *Camptotheca* subsequently lost the ability to produce loganin, and the resulting biosynthetic pathway evolved to produce camptothecin *via* strictosamide derived solely from secologanic acid (33).

Contrasting this route to TIAs through secologanic acid and strictosidinic acid in the *Camptotheca* lineage, the SLS ancestor of the core Asterids lost the ability to turnover loganic acid with the His131Phe mutation and, thereby, became specialized for the conversion of loganin into secologanin. Because (for the extant, multifunctional *Camptotheca* CYPs) the His131Phe mutation alone increased loganin activity while eliminating loganic acid activity and the His132Asp mutation decimated loganin activity, the His132Asp mutation likely occurred first. Selective pressure subsequently caused the SLS-rescuing His131Phe mutation to maintain

sufficient levels of secologanin for TIA production. The lineage leading to *Camptotheca* at some point in time substituted Ser324 in place of the ancestral Glu present in the SLSs of the core Asterids. The Lys270Thr replacement likely occurred after the differentiation of the Lamiids and Campanulids as both Campanulid SLSs (*Lonicera* and *Nothapodytes*) retain the ancestral Lys.

The high sequence homology, shared motifs at the four loci identified in this report within clades, and common biosynthetic intermediates (*i.e.*, loganin, secologanin, and strictosidine) of TIA-producing plants strongly suggest an evolutionarily shared origin of TIA biosynthesis. Although Rai *et al*. (38) rightly conclude that an ancient origin for TIA biosynthesis requires a repeated loss of such capabilities across plant species, the function of secologanin production *and* the shared development of CYP72As for this function when taken together are in greater agreement with the loss of TIA-production in multiple lineages than convergent evolution of these metabolites. Furthermore, the apparent loss of LAMT activity in the lineage leading to *Camptotheca* (37) and the differentiation of Lamiid SLSs by Thr270 from those of other Asterids bolster the claim of an ancient TIA origin up to secologanin and strictosidine production.

Camptothecin production in *Camptotheca*, *Nothapodytes*, and *Ophiorrhiza*, however, appears to have arisen by convergent evolution. The divergence of *Camptotheca* in its biosynthetic intermediates after loganic acid (18, 33) and the occurrence of camptothecin production in three distinct lineages (Cornales, *Camptotheca*; Aquifoliales, *Nothapodytes*; *Ophiorrhiza*, Gentianales) substantiate convergent evolution. The differential trajectories of whole-genome duplications and development of camptothecin biosynthetic genes observed in *Camptotheca* and *Ophiorriza* (37, 38) further strengthens this claim. Comparable chromosome-level genome assemblies and analyses for *Nothapodytes* can illuminate this matter more; however, continued elucidation of camptothecin biosynthesis in these plants alongside

comparison of the enzymes employed will clarify the extensiveness of this instance of convergent evolution.

Although the use of ASR to identify key developments in CYP specialization is well reported for animal CYPs (39–42), this work highlights the usefulness of ASR and related methodologies for investigating CYP specialization in plant biosynthetic pathways. In light of the accumulating knowledge of plant specialized metabolism, the large number of CYPs in these pathways, and the burgeoning collections of genomic resources, ASR is emerging as a powerful tool to investigate how particular enzymes and the biosynthetic pathways they constitute have evolved through time. Increasing such studies provides guidance for the redesign of enzymes and creation of new-to-nature biosynthetic pathways needed to achieve greener chemical production *via* synthetic biology.

Each of the four sites identified by this combined ASR and homology modeling approach demonstrated important effects in differentiating SLAS and SLS activity. None of the mutations increased SLAS activity in any way whereas the effects on SLS activity were more variable. Intriguingly for improving heterologous production of strictosidine and thereby TIAs, mutating His131 to Phe in either of the *Camptotheca* SLASs simultaneously increased secologanin production *in vitro* and eliminated detectable loganic acid binding or turnover. Conversely, mutating His132 to Asp reduced secologanin production to background levels while the ability to produce secologanic acid was retained. As genetic and biochemical knowledge of specialized metabolism in plants continues to increase, studying the evolutionary development of biosynthetic pathways using methods like ASR will aid the design of these pathways and the enzymes that constitute them.

3.4. EXPERIMENTAL PROCEDURES

3.4.1. General

PCR reagents, plasmid purification, restriction enzymes, and T4 DNA ligase were used as recommended by New England Biolabs (NEB) unless otherwise noted. Vectors used in this study included the bacterial expression vectors pCWori (43) and pET28a (Novagen). *E. coli* cell strains used in this study included Top10 (Invitrogen), DH5α, and BL21(DE3) (NEB). All reagents were used as received from commercial sources. Loganin (Cayman Chemical) and loganic acid (Arctom Chemicals) were 98+% pure.

LC-MS was conducted using a Shimadzu LC-MS2010EV system at the Institute for Genomic Biology, University of Illinois Urbana-Champaign. Spectroscopic measurements were recorded with a Cary UV-Vis Bio100 dual beam spectrophotometer or Molecular Devices SpectraMax M-series plate reader as appropriate.

3.4.2. Ancestral Sequence Reconstruction

CYP72A sequences annotated as secologanin synthases (SLSs; 161 hits), 7-deoxyloganic acid hydroxylases (7DLHs; 23 hits), SLS-like (30 hits), or 7DLH-like (0 hits) in GenBank were selected for the ancestral sequence reconstruction. After curation to remove duplicates and CYP72A genes from species not known to produce loganic acid, the sequences chosen for inclusion in this analysis totaled 21 CYP72A genes, including three from *Camptotheca* (Table 3.1). *Arabidopsis thaliana* CYP734A1, a member of the closest subfamily to CYP72A, was chosen to root the tree as was done for a phylogeny of the whole CYP72A subfamily (21). The sequences were aligned with MEGA X (44) using MUSCLE (45) with the default parameters.

The model analysis tool of MEGA X identified the Le-Gascuel model (46) with a discrete gamma distribution (LG+G) as the best substitution model to use for the maximum likelihood

phylogenetic tree. The initial tree was obtained by BioNJ; tree topology, branch lengths, and rate parameters were optimized in a bootstrap analysis with 1,000 replicates. The results from this bootstrap analysis served as the initial tree used to infer the ancestral states of nodes in the phylogeny using the LG+G model treating the rates among sites as a gamma distribution with four gamma categories. The node identifications (Fig. 3.3), a multiple sequence alignment featuring all extant and ancestral sequences (Appendix E), and the amino acid probabilities for all ancestral sequences (Appendix F) are included.

3.4.3. Homology Modeling

The three-dimensional protein models based on single-template homology are the same as previously reported (20, 47) and were constructed utilizing MOE (Chemical Computing Group Inc.). As previously detailed, the template structure was determined via a multiple sequence alignment in MOE with all substrate-free CYPs in the PDB as of January 2018. All CYPs reported herein were modeled on substrate-free CYP2D6 (PDB 2F9Q) (48). Loganic acid was docked into each CYP72A before final minimization. The final models with docked loganic acid are supplied as *.pdb files (Appendices B, CYP72A564; C, CYP72A565; and G, SLS, SLAS Ancestor).

3.4.4. Site Directed Mutagenesis

I employed a three-step PCR method to create full-length, C-terminally His_6-tagged mutant ORFs for expression using the primers given in Table 3.2. CYP72A564 (GenBank MN815881) and CYP72A565 (GenBank MN815882) cloned into pCW from a previous study (20) were used as templates for mutagenesis. All PCR reactions were performed with Q5 High Fidelity DNA Polymerase (NEB) according to the manufacturer's recommendations with the annealing temperature for the primary and tertiary PCR reactions set at 69°C. DpnI (20 U; NEB) digestions

in Q5 reaction buffer lasted at least 3 h at 37°C before heat inactivation at 80°C for 20 min. Buffer and excess reagents were removed from the primary and tertiary PCR products using a GeneJET PCR Purification Kit (Thermo Fisher Scientific) according to the manufacturer's recommendations.

For the first PCR step, the 5' block was produced using a 5' NdeI primer at the N-terminus of the CYP and a mutant internal primer; the 3' block was produced with a 3' His_6 XbaI primer corresponding to the C-terminus of the CYP with another mutant internal primer (Table 3.2). The mutant primers were designed to contain an overlap of at least four nucleotides for later generation of full-length inserts. Following cycles with a 69°C annealing temperature, excess reagents were removed from these 5' and 3' blocks using a PCR purification kit.

For the second PCR step, which created a pool of full-length, double-stranded DNAs, approximately equimolar amounts of the 5' and 3' products were combined and amplified by PCR for 20 cycles with a low (40°C) annealing temperature. DpnI digestion preceded a third PCR seeded from the second to eliminate contaminating wild type plasmid. This third PCR used the gene-specific 5' and 3' primers, 30 cycles, and a 69°C annealing temperature to amplify enough of the full-length mutant transcript for subsequent insertion into the pCW expression vector.

Tertiary PCR products were digested with NdeI (20 U; NEB) and XbaI (20 U; NEB) at 37°C in 1x CutSmart buffer (NEB) for at least 2 h. Where a partial digest was required, only 2 U of enzyme were added, and the enzymes were heat denatured for 20 min at 80°C after the final 15 min of the 37°C digestion. Restriction endonuclease-digested PCR products were purified from low-melt agarose (Research Products International) using β-agarase (NEB) digestion and isopropyl alcohol precipitation. Purified inserts were ligated into pCW using T4 DNA ligase

(NEB). Dideoxy-sequencing over the full length of each coding sequence confirmed mutagenesis at only the desired position(s).

3.4.5. CYP, CPR Protein Expression and Purification

Mutant *Camptotheca* CYPs were purified as previously reported for the wild type enzymes (20), except Mg^{2+} and ATP were omitted from the resuspension step. Typical yields from two liters of culture were 1-2 mL of 10-60 μM P450.

Camptotheca CPR1 was also purified as previously reported (20) except for the replacement of Tris-HCl containing buffers with sodium-MOPS (pH 7.3) buffers containing 10% glycerol (49).

3.4.6. CYP, CPR Characterization

The concentration of properly folded CYPs (the P450 fraction) in each preparation was determined using an extinction coefficient of 106,000 M^{-1} cm^{-1} at 450 nm in a CO-reduced minus CO-oxidized difference spectrum (50). CPR activities were determined using the cytochrome *c* reduction method of Guengerich *et al.* (50).

3.4.7. Substrate-Induced Spectral Shift Assays (Type I Binding)

Concentrated CYPs were diluted to 0.60 μM in 100 mM $NaPO_4$ (pH 7.5) and their spectra obtained from 300-700 nm against a reference cuvette containing only buffer. Successive aliquots of loganic acid, loganin, or buffer were added to both the sample and reference cuvettes, respectively, to account for absorbance not associated with the CYPs. To correct for variations occurring from the removal and repositioning of cuvettes, spectra were first internally normalized by subtracting the absorbance at 700 nm from all other wavelengths. Then, the difference spectra were obtained by subtracting the appropriate buffer control spectrum from the substrate-containing spectrum of the same sample volume.

Binding isotherms from these spectra plotted the difference in ΔAbs from peak (at ~389 nm) to trough (at ~419 nm) against substrate concentration. The binding constants are reported with standard error from the curve fitting regression conducted in OriginPro2019 to a one-site binding model: $\Delta Abs = (\Delta A_{max} * [S])/(K_s + [S])$ where ΔA_{max} is the maximum difference between peak and trough (proportional to the amount of CYP bound to the substrate), [S] is the substrate concentration, and K_s is the binding constant. 95% confidence intervals for each parameter were calculated using a *F*-test (Tables 3.4 and 3.5). Any fits where at least one parameter's confidence interval included zero were interpreted as representing no binding of the compound assayed.

3.4.8. Steady State Kinetics

To estimate the rate of product formation, the rate of NADPH consumption was monitored spectroscopically in polystyrene 96-well microtiter plates with each 100 μL reaction (0.211 cm pathlength) performed at 30°C with technical triplicates. Individual wells contained 0.5 mU/μL full-length *Camptotheca* CPR1, 25 pmol P450 (250. nM), 60 μM DLPC (final concentration of 0.6% v/v methanol), no substrate or 50.0 μM to 5.00 mM substrate in 100 mM $NaPO_4$ (pH 7.5). A master mix containing the CPR, P450, and DLPC was preincubated on ice for two hours before adding substrate within the microtiter plate wells. Reactions were initiated by the addition of NADPH (~0.8 mM final concentration), shaking for 15 s before reading and monitoring at 340 nm every 15 s with shaking for 3 s between reads.

Initial rates were estimated via a linear regression of the absorbance versus time data, converted from a.u. min^{-1} to μM NADPH min^{-1} using an extinction coefficient of 6,220 M^{-1} cm^{-1}. Subtracting the rate of consumption in the absence of substrate from substrate-containing assays corrected for the background rate of NADPH turnover by CPR. Plotting these corrected rates against substrate concentration in OriginPro2019 and fitting to the Michaelis-Menten equation

by nonlinear curve fitting yielded estimates of the kinetic parameters (Fig. 3.11). As with the substrate binding isotherms, 95% confidence intervals for each parameter were calculated by an *F*-test (Tables 3.8 and 3.9), and I interpreted any fits where at least one parameter's confidence interval included zero as representing no fit (and non-significant turnover) for the compound assayed.

3.4.9. In Vitro Reconstitutions

SLAS activity was reconstituted *in vitro* with full-length *Camptotheca* CPR1 in the presence of DLPC (20). 50 μL reactions in 100 mM $NaPO_4$ (pH 7.5) buffer were started by the addition of a master mix to 12.5 pmol P450 (250. nM final concentration) or buffer. In the final reaction volume, the master mix contained 60 μM DLPC (0.6% v/v methanol final concentration), 0.5 mU/μL CPR, 10 mU/μL glucose-6-phosphate dehydrogenase, 250. μM substrate, 500 μM NADPH, and 1 mM glucose-6-phosphate. Upon addition of the master mix, reactions were incubated at 30°C for 60 min before quenching with 50 μL methanol (for reactions with loganin) or acetonitrile (for reactions with loganic acid) containing 250. μM caffeine as an internal standard. Secoiridoid production was then assessed for each mutant and wild type enzyme from triplicate assays via liquid chromatography-mass spectrometry.

3.4.10. Liquid Chromatography-Mass Spectrometry

A Shimadzu LCMS-2010EV system was used to analyze *in vitro* reactions. A Phenomenex Gemini C_{18} 110 Å (5 μM x 150 mm x 2 mm) column maintained at 40°C separated analytes from a 10 μL injection using a gradient at a 0.200 mL/min flow rate. The gradient for assays containing loganin was 10% acetonitrile (ACN) in 0.2% (v/v) acetic acid/water, 0 min; ramp to 30% ACN at 25 min; ramp to 98% ACN at 26 min; ramp to 10% ACN at 34 min; hold at 10% ACN until 35 min. The gradient for assays containing loganic acid started at 2% ACN in 0.2%

(v/v) acetic acid/water, ramp to 15% ACN at 15 min., ramp to 60% ACN at 22.5 min., ramp to 98% ACN at 25 min, hold to 27.5 min, ramp as quickly as possible to 2% ACN, hold until 32.5 min. A photodiode array detector maintained at 40°C with 1.2 nm slit width monitored 190-800 nm at 1.5625 Hz.

Selective ion monitoring for caffeine ($[M+H]^+$ 195.05), loganin ($[M+Na]^+$ 413.15), and secologanin ($[M+Na]^+$ 411.15) after electrospray ionization was used to quantitate these compounds in assays starting with loganin. Parameters for the mass spectrometer were optimized using the instruments AutoTune function: Nebulizing gas flow, 0.18 L/min; heat block temperature, 200°C; interface temperature, 250°C; interface bias voltage, +4.5 kV; interface current, 1.40 μA; CDL temperature, 230°C; CDL voltage, -20.0 V; focus lens, -2.5 V; entrance lens, -45.0 V; RF gain, 4335; RF offset, 5145; prerod bias, -3.6 V; main-rod bias, -3.5 V; aperture voltage, +0.0 V; conversion dynode, -8.0 kV; detector voltage, -2.10 kV.

Selective ion monitoring for caffeine ($[M+H]^+$ 195.10), loganic acid ($[M+Na]^+$ 399.05), and secologanic acid ($[M+Na]^+$ 397.05) after electrospray ionization was used to quantitate these compounds in assays starting with loganic acid. Parameters for the mass spectrometer were optimized using the instruments AutoTune function and were identical to those used for the loganin-containing assays.

The secoiridoids and the internal standard (caffeine) were quantitated from their extracted ion chromatograms using LabSolutions software (Shimadzu). The secoiridoid signal was divided by the caffeine signal, the wild type reactions averaged, and each individual reaction divided by this value to yield the normalized secoiridoid production as a percentage of wild type activity. A one-way ANOVA with Tukey's HSD test for *post hoc* pairwise comparison (α=0.05) compared the activity of the mutants (Tables 3.10-3.13).

3.5. TABLES AND FIGURES

Table 3.1. Accession numbers of sequences used for ancestral sequence reconstruction.

GenBank Accession ID	Species	Identifiers and Notes	Label Used in Phylogenetic Trees
MN815881	*Camptotheca acuminata*	Secologanic acid synthase; CYP72A564	Caa SLAS CYP72A564 | MN815881
MN815882	*Camptotheca acuminata*	Secologanic acid synthase; CYP72A565	Caa SLAS CYP72A565 | MN815882
MN815883	*Camptotheca acuminata*	Secologanic acid synthase-like; CYP72A730	Caa CYP72A730 | MN815883
Q05047.1	*Catharanthus roseus*	Secologanin synthase; CYP72A1	Cra SLS CYP72A1 | Q05047.1
AGX93062.1	*Catharanthus roseus*	7-deoxyloganic acid hydroxylase; CYP72A224	Cra 7DLH | AGX93062.1
AGX93046.1	*Amsonia hubrichtii*	Secologanin synthase-like protein	Ahu SLS | AGX93046.1
AGX93058.1	*Amsonia hubrichtii*	7-deoxyloganic acid hydroxylase-like protein	Ahu 7DLH | AGX93058.1
AGX93045.1	*Cinchona calisaya*	Secologanin synthase-like protein	Cca SLS | AGX93045.1
AGX93057.1	*Cinchona calisaya*	7-deoxyloganic acid hydroxylase-like protein	Cca 7DLH | AGX93057.1
AGX93044.1	*Lonicera japonica*	Secologanin synthase-like protein	Lja SLS | AGX93044.1
AGX93056.1	*Lonicera japonica*	7-deoxyloganic acid hydroxylase-like protein	Lja 7DLH | AGX93056.1
AQW38832.1	*Nothapodytes nimmoniana*	Secologanin synthase-like protein	Nni SLS | AQW38832.1
ATN39844.1	*Olea europaea*	7-deoxylaganic acid hydroxylase-like protein	Oeu 7DLH | ATN39844.1
BAP90521.1	*Ophiorrhiza pumila*	Secologanin synthase	Opu SLS | BAP90521.1
AGX93047.1	*Rauvolfia serpentina*	Secologanin synthase-like protein	Rsa SLS | AGX93047.1
AGX93059.1	*Rauvolfia serpentina*	7-deoxylaganic acid hydroxylase-like protein	Rsa 7DLH | AGX93059.1
BBI55002.1	*Swertia japonica*	Secologanin synthase	Sja SLS | BBI55002.1
AGX93048.1	*Tabernaemontana elegans*	Secologanin synthase-like protein	Tel SLS | AGX93048.1
AGX93060.1	*Tabernaemontana elegans*	7-deoxylaganic acid hydroxylase-like protein	Tel 7DLH | AGX93060.1
AGX93049.1	*Vinca minor*	Secologanin synthase-like protein	Vmi SLS | AGX93049.1
AGX93061.1	*Vinca minor*	7-deoxylaganic acid hydroxylase-like protein	Vmi 7DLH | AGX93061.1
NP_180239.1	*Arabidopsis thaliana*	CYP734A1; outgroup used to root the tree	At CYP734A1 | NP_180239.1

Table 3.2. Primers used in this study.

Name	Sequence (5'→3')
Caa SLAS NdeI 5'	ATTACATATGGAAATTCAAATGGACGTTCTGTAC
CYP72A564 His_6 XbaI 3'	TAATTCTAGATCAGTGGTGGTGGTGGTGGTGACATTTGAGC CTGCGGAAGATC
CYP72A564 H131F fwd	TAACAACTTTCAAAAGAATtttCATGCATACAGTCCGTTGACC AAG
CYP72A564 H132D fwd	TAACAACTTTCAAAAGAATCACgATGCATACAGTCCGTTGA CCAAG
CYP72A564 H131F,H132D fwd	CAACCATAACAACTTTCAAAAGAATtttgATGCATACAGTCCG TTGACCAAG
CYP72A564 SRS1 rev	GTTATGGTTGGCCAAGATATCCTTG
CYP72A564 R270T fwd	AATCaccGAGAGGTTGTCAGGTATCATCAACAG
CYP72A564 R270T rev	GATTTCTCTGTCAACTTCCCTCATCC
CYP72A564 S324E fwd	CATCgaaGAATGCAAGCTATTCTACTTCG
CYP72A564 S324E rev	GATGACATCCTCAATACTCATTCCAGC
CYP72A565 His_6 XbaI 3'	TAATTCTAGATCAGTGGTGGTGGTGGTGGTGACACTTGATC TTGCGGAAGACCAC
CYP72A565 H131F fwd	CAACGACTTCATGAAGAATtttCATGCATACAACCCACTAAC CAAATTC
CYP72A565 H132D fwd	GAATCACgATGCATACAACCCACTAACCAAATTC
CYP72A565 H131F,H132D fwd	CAACCACAACGACTTCATGAAGAATtttgATGCATACAACCC ACTAACCAAATTC
CYP72A565 SRS1 rev	GTTGTGGTTGGCCAATATATCC
CYP72A565 K270T fwd	GATAaccGACAGGTTGTCGGGGATCATC
CYP72A565 K270T rev	TATCTCTCCATCAACTTCCCTCATCC
CYP72A565 S324E fwd	CATCgaaGAATGCAAGCTTTTCTACTTTGCG
CYP72A565 S324E rev	GATGACATCTTCAATACTCATGCCAGC

Lowercase text denotes nucleotide mutation. Highlighted text denotes mutated codon.

Table 3.3. Type I Binding Isotherm Parameters of CYP72A564 and CYP72A565 Mutants

CYP	Mutation	K_s / *mM*		ΔA_{max} / *mA.U.*	
		loganic acid	loganin	loganic acid	loganin
72A564	Wild Type*	0.87 ± 0.06 [a]	0.26 ± 0.02 [a]	39.5 ± 1.3 [a]	46.6 ± 1.2 [a]
	H131F	n.d.	0.055 ± 0.003 [b]	n.d.	62.5 ± 0.8 [b]
	H132D	n.d.	1.19 ± 0.05 [c]	n.d.	42.7 ± 1.0 [a]
	H131F,H132D	n.d.	0.26 ± 0.01 [a]	n.d.	64.8 ± 1.1 [b]
	R270T	1.5 ± 0.4 [a,b]	0.37 ± 0.01 [d]	40.6 ± 6.0 [a]	46.6 ± 0.4 [a]
	S324E	2.5 ± 1.1 [b]	1.3 ± 0.3 [c]	15.9 ± 4.7 [a]	30.4 ± 4.8 [a]
72A565	Wild Type*	2.4 ± 0.4 [a]	0.84 ± 0.07 [a]	43.9 ± 4.6 [a]	51.0 ± 2.2 [a,b]
	H131F	n.d.	0.28 ± 0.01 [b]	n.d.	58.2 ± 1.0 [a]
	H132D	n.d.	3.2 ± 1.3 [c]	n.d.	22.9 ± 6.5 [a,b,c,d]
	H131F,H132D	n.d.	0.35 ± 0.04 [b]	n.d.	42.5 ± 2.0 [b]
	K270T	2.0 ± 0.6 [a]	1.2 ± 0.1 [d]	22.4 ± 4.5 [a]	33.7 ± 1.3 [c]
	S324E	n.d.	0.97 ± 0.25 [a,c,d]	n.d.	17.5 ± 2.4 [d]

n.d., no data because one or more fit parameters were not significantly different than zero
Different superscript letters in a column for either CYP denote significantly different fit parameters by comparing 95% confidence intervals calculated via an *F*-test (see Tables 3.4-3.5).
*Data from Miller *et al.* (20)

Table 3.4. 95% Confidence Intervals of CYP72A564 Substrate Binding Isotherms.

Substrate	Mutation	K_s / *mM*			ΔA_{max} / *mA.U.*		
		Lower	Upper	Group	Lower	Upper	Group
loganic acid	Wild Type	0.763	1.003	a	36.8	42.4	a
	H131F	--	0.002		1.5	2.2	
	H132D	--	--		--	--	
	H131F,H132D	--	0.579		0.1	0.8	
	K270T	0.916	2.561	a,b	30.9	57.9	a
	S324E	1.157	8.284	b	9.8	40.5	a
loganin	Wild Type	0.218	0.302	a	44.0	49.4	a
	H131F	0.048	0.062	b	60.8	64.2	b
	H132D	1.097	1.302	c	40.7	44.9	a
	H131F,H132D	0.232	0.286	a	62.5	67.3	b
	K270T	0.354	0.389	d	45.4	47.2	a
	S324E	0.767	2.421	c	22.7	44.9	a

Table 3.5. 95% Confidence Intervals of CYP72A565 Substrate Binding Isotherms.

Substrate	Mutation	K_s / *mM*			ΔA_{max} / *mA.U.*		
		Lower	Upper	Group	Lower	Upper	Group
loganic acid	Wild Type	1.77	3.47	a	35.7	56.8	a
	H131F	--	0.01		0.4	0.8	
	H132D	--	0.98		--	4.3	
	H131F,H132D	--	0.02		0.7	1.4	
	K270T	1.06	4.65	a	15.2	40.8	a
	S324E	--	--		--	--	
loganin	Wild Type	0.700	1.016	a	46.5	56.2	a,b
	H131F	0.256	0.312	b	56.1	60.3	a
	H132D	1.596	10.707	c	14.2	60.6	a,b,c,d
	H131F,H132D	0.264	0.453	b	38.4	47.3	b
	K270T	1.039	1.395	d	31.1	36.8	c
	S324E	0.540	1.864	a,c,d	13.1	25.5	d

Table 3.6. Steady State Kinetic Parameters of CYP72A564 Mutants

Substrate	Mutation	K_M	V_{max}	k_{cat}	k_{cat}/K_M
		mM	*μM min^{-1}*	*min^{-1}*	*min^{-1} mM^{-1}*
loganic acid	Wild Type*	2.67 ± 0.80^{a}	3.49 ± 0.53^{a}	13.9 ± 2.1	5.2 ± 1.7
	H131F	n.d.	n.d.	n.d.	n.d.
	H132D	n.d.	n.d.	n.d.	n.d.
	H131F,H132D	n.d.	n.d.	n.d.	n.d.
	R270T	3.54 ± 0.51^{a}	5.00 ± 0.40^{a}	20.0 ± 1.6	5.65 ± 0.94
	S324E	n.d.	n.d.	n.d.	n.d.
loganin	Wild Type*	0.24 ± 0.02^{a}	3.00 ± 0.08^{a}	12.0 ± 0.3	49.5 ± 4.8
	H131F	0.05 ± 0.01^{b}	1.98 ± 0.04^{b}	7.9 ± 0.2	149 ± 18
	H132D	1.44 ± 0.06^{c}	3.41 ± 0.06^{c}	13.7 ± 0.3	9.5 ± 0.5
	H131F,H132D	0.38 ± 0.16^{a,d}	1.60 ± 0.20^{b}	6.4 ± 0.8	17.1 ± 7.4
	R270T	1.03 ± 0.32^{c,d}	4.95 ± 0.61^{d}	19.8 ± 2.5	19.2 ± 6.5
	S324E	1.3 ± 0.1^{c}	2.50 ± 0.08^{e}	10.0 ± 0.3	5.2 ± 1.7

n.d., no data because one or more fit parameters were not significantly different than zero
Different superscript letters in a column for either CYP denote significantly different fit parameters by comparing 95% confidence intervals calculated via an *F*-test (see Table 3.8).
*Data from Miller *et al.* (20)

Table 3.7 Steady State Kinetic Parameters of CYP72A565 Mutants

Substrate	Mutation	K_M	V_{max}	k_{cat}	k_{cat}/K_M
		mM	*μM min^{-1}*	*min^{-1}*	*min^{-1} mM^{-1}*
loganic acid	Wild Type*	3.57 ± 0.42^{a}	4.86 ± 0.32^{a}	19.4 ± 1.3	5.5 ± 0.7
	H131F	n.d.	n.d.	n.d.	n.d.
	H132D	n.d.	n.d.	n.d.	n.d.
	H131F,H132D	n.d.	n.d.	n.d.	n.d.
	K270T	n.d.	n.d.	n.d.	n.d.
	S324E	n.d.	n.d.	n.d.	n.d.
loganin	Wild Type*	0.72 ± 0.07^{a}	3.82 ± 0.13^{a}	15.3 ± 0.5	21.2 ± 2.1
	H131F	0.29 ± 0.08^{a}	1.92 ± 0.15^{b}	7.7 ± 0.6	26.7 ± 7.8
	H132D	5.3 ± 2.1^{b,c}	5.2 ± 1.2^{a,c,d}	20.6 ± 4.9	3.9 ± 1.8
	H131F,H132D	1.7 ± 0.1^{b}	7.17 ± 0.19^{c}	31.8 ± 1.4	16.8 ± 1.1
	K270T	2.0 ± 0.7^{a,b,c}	2.59 ± 0.45^{b,d}	10.3 ± 1.8	5.2 ± 2.2
	S324E	3.6 ± 0.7^{c}	4.84 ± 0.50^{a}	19.4 ± 2.0	5.4 ± 1.2

n.d., no data because one or more fit parameters were not significantly different than zero
Different superscript letters in a column for either CYP denote significantly different fit parameters by comparing 95% confidence intervals calculated via an *F*-test (see Table 3.9).
*Data from Miller *et al.* (20)

Table 3.8. 95% Confidence Intervals of CYP72A564 Steady State Kinetic Parameters.

Substrate	Mutation	K_M / *mM*			V_{max} / μM min^{-1}		
		Lower	Upper	Group	Lower	Upper	Group
loganic acid	Wild Type	1.53	5.20	a	2.65	5.00	a
	H131F	--	--		--	--	
	H132D	--	--		--	--	
	H131F,H132D	--	--		--	--	
	R270T	2.64	4.91	a	4.28	6.01	a
	S324E	--	--		--	--	
loganin	Wild Type	0.197	0.296	a	2.84	3.16	a
	H131F	0.041	0.068	b	1.89	2.06	b
	H132D	1.317	1.582	c	3.28	3.55	c
	H131F,H132D	0.131	0.922	a,d	1.21	2.13	b
	R270T	0.519	2.017	c,d	3.84	6.55	d
	S324E	1.063	1.485	c	2.34	2.69	e

Table 3.9. 95% Confidence Intervals of CYP72A565 Steady State Kinetic Parameters.

Substrate	Mutation	K_M / *mM*			V_{max} / μM min^{-1}		
		Lower	Upper	Group	Lower	Upper	Group
loganic acid	Wild Type	2.80	4.64	a	4.27	5.63	a
	H131F	--	--		--	--	
	H132D	--	--		--	--	
	H131F,H132D	--	--		--	--	
	K270T	--	--		--	--	
	S324E	--	--		--	--	
loganin	Wild Type	0.578	0.865	a	3.54	4.09	a
	H131F	0.119	0.457	a	1.61	2.24	b
	H132D	0.966	9.687	b,c	2.58	7.72	a,c,d
	H131F,H132D	1.501	1.920	b	6.78	7.57	c
	K270T	0.425	3.526	a,b,c	1.65	3.52	b,d
	S324E	2.155	5.006	c	3.79	5.90	a

Table 3.10. Tukey's HSD pairwise comparison of secologanic acid production by CYP72A564 mutants.

Comparison	Difference in Means	SEM	q Value	Probability	Alpha	Significant?
WT -P450	0.9838	0.09348	14.88	7.7E-07	0.05	Yes
H131F -P450	-0.006220	0.09348	0.09412	1.00	0.05	No
H131F WT	-0.9900	0.09348	14.98	7.1E-07	0.05	Yes
H132D -P450	0.9516	0.09348	14.40	1.2E-06	0.05	Yes
H132D WT	-0.03213	0.09348	0.4861	1.00	0.05	No
H132D H131F	0.9579	0.09348	14.49	1.1E-06	0.05	Yes
H131F,H132D -P450	0.02484	0.09348	0.3758	1.00	0.05	No
H131F,H132D WT	-0.9589	0.09348	14.51	1.1E-06	0.05	Yes
H131F,H132D H131F	0.03106	0.09348	0.4699	1.00	0.05	No
H131F,H132D H132D	-0.9268	0.09348	14.02	1.7E-06	0.05	Yes
R270T -P450	0.7681	0.09348	11.62	1.6E-05	0.05	Yes
R270T WT	-0.2157	0.09348	3.263	0.31	0.05	No
R270T H131F	0.7743	0.09348	11.71	1.5E-05	0.05	Yes
R270T H132D	-0.1836	0.09348	2.777	0.48	0.05	No
R270T H131F,H132D	0.7432	0.09348	11.24	2.4E-05	0.05	Yes
S324E -P450	0.3929	0.09348	5.944	0.012	0.05	Yes
S324E WT	-0.5908	0.09348	8.938	3.0E-04	0.05	Yes
S324E H131F	0.3992	0.09348	6.038	0.011	0.05	Yes
S324E H132D	-0.5587	0.09348	8.452	5.2E-04	0.05	Yes
S324E H131F,H132D	0.3681	0.09348	5.569	0.019	0.05	Yes
S324E R270T	-0.3751	0.09348	5.675	0.017	0.05	Yes

Table 3.11. Tukey's HSD pairwise comparison of secologanic acid production by CYP72A565 mutants.

Comparison	Difference in Means	SEM	q Value	Probability	Alpha	Significant?
WT -P450	0.9886	0.04183	33.42	< 1E-9	0.05	Yes
H131F -P450	0.004620	0.04183	0.1560	1	0.05	No
H131F WT	-0.9839	0.04183	33.26	< 1E-9	0.05	Yes
H132D -P450	0.5355	0.04183	18.10	6.4E-08	0.05	Yes
H132D WT	-0.4530	0.04183	15.31	5.2E-07	0.05	Yes
H132D H131F	0.5309	0.04183	17.95	6.9E-08	0.05	Yes
H131F,H132D -P450	6.430E-04	0.04183	0.02174	1	0.05	No
H131F,H132D WT	-0.9879	0.04183	33.40	< 1E-9	0.05	Yes
H131F,H132D H131F	-0.003970	0.04183	0.1343	1	0.05	No
H131F,H132D H132D	-0.5349	0.04183	18.08	6.4E-08	0.05	Yes
K270T -P450	0.2373	0.04183	8.020	8.8E-04	0.05	Yes
K270T WT	-0.7513	0.04183	25.40	1.5E-07	0.05	Yes
K270T H131F	0.2326	0.04183	7.864	0.0011	0.05	Yes
K270T H132D	-0.2983	0.04183	10.08	8.2E-05	0.05	Yes
K270T H131F,H132D	0.2366	0.04183	7.998	9.0E-04	0.05	Yes
S324E -P450	0.3897	0.04183	13.18	3.7E-06	0.05	Yes
S324E WT	-0.5988	0.04183	20.24	1.2E-07	0.05	Yes
S324E H131F	0.3851	0.04183	13.02	4.2E-06	0.05	Yes
S324E H132D	-0.1458	0.04183	4.928	0.044	0.05	Yes
S324E H131F,H132D	0.3891	0.04183	13.15	3.7E-06	0.05	Yes
S324E K270T	0.1525	0.04183	5.155	0.033	0.05	Yes

Table 3.12. Tukey's HSD pairwise comparison of secologanin production by CYP72A564 mutants.

Comparison	Difference In Means	SEM	q Value	Probability	Alpha	Significant?
WT -P450	0.9997	0.09585	14.74984	8.7E-7	0.05	Yes
H131F -P450	1.8468	0.09585	27.24808	4.1E-8	0.05	Yes
H131F WT	0.8471	0.09585	12.49824	6.9E-6	0.05	Yes
H132D -P450	0.08288	0.09585	1.22285	0.97	0.05	No
H132D WT	-0.91682	0.09585	13.52699	2.6E-6	0.05	Yes
H132D H131F	-1.76392	0.09585	26.02523	3.9E-6	0.05	Yes
H131F,H132D -P450	0.79978	0.09585	11.80012	1.4E-5	0.05	Yes
H131F,H132D WT	-0.19992	0.09585	2.94972	0.41	0.05	No
H131F,H132D H131F	-1.04702	0.09585	15.44796	4.7E-7	0.05	Yes
H131F,H132D H132D	0.7169	0.09585	10.57727	4.8E-5	0.05	Yes
R270T -P450	0.78529	0.09585	11.58633	1.7E-5	0.05	Yes
R270T WT	-0.21441	0.09585	3.16352	0.34	0.05	No
R270T H131F	-1.06151	0.09585	15.66176	3.8E-7	0.05	Yes
R270T H132D	0.70241	0.09585	10.36347	6.0E-5	0.05	Yes
R270T H131F,H132D	-0.01449	0.09585	0.2138	1.00	0.05	No
S324E -P450	0.19991	0.09585	2.94955	0.41	0.05	No
S324E WT	-0.79979	0.09585	11.80029	1.4E-5	0.05	Yes
S324E H131F	-1.64689	0.09585	24.29853	1.7E-6	0.05	Yes
S324E H132D	0.11703	0.09585	1.7267	0.87	0.05	No
S324E H131F,H132D	-0.59987	0.09585	8.85057	3.3E-4	0.05	Yes
S324E R270T	-0.58538	0.09585	8.63678	4.2E-4	0.05	Yes

Table 3.13. Tukey's HSD pairwise comparison of secologanin production by CYP72A565 mutants.

Comparison	Difference in Means	SEM	q Value	Probability	Alpha	Significant?
WT -P450	0.99927	0.03613	39.10973	< 1E-9	0.05	Yes
H131F -P450	1.20867	0.03613	47.30507	< 1E-9	0.05	Yes
H131F WT	0.2094	0.03613	8.19534	7.1E-4	0.05	Yes
H132D -P450	0.05168	0.03613	2.0226	0.78	0.05	No
H132D WT	-0.94759	0.03613	37.08713	< 1E-9	0.05	Yes
H132D H131F	-1.15699	0.03613	45.28247	< 1E-9	0.05	Yes
H131F,H132D -P450	0.45526	0.03613	17.81791	7.4E-8	0.05	Yes
H131F,H132D WT	-0.54402	0.03613	21.29182	1.3E-7	0.05	Yes
H131F,H132D H131F	-0.75341	0.03613	29.48717	3.9E-8	0.05	Yes
H131F,H132D H132D	0.40358	0.03613	15.79531	3.4E-7	0.05	Yes
K270T -P450	0.35981	0.03613	14.08213	1.5E-6	0.05	Yes
K270T WT	-0.63947	0.03613	25.02759	1.4E-7	0.05	Yes
K270T H131F	-0.84886	0.03613	33.22294	< 1E-9	0.05	Yes
K270T H132D	0.30813	0.03613	12.05954	1.1E-5	0.05	Yes
K270T H131F,H132D	-0.09545	0.03613	3.73577	0.19	0.05	No
S324E -P450	0.19876	0.03613	7.77905	1.2E-3	0.05	Yes
S324E WT	-0.80051	0.03613	31.33067	< 1E-9	0.05	Yes
S324E H131F	-1.00991	0.03613	39.52602	< 1E-9	0.05	Yes
S324E H132D	0.14708	0.03613	5.75646	0.015	0.05	Yes
S324E H131F,H132D	-0.2565	0.03613	10.03885	8.6E-5	0.05	Yes
S324E K270T	-0.16105	0.03613	6.30308	7.6E-3	0.05	Yes

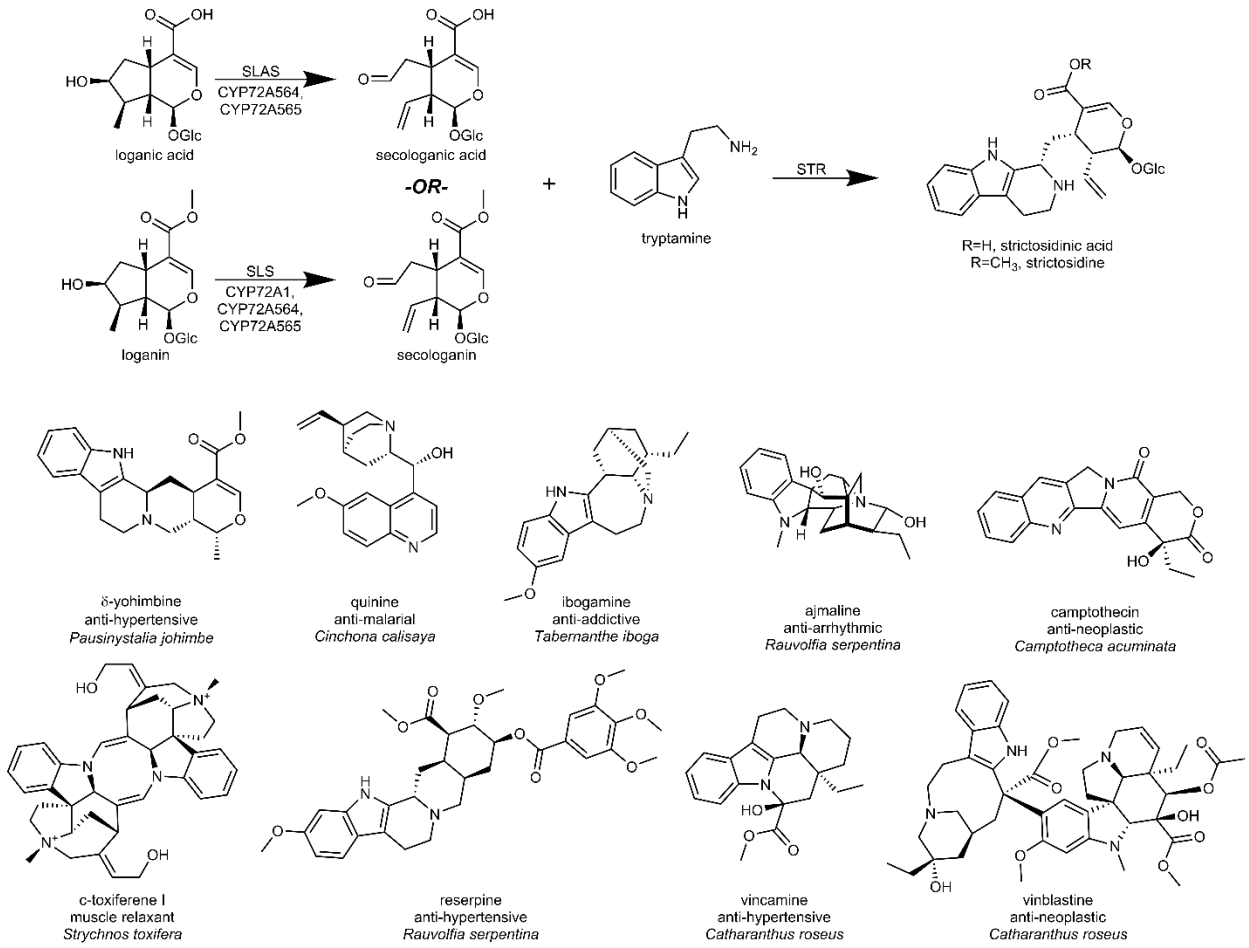


Figure 3.1. Production of strictosidinic acid derivatives and the terpene indole alkaloids (TIAs) derived from them.

CYP72As denoted secologanin synthases (SLSs) and secologanic acid synthases (SLASs) produce secoiridoids from loganin and loganic acid, respectively, that react with tryptamine to form strictosidinic acid derivatives. Species-specific pathways then modify these core structures into a variety of biologically active, medicinally useful TIAs.

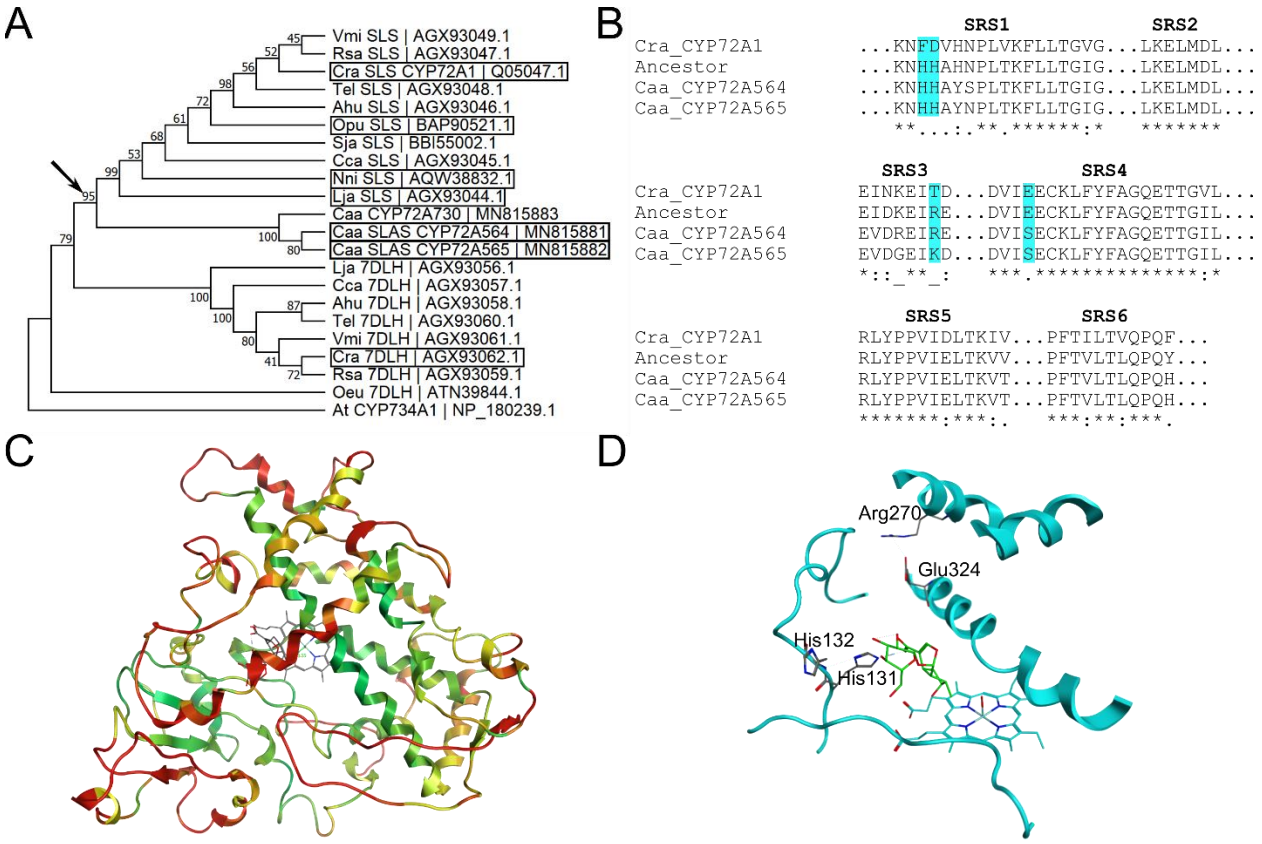


Figure 3.2: Ancestral sequence reconstruction (ASR) of the secologanin/secologanic acid synthase (SLS/SLAS) common ancestor.

(A) Maximum likelihood, bootstrap consensus tree of CYP72As annotated as SLS, 7-deoxyloganic acid hydroxylase (7DLH), SLS-like, or 7DLH-like from strictosidine-producing species used for the ASR. The evolutionary history was inferred using the model of Le and Gascuel (46) with a discrete Gamma distribution using MEGA X (44). Genes surrounded by a box denote those for which the given activity has been reported in the literature (4, 5, 7, 16–20). The arrow points to the node corresponding to the SLS/SLAS Ancestor featured in this study. Abbreviations and descriptions for the genes are located in Table 3.2. (B) Multiple sequence alignment of the predicted substrate recognition sequences (SRSs) comparing the *Catharanthus* CYP72A1, *Camptotheca* CYP72A564 and CYP72A565, and the predicted SLS/SLAS common ancestor. Cyan boxes indicate residues mutated in the *Camptotheca* enzymes for analysis. (C) Ribbon cartoon of the SLS/SLAS common ancestor docked with loganic acid (Appendix G). Coloration corresponds to the average backbone RMSD of the ancestor compared to models for *Camptotheca* CYP72A564 and CYP72A565: Green, 0.0 Å; Yellow, 2.0 Å; Red, 4.0 Å. (D) SLS/SLAS common ancestor homology model docked with loganic acid (green) with the predicted SRSs (cyan ribbon cartoon) and the four amino acids selected for mutation in the *Camptotheca* enzymes shown.

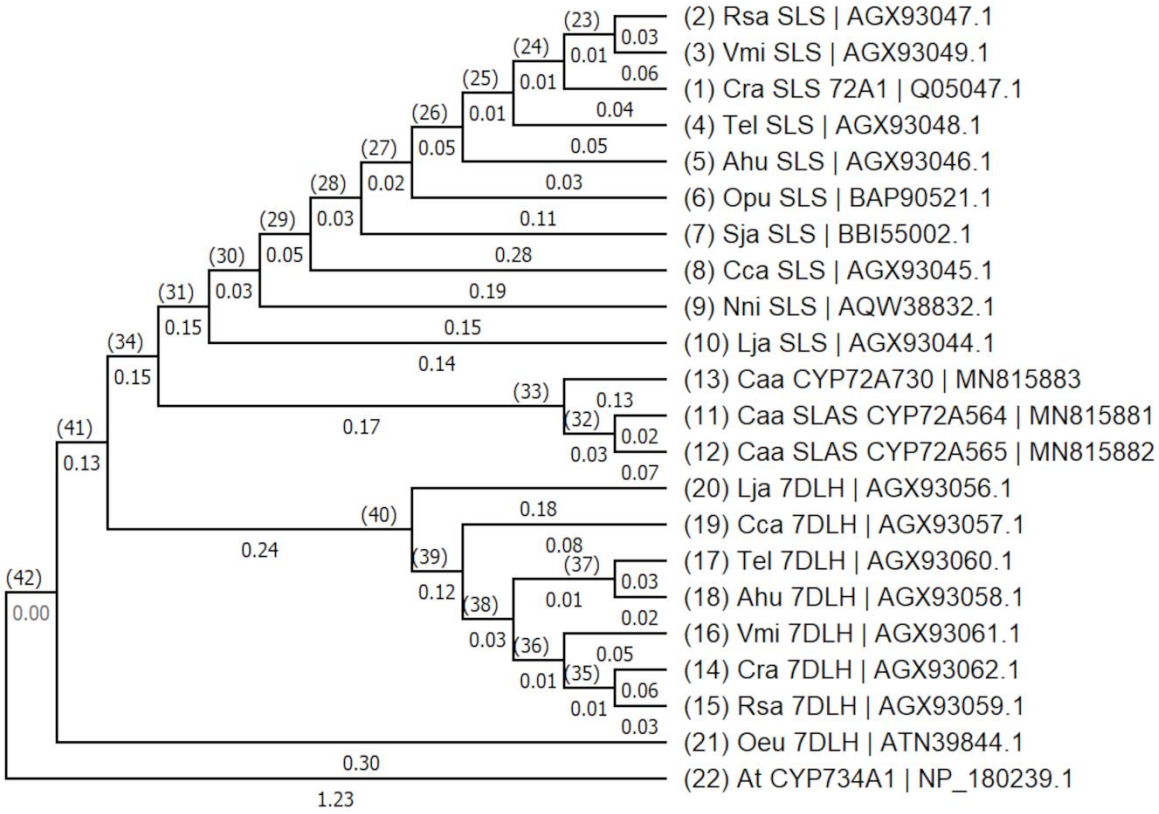


Figure 3.3. Bootstrap consensus tree containing node identifications (in parentheses) and branch lengths (below the branch) for the ancestral sequence reconstruction.

```
>SLS,SLAS_Ancestor Node_34
MEMDLLYKSIAASVLVALLVWAWRVLNWAWFTPKRLEKRLRQQGFKGNPYRLLVGDLKESSMML
KEAMSKPIPFSNDIVPRIMPHIVKTIETYGKNSFTWMGRMPRIHIMEPELIREVLTNSNKFQKN
HHAHNPLTKFLLTGIGSLEGDKWAKHRRIISPAFHLEKLKTMLPAFYISYDDLLSKWEKVASTE
GSVEVDVFPTFDTLTSDVISRVAFGSNYEEGGKIFQLLKELMDLTIEVMRSVYIPGWSFLPTKR
NQRMREIDKEIRERLSNIINKRVKAMKAGEPSGDDLLGVLLESNFKEIQRQGNKKNAGMSIEDV
IEECKLFYFAGQETTGILLTWTMVLLSRHPEWQERAREEVLQVFGNGKPDFDRLNHLKIVSMIL
YEVLRLYPPVIELTKVVHEETKLGNLTIPAGVQIMMPTILLHRDKEIWGDDAMEFNPGRFAEGV
AKATKSQVSYIPFSWGPRICIGQNFALLQAKMALAMILQRFSFDLSPSYAHAPFTVLTLQPQYG
AHVIFRKLKS
```

Key

100%	95-100%
90-95%	80-90%
70-80%	60-70%
0 -60%	

Figure 3.4. Amino acid sequence of SLS, SLAS common ancestor color coded by probability from ancestral sequence reconstruction.

```
Cra_CYP72A1       MEMDMDTIRKAIAATIFALVMAWAWRVLDWAWFTPKRIEKRLRQQGFRGNPYRFLVGDVKESGKMHQEAL    70
Ancestor          --MEMDLLYKSIAASVLVALLVWAWRVLNWAWFTPKRLEKRLRQQGFKGNPYRLLVGDLKESSMMLKEAM    68
Caa_CYP72A564     MEIQMDVLYKSIAASVAVVFLVYAWKMLNWAYLTPKRIEKCLRKQGFKGNSYRLLVGDLKESSMMLKEAM    70
Caa_CYP72A565     MEIQMDVLYKSIAASVAVVFLVYAWKMLNWAYLTPKRIEKCLRKQGFKGNSYRLLVGDLKESSMMLKETM    70
                    ::** : *:***:: . .:.:**::*:**::****:** **:***:** **:****:***. * :*::

Cra_CYP72A1       SKPMEFNNDIVPRLMPHINHTINTYGRNSFTWMGRIPRIHVMEPELIKEVLTHSSKYQKNFDVHNPLVKF   140
Ancestor          SKPIPFSNDIVPRIMPHIVKTIETYGKNSFTWMGRMPRIHIMEPELIREVLTNSNKFQKNHHAHNPLTKF   138
Caa_CYP72A564     SKPIPVSQDIVQRLMPHVVKTIQTYGKNSFTWIGRMPRVHIMEPELIKDILANHNNFQKNHHAYSPLTKF   140
Caa_CYP72A565     SKPINVSEDIVQRVMPHVIKTIDTYGKNSFTWIGRMPRVHIMEPDLIKDILANHNDFMKNHHAYNPLTKF   140
                  ***: ..:*** *:***: :**:***:*****:**:**:*:***:**:::*:: ..: **...:.**.**

Cra_CYP72A1       LLTGVGSFEGAKWSKHRRIISPAFTLEKLKSMLPAFAICYHDMLTKWEKIAEKQGSHEVDIFPTFDVLTS   210
Ancestor          LLTGIGSLEGDKWAKHRRIISPAFHLEKLKTMLPAFYISYDDLLSKWEKVASTEGSVEVDVFPTFDTLTS   208
Caa_CYP72A564     LLTGIGSLEGEKWAKHRRIISPSFHLEKLKTMLPAFYVSYDELLGKWERESSTKGSVEVDLFPTFDTLTS   210
Caa_CYP72A565     LLTGIGSLEGDKWAKHRRIISPSFHLEKLKTMLPAFYVSYDDLLTKWEQQCSSKGSVEIDLFPTFDTLTS   210
                  ****:**:** **:********:* *****:***** :.*.::* ***: ...:** *:*:*****.***

Cra_CYP72A1       DVISKVAFGSTYEEGGKIFRLLKELMDLTIDCMRDVYIPGWSYLPTKRNKRMKEINKEITDMLRFIINKR   280
Ancestor          DVISRVAFGSNYEEGGKIFQLLKELMDLTIEVMRSVYIPGWSFLPTKRNQRMREIDKEIRERLSNIINKR   278
Caa_CYP72A564     DVISRVAFGSSYGEGGRIFILLKELMDLTVDVMRSVYVPGWSLLPTKRNQRMREVDREIRERLSGIINSR   280
Caa_CYP72A565     DVISRVAFGSSYGEGGRIFILLKELMDLTVDVMRSVYVPGSSFLPTKRNNRMREVDGEIKDRLSGIINSR   280
                  ****:*****.* ***:** *********:: **.**:** * ******:**:*:: ** : *  ***.*

Cra_CYP72A1       MKALKAGEPGEDDLLGVLLESNIQEIQKQGNKKDGGMSINDVIEECKLFYFAGQETTGVLLTWTTILLSK   350
Ancestor          VKAMKAGEPSGDDLLGVLLESNFKEIQRQGNKKNAGMSIEDVIEECKLFYFAGQETTGILLTWTMVLLSR   348
Caa_CYP72A564     VKAMKAGEPSGDDLLGTLLESNFREIERLGNKKNAGMSIEDVISECKLFYFAGQETTGILLTWTCVLLSR   350
Caa_CYP72A565     VKAMKAGEPSGEDLLGTLLESNFKEIERLGNKKNAGMSIEDVISECKLFYFAGQETTGILLTWTCVLLSR   350
                  :**:*****. :****.*****::**:: ****:.****:***.***************:***** :***:

Cra_CYP72A1       HPEWQERAREEVLQAFGKNKPEFERLNHLKYVSMILYEVLRLYPPVIDLTKIVHKDTKLGSYTIPAGTQV   420
Ancestor          HPEWQERAREEVLQVFGNGKPDFDRLNHLKIVSMILYEVLRLYPPVIELTKVVHEETKLGNLTIPAGVQL   418
Caa_CYP72A564     HPEWQERAREEIFQVFGNGKLDFDRVQGLKIVPMILYEVLRLYPPVIELTKVTYEEQKLGNLTIPAGVQL   420
Caa_CYP72A565     HPEWQERAREEIFQVFGNGKVDFDRVQNLKIVPMILYEVLRLYPPVIELTKVTYEEQKLGNLTIPAGVQL   420
                  ***********::*.**:.* :*:*:: ** * ***************:***:.::: ***. *****.*:

Cra_CYP72A1       MLPTVMLHREKSIWGEDAMEFNPMRFVDGVANATKNNVTYLPFSWGPRVCLGQNFALLQAKLGLAMILQR   490
Ancestor          MMPTILLHRDKEIWGDDAMEFNPGRFAEGVAKATKSQVSYIPFSWGPRICIGQNFALLQAKMALAMILQR   488
Caa_CYP72A564     MMPSILLHRDKEMWGDDATEFNPGRFAEGVAKAVKSPFFYIPFSWGPRICVGQNFALLQAKMALAMILQR   490
Caa_CYP72A565     MMPSILLHRDQEMWGADSKEFNPGRFADGISKAVKSPFFYIPFSWGPRICVGQNFALLQAKMALTMILQR   490
                  *:*:::***::.:** *: **** **.:*:::*.*. . *:********:*:**********:.*:*****

Cra_CYP72A1       FKFDVAPSYVHAPFTILTVQPQFGSHVIYKKLES   524
Ancestor          FSFDLSPSYAHAPFTVLTLQPQYGAHVIFRKLKS   522
Caa_CYP72A564     FSFDLSPTYAHAPFTVLTLQPQHGAQVIFRRLKC   524
Caa_CYP72A565     FTFDLSPTYAHAPFTVLTLQPQHGAQVVFRKIKC   524
                  *.**::*:*.*****:**:***.*::*:::::::.
```

Figure 3.5. Multiple sequence alignment of *Catharanthus* CYP72A1, *Camptotheca* CYP72A564 and CYP72A565, and their common ancestor.

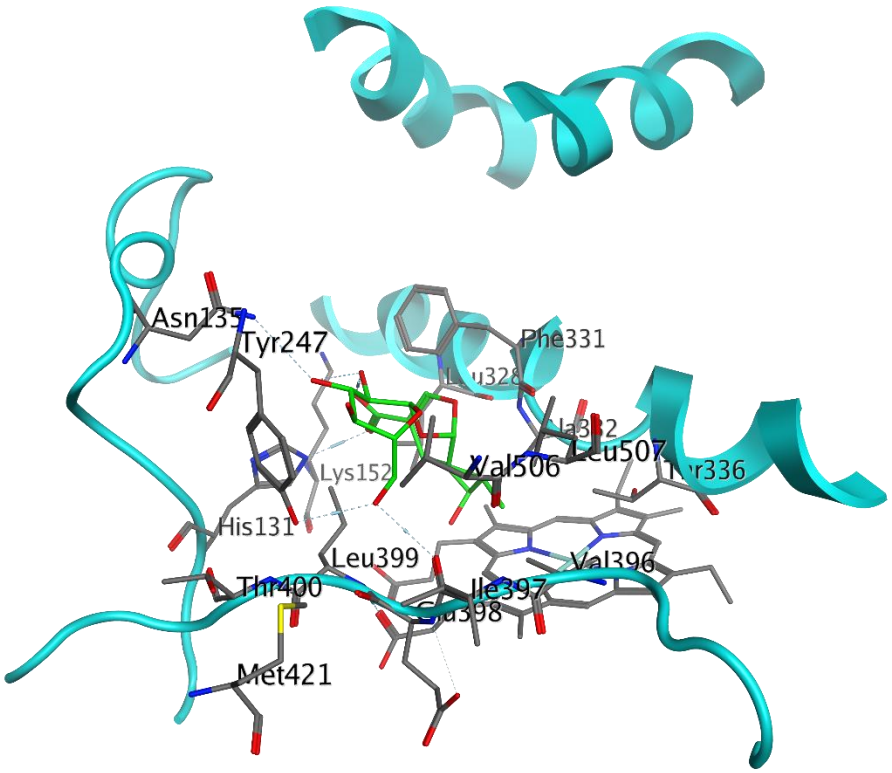


Figure 3.6. Amino acid contacts of docked loganic acid in homology model of SLS, SLAS common ancestor.
Residues within 4.5 Å of loganic acid (green) are shown with SRSs 1-5 shown in ribbon (cyan). Appendix G contains the full structure file.

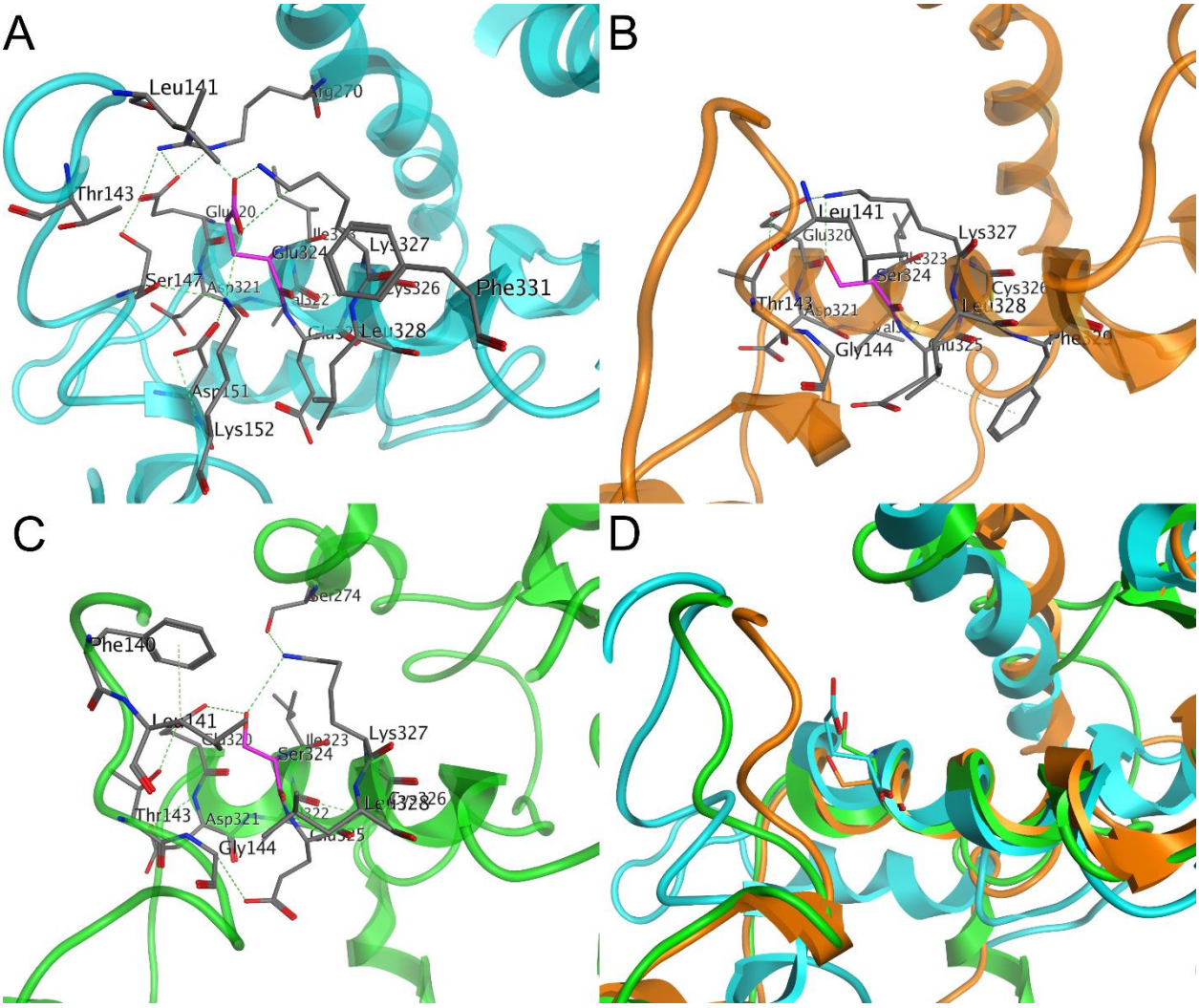


Figure 3.7. Hydrogen bonding networks around site 324 in homology models.
All residues within 4.5 Å of the amino acid at site 324 (magenta) are shown for the SLS, SLAS ancestor (A), CYP72A564 (B), CYP72A565 (C). (D) A ribbon overlay of all three models with site 324 shown. Hydrogen bonds are shown as green dotted lines.

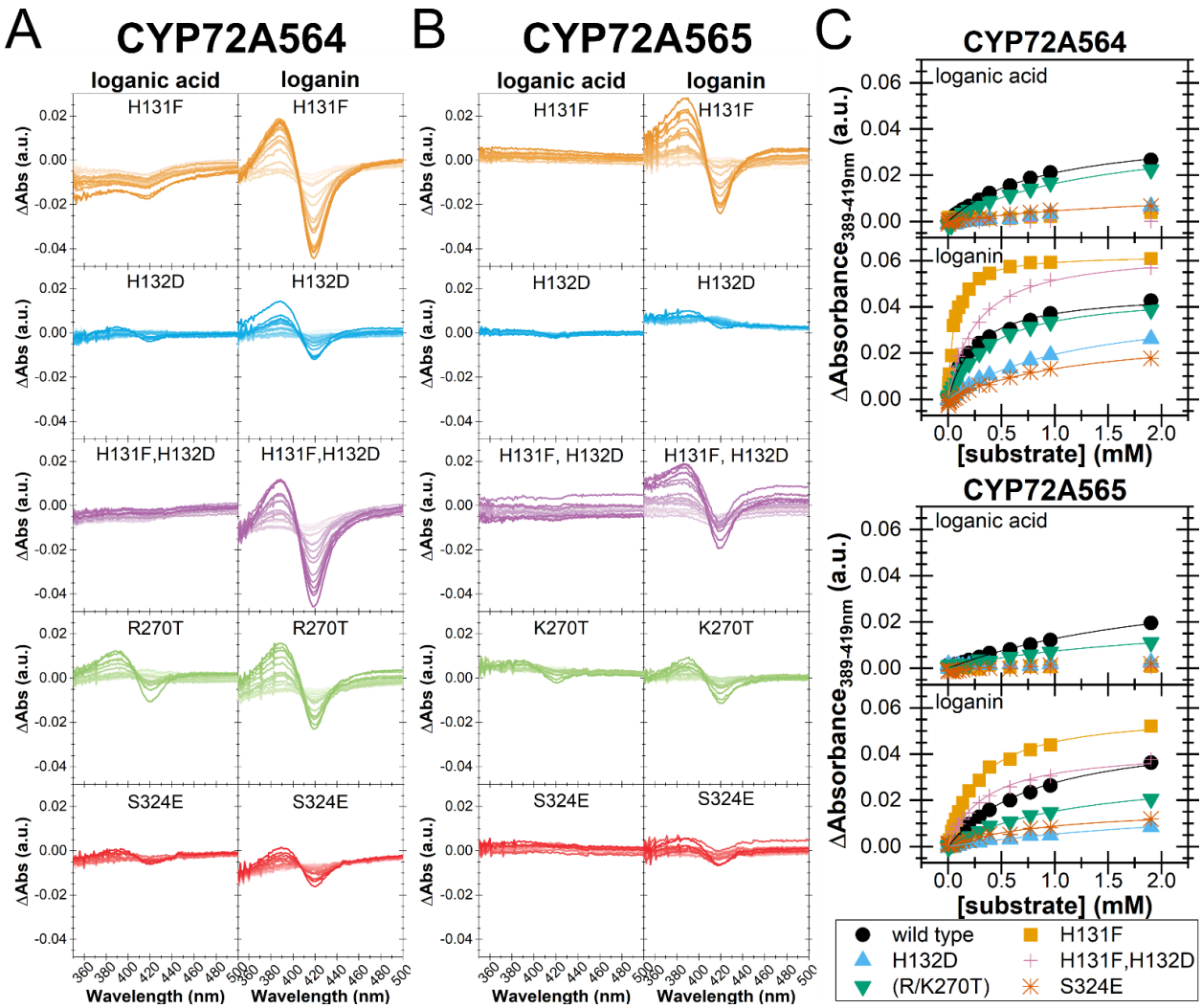


Figure 3.8: Type I binding spectra of *Camptotheca* CYP72A564 and CYP72A565 mutants. Difference spectra from 998 nM (lighter color) to 1.90 mM (darker color) loganic acid or loganin for CYP72A564 (A) and CYP72A565 (B) mutants. (C) Binding isotherms derived by plotting the difference between the peak (389 nm) and trough (419 nm) of the difference spectra against the substrate concentration using OriginPro 2019. Table 3.3 records the fit parameters.

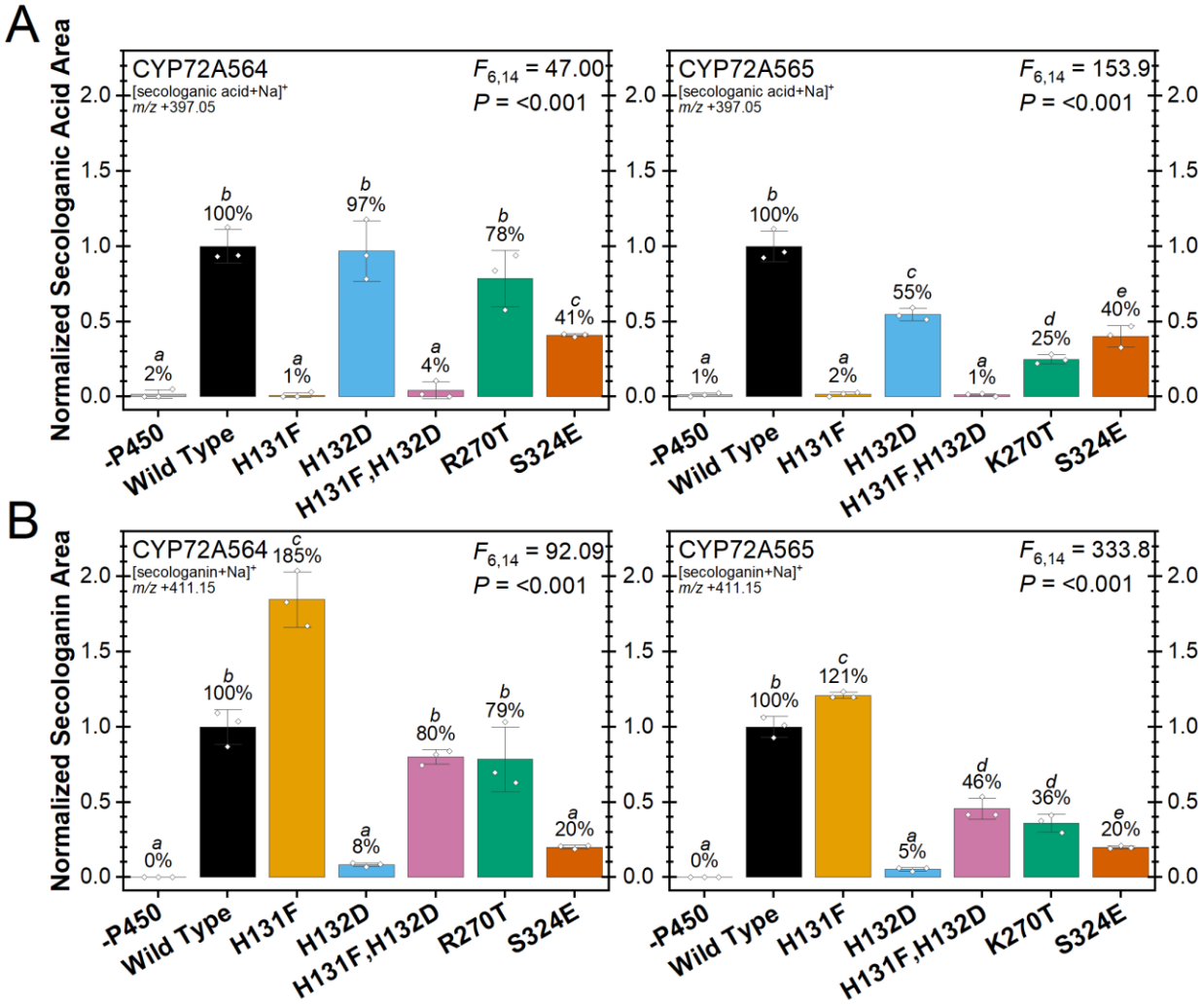


Figure 3.9: Secologanic acid, secologanin produced by *Camptotheca* CYP72A564 and CYP72A565 mutants.

Triplicate *in vitro* reactions of the listed CYP mutants with *Camptotheca* CPR1 were assayed for secologanic acid produced from loganic acid (A) and secologanin produced from loganin (B) by LC-MS normalized to an internal standard. The percentages of wild type activity (bars ± standard deviation with ◊ denoting individual replicates) were compared by a one-way ANOVA with Tukey's HSD test used for *post hoc* analysis (α = 0.05). Figure 3.10 displays the extracted ion chromatograms used for quantitation. Tables 3.10-3.13 contain the output from the pairwise comparisons.

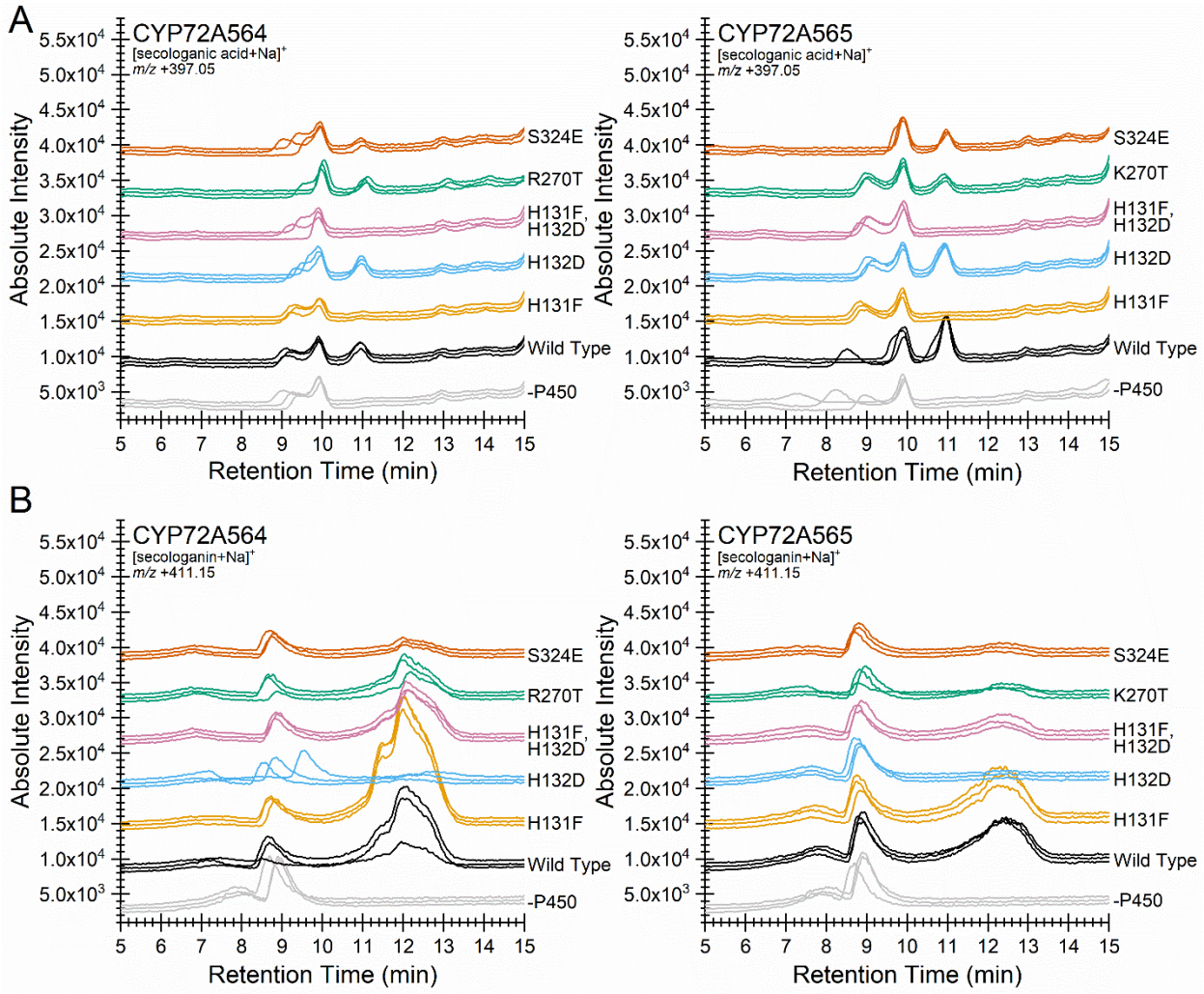


Figure 3.10. Extracted ion chromatograms of secoiridoids from *in vitro* reactions. Selective ion monitoring at *m/z* +397.05 $[secologanic\ acid+Na]^+$ (A) and *m/z* +411.15 $[secologanin+Na]^+$ (B) were used to assay secoiridoid production for CYP72A564 (left) and CYP72A565 (right) mutants. Reactions included 250. nM P450, 0.5 mU μL^{-1} *Camptotheca* CPR1, 1 mM DLPC, 250. µM substrate, 500 µM NADPH, 1.00 mM glucose-6-phosphate, and 0.1 U μL^{-1} G6PDH in 100. mM $NaPO_4$ (pH 7.5). Retention times: loganic acid, 10.0 min; secologanic acid, 11.0 min; loganin, 8.8 min; secologanin, 12.5 min.

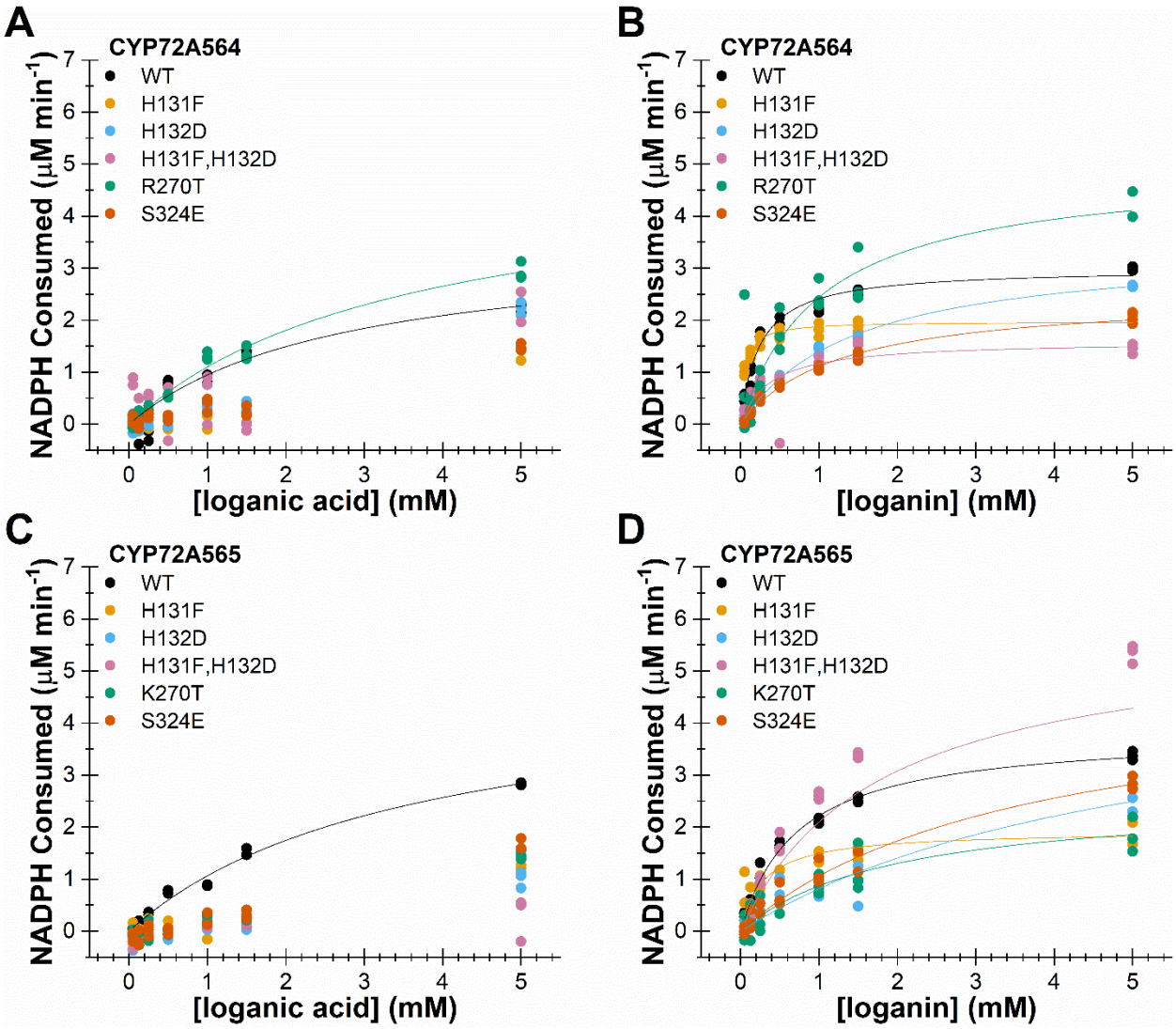


Figure 3.11. Steady-state kinetics of loganic acid and loganin turnover for all CYP72A564 and CYP72A565 mutants.

3.6. REFERENCES

1. Leonard, E., Runguphan, W., O'Connor, S., and Prather, K. J. (2009) Opportunities in metabolic engineering to facilitate scalable alkaloid production. *Nat. Chem. Biol.* **5**, 292–300
2. Glenn, W. S., Runguphan, W., and O'Connor, S. E. (2013) Recent progress in the metabolic engineering of alkaloids in plant systems. *Curr. Opin. Biotechnol.* **24**, 354–365
3. Miettinen, K., Dong, L., Navrot, N., Schneider, T., Burlat, V., Pollier, J., Woittiez, L., van der Krol, S., Lugan, R., Ilc, T., Verpoorte, R., Oksman-Caldentey, K.-M., Martinoia, E., Bouwmeester, H., Goossens, A., Memelink, J., and Werck-Reichhart, D. (2014) The seco-iridoid pathway from *Catharanthus roseus*. *Nat. Commun.* **5**, 3606
4. Irmler, S., Schröder, G., St-Pierre, B., Crouch, N. P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J. (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: New enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797–804
5. Dugé de Bernonville, T., Foureau, E., Parage, C., Lanoue, A., Clastre, M., Londono, M. A., Oudin, A., Houillé, B., Papon, N., Besseau, S., Glévarec, G., Atehortùa, L., Giglioli-Guivarc'h, N., St-Pierre, B., De Luca, V., O'Connor, S. E., and Courdavault, V. (2015) Characterization of a second secologanin synthase isoform producing both secologanin and secoxyloganin allows enhanced *de novo* assembly of a *Catharanthus roseus* transcriptome. *BMC Genomics*. **16**, 619
6. Geu-Flores, F., Sherden, N. H., Courdavault, V., Burlat, V., Glenn, W. S., Wu, C., Nims, E., Cui, Y., and O'Connor, S. E. (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis. *Nature*. **492**, 138–142
7. Salim, V., Yu, F., Altarejos, J., and De Luca, V. (2013) Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J.* **76**, 754–765
8. Asada, K., Salim, V., Masada-Atsumi, S., Edmunds, E., Nagatoshi, M., Terasaka, K., Mizukami, H., and De Luca, V. (2013) A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle. *Plant Cell*. **25**, 4123–4134
9. de Waal, A., Meijer, A. H., and Verpoorte, R. (1995) Strictosidine synthase from *Catharanthus roseus*: Purification and characterization of multiple forms. *Biochem. J.* **306**, 571–580
10. De Luca, V., Salim, V., Thamm, A., Masada, S. A., and Yu, F. (2014) Making iridoids/secoiridoids and monoterpenoid indole alkaloids: Progress on pathway elucidation. *Curr. Opin. Plant Biol.* **19**, 35–42
11. De Luca, V., Salim, V., Levac, D., Atsumi, S. M., and Yu, F. (2012) Discovery and functional analysis of monoterpenoid indole alkaloid pathways in plants. in *Methods in Enzymology* (Hopwood, D. A. ed), pp. 207–229, Natural Product Biosynthesis by Microorganisms and Plants, Part A, Academic Press, **515**, 207–229
12. Murata, J., Roepke, J., Gordon, H., and Luca, V. D. (2008) The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell*. **20**, 524–542

13. Collu, G., Unver, N., Peltenburg-Looman, A. M. G., van der Heijden, R., Verpoorte, R., and Memelink, J. (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis. *FEBS Lett.* **508**, 215–220
14. De Luca, V., Marineau, C., and Brisson, N. (1989) Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: Comparison with animal dopa decarboxylases. *Proc. Natl. Acad. Sci.* **86**, 2582–2586
15. Brown, S., Clastre, M., Courdavault, V., and O'Connor, S. E. (2015) De novo production of the plant-derived alkaloid strictosidine in yeast. *Proc. Natl. Acad. Sci.* **112**, 3205–3210
16. Yamamoto, H., Katano, N., Ooi, A., and Inoue, K. (2000) Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. *Phytochemistry*. **53**, 7–12
17. Rather, G. A., Sharma, A., Misra, P., Kumar, A., Kaul, V., and Lattoo, S. K. (2020) Molecular characterization and overexpression analyses of secologanin synthase to understand the regulation of camptothecin biosynthesis in *Nothapodytes nimmoniana* (Graham.) Mabb. *Protoplasma*. **257**, 391–405
18. Yang, M., Wang, Q., Liu, Y., Hao, X., Wang, C., Liang, Y., Chen, J., Xiao, Y., and Kai, G. (2021) Divergent camptothecin biosynthetic pathway in *Ophiorrhiza pumila*. *BMC Biol.* **19**, 122
19. Yang, Y., Li, W., Pang, J., Jiang, L., Qu, X., Pu, X., Zhang, G., and Luo, Y. (2019) Bifunctional cytochrome P450 enzymes involved in camptothecin biosynthesis. *ACS Chem. Biol.* **14**, 1091–1096
20. Miller, J. C., Hollatz, A. J., and Schuler, M. A. (2021) P450 variations bifurcate the early terpene indole alkaloid pathway in *Catharanthus roseus* and *Camptotheca acuminata*. *Phytochemistry*. **183**, 112626
21. Prall, W., Hendy, O., and Thornton, L. E. (2016) Utility of a phylogenetic perspective in structural analysis of CYP72A enzymes from flowering plants. *PLOS ONE*. **11**, e0163024
22. Zhang, C., Zhang, T., Luebert, F., Xiang, Y., Huang, C.-H., Hu, Y., Rees, M., Frohlich, M. W., Qi, J., Weigend, M., and Ma, H. (2020) Asterid phylogenomics/phylotranscriptomics uncover morphological evolutionary histories and support phylogenetic placement for numerous whole-genome duplications. *Mol. Biol. Evol.* **37**, 3188–3210
23. Rodríguez-López, C. E., Hong, B., Paetz, C., Nakamura, Y., Koudounas, K., Passeri, V., Baldoni, L., Alagna, F., Calderini, O., and O'Connor, S. E. (2021) Two bi-functional cytochrome P450 CYP72 enzymes from olive (*Olea europaea*) catalyze the oxidative C-C bond cleavage in the biosynthesis of secoxy-iridoids – flavor and quality determinants in olive oil. *New Phytol.* **229**, 2288–2301
24. Colthart, A. M., Tietz, D. R., Ni, Y., Friedman, J. L., Dang, M., and Pochapsky, T. C. (2016) Detection of substrate-dependent conformational changes in the P450 fold by nuclear magnetic resonance. *Sci. Rep.* **6**, 1–11
25. Tietz, D. R., Colthart, A. M., Pochapsky, S. S., and Pochapsky, T. C. (2017) Substrate recognition by two different P450s: Evidence for conserved roles in a common fold. *Sci. Rep.* **7**, 1–9
26. Katano, N., Yamamoto, H., Iio, R., and Inoue, K. (2001) 7-Deoxyloganin 7-hydroxylase in *Lonicera japonica* cell cultures. *Phytochemistry*. **58**, 53–58

27. Dunn, A. R., Hays, A.-M. A., Goodin, D. B., Stout, C. D., Chiu, R., Winkler, J. R., and Gray, H. B. (2002) Fluorescent probes for cytochrome P450 structural characterization and inhibitor screening. *J. Am. Chem. Soc.* **124**, 10254–10255
28. Fisher, M. T., and Sligar, S. G. (1985) Control of heme protein redox potential and reduction rate: linear free energy relation between potential and ferric spin state equilibrium. *J. Am. Chem. Soc.* **107**, 5018–5019
29. Denisov, I. G., Makris, T. M., Sligar, S. G., and Schlichting, I. (2005) Structure and chemistry of cytochrome P450. *Chem. Rev.* **105**, 2253–2278
30. Denisov, I. G., and Sligar, S. G. (2015) Activation of molecular oxygen in cytochromes P450. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed), pp. 69–109, Springer International Publishing, Cham, 10.1007/978-3-319-12108-6_3
31. Madyastha, K. M., Guarnaccia, R., Baxter, C., and Coscia, C. J. (1973) *S*-Adenosyl-L-methionine: loganic acid methyltransferase: A carboxyl-alkylating enzyme from *Vinca rosea*. *J. Biol. Chem.* **248**, 2497–2501
32. Stöckigt, J., Antonchick, A. P., Wu, F., and Waldmann, H. (2011) The Pictet–Spengler Reaction in Nature and in Organic Chemistry. *Angew. Chem. Int. Ed.* **50**, 8538–8564
33. Sadre, R., Magallanes-Lundback, M., Pradhan, S., Salim, V., Mesberg, A., Jones, A. D., and DellaPenna, D. (2016) Metabolite diversity in alkaloid biosynthesis: A multilane (diastereomer) highway for camptothecin synthesis in *Camptotheca acuminata*. *Plant Cell*. **28**, 1926–1944
34. Caputi, L., Franke, J., Farrow, S. C., Chung, K., Payne, R. M. E., Nguyen, T.-D., Dang, T.-T. T., Carqueijeiro, I. S. T., Koudounas, K., Bernonville, T. D. de, Ameyaw, B., Jones, D. M., Vieira, I. J. C., Courdavault, V., and O'Connor, S. E. (2018) Missing enzymes in the biosynthesis of the anticancer drug vinblastine in Madagascar periwinkle. *Science*. **360**, 1235–1239
35. Qu, Y., Safonova, O., and De Luca, V. (2019) Completion of the canonical pathway for assembly of anticancer drugs vincristine/vinblastine in *Catharanthus roseus*. *Plant J.* **97**, 257–266
36. Wu, F., Kerčmar, P., Zhang, C., and Stöckigt, J. (2016) Sarpagan-ajmalan-type indoles: Biosynthesis, structural biology, and chemo-enzymatic significance. in *The Alkaloids: Chemistry and Biology* (Knölker, H.-J. ed), pp. 1–61, Academic Press, **76**, 1–61
37. Kang, M., Fu, R., Zhang, P., Lou, S., Yang, X., Chen, Y., Ma, T., Zhang, Y., Xi, Z., and Liu, J. (2021) A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* **12**, 3531
38. Rai, A., Hirakawa, H., Nakabayashi, R., Kikuchi, S., Hayashi, K., Rai, M., Tsugawa, H., Nakaya, T., Mori, T., Nagasaki, H., Fukushi, R., Kusuya, Y., Takahashi, H., Uchiyama, H., Toyoda, A., Hikosaka, S., Goto, E., Saito, K., and Yamazaki, M. (2021) Chromosome-level genome assembly of *Ophiorrhiza pumila* reveals the evolution of camptothecin biosynthesis. *Nat. Commun.* **12**, 405
39. Bart, A. G., Harris, K. L., Gillam, E. M. J., and Scott, E. E. (2020) Structure of an ancestral mammalian family 1B1 cytochrome P450 with increased thermostability. *J. Biol. Chem.* **295**, 5640–5653
40. Gumulya, Y., Baek, J.-M., Wun, S.-J., Thomson, R. E. S., Harris, K. L., Hunter, D. J. B., Behrendorff, J. B. Y. H., Kulig, J., Zheng, S., Wu, X., Wu, B., Stok, J. E., De Voss, J. J.,

Schenk, G., Jurva, U., Andersson, S., Isin, E. M., Bodén, M., Guddat, L., and Gillam, E. M. J. (2018) Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888

41. Gumulya, Y., Huang, W., D'Cunha, S. A., Richards, K. E., Thomson, R. E. S., Hunter, D. J. B., Baek, J.-M., Harris, K. L., Boden, M., De Voss, J. J., Hayes, M. A., Isin, E. M., Andersson, S., Jurva, U., and Gillam, E. M. J. (2019) Engineering thermostable CYP2D enzymes for biocatalysis using Combinatorial Libraries of Ancestors for Directed Evolution (CLADE). *ChemCatChem*. **11**, 841–850
42. Hartz, P., Strohmaier, S. J., EL-Gayar, B. M., Abdulmughni, A., Hutter, M. C., Hannemann, F., Gillam, E. M. J., and Bernhardt, R. (2021) Resurrection and characterization of ancestral CYP11A1 enzymes. *FEBS J.* **288**, 6510–6527
43. Gegner, J. A., and Dahlquist, F. W. (1991) Signal transduction in bacteria: CheW forms a reversible complex with the protein kinase CheA. *Proc. Natl. Acad. Sci.* **88**, 750–754
44. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549
45. Edgar, R. C. (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics*. **5**, 113
46. Le, S. Q., and Gascuel, O. (2008) An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **25**, 1307–1320
47. Rupasinghe, S., and Schuler, M. A. (2006) Homology modeling of plant cytochrome P450s. *Phytochem. Rev.* **5**, 473–505
48. Rowland, P., Blaney, F. E., Smyth, M. G., Jones, J. J., Leydon, V. R., Oxbrow, A. K., Lewis, C. J., Tennant, M. G., Modi, S., Eggleston, D. S., Chenery, R. J., and Bridges, A. M. (2006) Crystal structure of human cytochrome P450 2D6. *J. Biol. Chem.* **281**, 7614–7622
49. Jin, Z., Cong, Y., Zhu, S., Xing, R., Zhang, D., Yao, X., Wan, R., Wang, Y., and Yu, F. (2019) Two classes of cytochrome P450 reductase genes and their divergent functions in *Camptotheca acuminata* Decne. *Int. J. Biol. Macromol.* **138**, 1098–1108
50. Guengerich, F. P., Martin, M. V., Sohl, C. D., and Cheng, Q. (2009) Measurement of cytochrome P450 and NADPH–cytochrome P450 reductase. *Nat. Protoc.* **4**, 1245–1251

CHAPTER 4: THREE AMINO ACID DELETION IN *CAMPTOTHECA* CYP72A730 FAILS TO EXPLAIN LOSS OF SLS, SLAS ACTIVITY

Developing new enzyme functions is an important part of evolution as organisms adapt to changing environments. Often facilitated by gene duplication, one coding sequence continues in its original role(s) while another copy diverges from its initial activity *en route* to developing new and, possibly, losing its original capabilities. A whole-genome duplication in *Camptotheca acuminata* has deposited multiple copies of most genes marked for involvement in camptothecin biosynthesis, including two secologanic acid synthases (SLASs; CYP72A564 and CYP72A565). CYP72A730, apparently paralogous to CYP72A564, was postulated to be another SLAS candidate, but biochemical analyses have demonstrated it lacked this enzymatic activity. Curious as to which changes rendered CYP72A730 inactive as a secologanic acid synthase (SLAS), I investigated whether a three amino acid deletion in this protein's F-G loop was deleterious. Reciprocal site-directed mutagenesis on CYP72A564 and CYP72A730, spectral binding assays, and *in vitro* turnover of loganic acid and loganin showed that neither deleting the amino acids from CYP72A564 nor inserting them into CYP72A730 significantly affected the activity (or inactivity) of either protein. Comparing CYP72A730 first to the *Camptotheca* SLASs then to secologanin synthases (SLSs) from other species, I subsequently identified eleven residues conserved amongst the active enzymes (SLASs and SLSs) but drastically different in CYP72A730. Half of these involve changes to or from Gly and Pro, suggesting that the CYP72A730 fold may contain disruptions in secondary structure compared to its closest homologues. Additional experiments combining *Camptotheca* cell culture, virus-induced gene silencing, and metabolomics are likely needed to unveil the *in vivo* activity of CYP72A730, should it have any.

4.1. INTRODUCTION

The development of new enzyme functions is an important aspect of adapting to changing environments. For plants that frequently utilize specialized metabolites to fend off microbial invaders and herbivores, the development of detoxification mechanisms in these other species demands that plants produce new metabolites to which resistance is not widespread. This requirement for new compounds presents competing demands to maintain the activities that yield specialized metabolites currently produced *and* acquire new activities yielding new molecules.

Through the course of gene (or genome) duplications, the presence of multiple copies of the same gene enables one to maintain the current activity and the other to mutate and evolve other activities, a process often labeled neofunctionalization (1, 2). The derived protein in the course of mutating frequently loses its ancestral activity before it has developed a new catalytic role, producing a pseudoenzyme that is expressed by an organism but lacks catalytic activity (1, 2).

My search for the secologanic acid synthases (SLASs) of *Camptotheca acuminata* identified two catalytically-active SLASs and a third, highly homologous protein (CYP72A730) lacking either of these activities (3). The inactivity of CYP72A730 is particularly puzzling in light of its high mRNA expression patterns in tissues with the greatest amount of terpene indole alkaloid (TIA) biosynthesis. A whole-genome duplication in *Camptotheca* has produced multiple copies of the biosynthetic genes that produce camptothecin (4), including the loci encoding these CYP72As. Together, these data suggest that CYP72A730 is a pseudoenzyme evolving from a common SLAS ancestor.

From this hypothesis, I desired to investigate which changes from the common CYP72A564 and CYP72A730 ancestor (cf. Figure 3.3) have led to the former's retention of

7DLH, SLS, and SLAS activity (3, 5) and the latter's apparent inactivity (3). I proposed that the three amino acid deletion between substrate recognition sequence 3 (SRS3) and SRS4 in CYP72A730 (Fig. 4.1) were the main cause for its loss of SLS and SLAS activity. A combination of site-directed mutagenesis and *in vitro* analyses with the purified mutant proteins revealed that this small deletion is not responsible for the differences in activity among these CYP72As.

4.2. RESULTS AND DISCUSSION

To evaluate the importance of the three amino acid deletion found in CYP72A730 compared to the SLSs and SLASs (Fig. 4.1; Appendix E), I mutated CYP72A730 to include the three missing amino acids (CYP72A730ins) as outlined in the Section 4.3.1. Reciprocally removing these three amino acids from CYP72A564 resulted in a mutated SLAS resembling the backbone structure of CYP72A730 (CYP72A564del). These two mutant proteins, after expression in *E. coli* and purification, properly folded to yield strong P450 peaks in reduced-CO difference spectra (Fig. 4.2).

4.2.1. Insertion and deletion mutations slightly perturb ligand binding

To evaluate the importance of the deletion present in CYP72A730, I subsequently compared the substrate-induced binding spectra and *in vitro* enzyme assays of these mutated proteins and the wild type CYPs. As with the wild type protein, additions of loganic acid or loganin to CYP72A564del induced Type I spectra (Fig. 4.3). Deleting these three amino acids decreased affinity for both of these iridoids by approximately 50% (*i.e.*, the K_s increased by about 50%; Table 4.2). These small changes in binding affinity suggested that CYP72A564del would exhibit only small losses in secoiridoid production, if any.

In contrast, additions of loganic acid and loganin to CYP72A730ins had minimal effects on substrate binding when probed for Type I binding spectra. Like the wild type CYP72A730, no spin shift in the heme Soret was observed for CYP72A730ins at loganic acid concentrations of up to 1.90 mM (Fig. 4.3). At the higher concentrations of loganin studied (0.3-1.9 mM), CYP72A730ins showed small amounts of spin-shift in the heme Soret consistent with water displacement from the oxidized heme iron. The minimal levels of spin-shift with this substrate and the sampling of concentrations below the apparent K_s prevented accurate estimation of loganin binding parameters (Table 4.2). Based on these findings, although I predicted secologanic acid production from neither CYP72A730 nor CYP72A730ins, the insertion of three amino acids might allow CYP72A730ins to produce small amounts of secologanin.

4.2.2. Insertion, deletion mutations do not affect secoiridoid production

In vitro reconstitution of these CYP72A wild type and mutant proteins with full-length *Camptotheca* CPR1, DLPC, 250 μM substrate, and an NADPH regenerating system was used to determine whether the insertion and deletion mutations affected SLAS and SLS activity. Deleting the three amino acids absent in CYP72A730 from CYP72A564 (*i.e.*, CYP72A564del) did not greatly affect the amount of secologanic acid or secologanin produced (Fig. 4.4, 4.5). Adding these same amino acids to CYP72A730 (CYP72A730ins) did not allow the protein to produce either secologanic acid or secologanin. This inability to produce significant amounts of secologanin from 250 μM loganin contrasts my earlier predictions rooted in the Type I binding spectra observed around this concentration and above for loganin.

4.2.3. Alternative explanations for inactivity of CYP72A730

Whether the three amino acids present in the *Camptotheca* SLASs but absent in CYP72A730 were added (CYP72A730ins) or removed (CYP72A564del), changes in the mutant

proteins' activities towards loganic acid and loganin were largely nonexistent. The minimal induction of Type I binding observed for CYP72A730ins at the highest concentrations of loganin did not correspond to detectable secologanin production. Similarly, the loss in binding affinity of CYP72A564del did not lead to major changes in the production of either secologanic acid or secologanin compared to the wild type CYP72A564.

The three amino acid deletion present in CYP72A730 compared to CYP72A564 and CYP72A565 is an insufficient explanation of the former's inability to turnover loganic acid or loganin as witnessed by the results reported herein after site-directed mutagenesis. Moreover, the lack of strong losses of activity for CYP72A564del suggest that this deletion does not contribute to the SLAS inactivity of CYP72A730. Consequently, one or (more likely) some combination of the 39 different residues present in CYP72A730 conserved in the *Camptotheca* SLASs (Fig. 4.1) are responsible for creating this pseudozyme.

By comparing the identity of SLSs at these sites with CYP72A730 and the *Camptotheca* SLASs (Table 4.5; see also Appendix E), eleven of the differences appear stark; six present changes in biochemical features that might influence activity; twelve others are probably not important differences; and ten are variable sites amongst these proteins. The eleven sites with starkly different residue properties include changes in charge (Tyr117, Glu176, Arg221, Glu224, Gly270, Thr345, Arg460), size (Pro164, Glu224, Gly270, Ser325, Thr345), hydrophobicity (Tyr117, Glu176, Arg221, Glu224, Ser325, Arg460), and backbone flexibility (Pro164, Ala201, Arg221, Glu224, Gly270, Ser347).[1] This assortment of the most drastic changes (with the exception of Ser325 and perhaps Gly270) are outside of the predicted substrate recognition

[1] All numbered according to CYP72A730.

sequences (SRSs), rendering predictions about their effects on activity difficult. The locations where CYP72A730 contains Gly or Pro (Pro164, Gly270), lacks a Gly (Arg 221, Glu224), or lacks a Pro (Ala201, Ser347) are particularly intriguing given their large number and the propensity of Gly and Pro to disrupt secondary structures.

4.2.4. Avenues to discern in vivo activity of CYP72A730

The intractability of CYP72A730 to restoring an ancestral activity shared with CYP72A564 and CYP72A565 further highlights the pseudozyme nature of this protein. Whether it is enzymatically active in any capacity within *Camptotheca* remains unknown. The genomic locus encoding this protein demonstrates mRNA expression across tissue types, especially in immature and mature fruit (see Fig. 2.2), supporting a role in the specialized metabolism known to occur in these tissues. Moreover, the energy required to produce mature mRNA and presumably a polypeptide from it makes it unlikely for this protein to lack a biological function.

Given the pattern of expression for CYP72A730, the biological function of this protein is likely involved in another part of camptothecin production. Combining *Camptotheca* cell culture with virus-induced gene silencing (VIGS) targeting *cyp72a730* is a straightforward means of assessing this *in vivo* activity when combined with metabolomics to assess which metabolites are affected by knockdown. Such a method was important in identifying which P450s modified nepetalactol to form 7-deoxyloganetic acid (6) and 7-hydroxylated 7-deoxyloganin to form loganic acid (7) in *Catharanthus roseus*. VIGS is amenable to *Camptotheca* seedlings (8), and cell culturing with a variety of different cell types (9–12) from *Camptotheca* is known. Coalescing these existent cell culture, VIGS, and metabolomics technologies provides a means of assaying CYP72A730 function *and* testing genes downstream of strictosidinic acid production

in camptothecin biosynthesis for which there are no known homologues in other organisms with quality throughput.

4.3. EXPERIMENTAL

4.3.1. Site-directed mutagenesis

Site-directed mutagenesis was conducted using the primers given in Table 4.1 using the three-step PCR procedure outlined in Chapter 3. Annealing temperatures for the first and third PCR step were set to 55°C instead of 69°C.

4.3.2. Protein expression, purification, and characterization

All proteins (CYPs and CPRs) were expressed and purified using the procedures given in Chapter 2 with the adaptations noted in Chapter 3.

4.3.3. In vitro assays

Substrate-induced binding spectra were conducted as noted in Chapter 2. *In vitro* reconstitutions assayed by LC-MS followed the procedures given in Chapter 3 with a two-hour long incubation time at 30°C before quenching.

4.4. TABLES AND FIGURES

Table 4.1. Primers used in this study.

CYP Gene	Primer name	Sequence (5' to 3')
72A564	CaaSLS pCW NdeI 5'	ATTACAT**ATG**GAAATTCAAATGGACG TTCTGTAC
72A564	CaaSLS364 pCW His6 XbaI 3'	TAATTCTAGATCAGTGGTGGTGGTGGT GGTGACATTTGAGCCTGCGGAAGATC
72A564del	CaaSLS364_del_f1	caacagcagaATGAAAGCAGGAGAGCCTTC TGG
72A564del	CaaSLS364_del_r1	ctgctttcatTCTGCTGTTGATGATACCTGAC AACC
72A730	CaaSLS364i2 pCW NdeI 5'	ATTACAT**ATG**GAAATGGAGCTAATGT ACAAGTC
72A730	CaaSLS364i2 pCW His6 XbaI 3'	TAATTCTAGATCAGTGGTGGTGGTGGT GGTGATATTTGAGTTTGCGGAAGATCA CTTG
72A730ins	CaaSLS364i2_ins_f1	gtgaaggcaATGAAGACAGGAGAGCCTTT TGG
72A730ins	CaaSLS364i2_ins_r1	tgccttcacTCTACTATTTATGATGTTGGAC AATCCCTC

Table 4.2. Non-linear curve fitting parameters from Type I binding spectral titrations.
Parameters ± S.E. estimated from non-linear curve fit of spectral response against varying substrate concentrations (*N*=17)

	loganic acid		**loganin**	
CYP	K_s	ΔAbs_{max}	K_s	ΔAbs_{max}
	μM	*mAU*	*μM*	*mAU*
72A564	874 ± 57	39.5 ± 1.3	256 ± 19	46.6 ± 1.2
72A564del	1.18 ± 0.14	60.9 ± 5.7	375 ± 47	67.8 ± 3.4
72A730	n.d.[a]	n.d.[a]	n.d.[a]	n.d.[a]
72A730ins	n.d.[a]	n.d.[a]	3,500 ± 2,200[b]	21 ± 10.

[a] n.d., unable to fit data
[b] The apparent K_s is above the highest concentration (1.90 mM) used rendering the nonlinear curve fitting inaccurate.

Table 4.3. Tukey's HSD pairwise comparison of secologanic acid production.

Comparison	Difference in Means	SEM	q Value	Probability	Alpha	Significant?
72A564 WT -P450	0.9911	0.0422	33.22	< 1E-9	0.05	Yes
72A564 del -P450	0.8802	0.0422	29.50	< 1E-9	0.05	Yes
72A564 del 72A564 WT	-0.1109	0.0422	3.717	0.138	0.05	No
72A730 WT -P450	0.02620	0.0422	0.8781	0.968	0.05	No
72A730 WT 72A564 WT	-0.9649	0.0422	32.34	< 1E-9	0.05	Yes
72A730 WT 72A564 del	-0.8540	0.0422	28.62	< 1E-9	0.05	Yes
72A730 ins -P450	0.01351	0.0422	0.4529	0.997	0.05	No
72A730 ins 72A564 WT	-0.9776	0.0422	32.76	< 1E-9	0.05	Yes
72A730 ins 72A564 del	-0.8667	0.0422	29.05	< 1E-9	0.05	Yes
72A730 ins 72A730 WT	-0.01269	0.0422	0.4252	0.998	0.05	No

Table 4.4. Tukey's HSD pairwise comparison of secologanin production.

Comparison	Difference in Means	SEM	q Value	Probability	Alpha	Significant?
72A564 WT -P450	0.9936	0.04883	28.78	< 1E-9	0.05	Yes
72A564 del -P450	1.0791	0.04883	31.25	< 1E-9	0.05	Yes
72A564 del 72A564 WT	0.08553	0.04883	2.477	0.448	0.05	No
72A730 WT -P450	0.00656	0.04883	0.1900	1.000	0.05	No
72A730 WT 72A564 WT	-0.9870	0.04883	28.59	< 1E-9	0.05	Yes
72A730 WT 72A564 del	-1.0726	0.04883	31.06	< 1E-9	0.05	Yes
72A730 ins -P450	0.00355	0.04883	0.1029	1.000	0.05	No
72A730 ins 72A564 WT	-0.9900	0.04883	28.67	< 1E-9	0.05	Yes
72A730 ins 72A564 del	-1.0756	0.04883	31.15	< 1E-9	0.05	Yes
72A730 ins 72A730 WT	-0.00301	0.04883	0.08708	1.000	0.05	No

Table 4.5. Divergent residues of *Camptotheca* CYP72A730.

Camptotheca CYP72A730 Residue	*Camptotheca* CYP72A730 Number	*Camptotheca* SLAS Residue	*Camptotheca* SLAS Number	Comparison to All SLSs[a]
Met	101	Ile	103	all but one Met (other Asp)
Gln	111	Glu	113	all charged (all but one Asp/Glu)
Tyr	117	Asp	119	all Glu
Val	120	Ala	122	all but one Thr (other Phe)
Arg	126	Gln/Met	128	largely Gln (two Val)
Ser	132	Tyr	134	largely His (one Gln)
Pro	164	Leu	166	all Leu
Glu	176	Val	178	all hydrophobic of some kind
Val	199	Leu	201	all Val or Ile
Tyr	200	Phe	202	all Phe
Ala	201	Pro	203	all Pro
Val	206	Leu	208	all Leu
Ile	214	Val	216	all Val
Arg	221	Gly	223	all Asp/Glu
Glu	224	Gly	226	all Gly
Ile	234	Met	236	all Met
Phe	252	Leu	254	all but one Leu (other Val)
Thr	256	Arg	258	all but one Arg (other Thr)
Ile	260	Met	262	all but one Met (other Ile)
Lys	262	Glu	264	mostly Glu (Ala, Asn, Gln)
Gln	268	Arg/Lys	270	Thr, Lys, or Arg
Gly	270	Arg	272	all Met
Asn	273	Gly	275	highly variable (Asn common)
Thr	281	Ala	286	all but one Ala (other Ser)
Phe	285	Ser	290	highly variable
Leu	298	Phe	303	some larger hydrophobic residue
Ser	299	Arg/Lys	304	longer hydrophilic (Gln common)
Asp	315	Glu	320	mostly Asn (Asp, His also present)
Val	323	Leu	328	all Leu
Ser	325	Tyr	330	all Tyr
Val	327	Ala	332	all Ala
Thr	345	Arg	350	all Lys
Ser	347	Pro	352	all Pro
Arg	460	Trp	465	all Trp
Leu	466	Val	471	all but one Leu (other Ile)
Leu	499	Phe	504	larger hydrophobic residue (mostly Phe)
Ser	501	Val	506	Ile or Val
Val	510	Ala	515	all Ser
Tyr	519	Cys	524	Ser where that long

[a] See Appendix E for a full multiple sequence alignment.

```
CYP72A730   --MEMELMYKSITASLAVVFLVYFWRALNWAWFKPKMIEKRLRQQGFNGNPYRLLVGDLK  58
CYP72A564   MEIQMDVLYKSIAASVAVVFLVYAWKMLNWAYLTPKRIEKCLRKQGFKGNSYRLLVGDLK  60
CYP72A565   MEIQMDVLYKSIAASVAVVFLVYAWKMLNWAYLTPKRIEKCLRKQGFKGNSYRLLVGDLK  60
              ::*:::****:**:******* *: ****::.** *** **:***:** *********

CYP72A730   ESSMMLKETMSKPIPVSQDIVQRLMPHVVKTIETYGKNSFTWMGRMPRVHIMQPELIKYI  118
CYP72A564   ESSMMLKEAMSKPIPVSQDIVQRLMPHVVKTIQTYGKNSFTWIGRMPRVHIMEPELIKDI  120
CYP72A565   ESSMMLKETMSKPINVSEDIVQRVMPHVIKTIDTYGKNSFTWIGRMPRVHIMEPDLIKDI  120
            ********:***** **:*****:****:***:*********:*********:*:*** *
                        SRS1
CYP72A730   LVNHTDFRKNHHASNPMTKFLLTGIGSLEGDKWAKHRRIISPSFHPEKLKTMLPTFYESY  178
CYP72A564   LANHNNFQKNHHAYSPLTKFLLTGIGSLEGEKWAKHRRIISPSFHLEKLKTMLPAFYVSY  180
CYP72A565   LANHNDFMKNHHAYNPLTKFLLTGIGSLEGDKWAKHRRIISPSFHLEKLKTMLPAFYVSY  180
            *.**.:* ***** .*:**************:********** ********:** **
                                                                  SRS2
CYP72A730   DELLSKWERECSTKGSVEVDVYATFDTVTSDVISRIAFGSSYREGERIFILLKELIDLTV  238
CYP72A564   DELLGKWERESSTKGSVEVDLFPTFDTLTSDVISRVAFGSSYGEGGRIFILLKELMDLTV  240
CYP72A565   DDLLTKWEQQCSSKGSVEIDLFPTFDTLTSDVISRVAFGSSYGEGGRIFILLKELMDLTV  240
            *:** ***::.*:*****:*:: ****:*******:****** ** *********:****
                                  SRS3
CYP72A730   DVMRSVYVPGWSLFPSKTNQRIRKVDREIQEGLSNIINSR---MKTGEPFGNDLLGTLLE  295
CYP72A564   DVMRSVYVPGWSLLPTKRNQRMREVDREIRERLSGIINSRVKAMKAGEPSGDDLLGTLLE  300
CYP72A565   DVMRSVYVPGSSFLPTKRNNRMREVDGEIKDRLSGIINSRVKAMKAGEPSGEDLLGTLLE  300
            ********** *::*:* *:*:*:** **:: **.*****   **:*** *:********
                                  SRS4
CYP72A730   SNLSEIERLGNKKNAGMSIDDVISECKVFSFVGQETTGILLTWTCVLLSTHSEWQERARE  355
CYP72A564   SNFREIERLGNKKNAGMSIEDVISECKLFYFAGQETTGILLTWTCVLLSRHPEWQERARE  360
CYP72A565   SNFKEIERLGNKKNAGMSIEDVISECKLFYFAGQETTGILLTWTCVLLSRHPEWQERARE  360
            **: ***************:*******:* *.***************** * ********
                                                  SRS5
CYP72A730   EIFQVFGNGKLDFDRVQGLKIVPMILYEVLRLYPPVIELTKVTYEEQKLGNLTIPAGVQL  415
CYP72A564   EIFQVFGNGKLDFDRVQGLKIVPMILYEVLRLYPPVIELTKVTYEEQKLGNLTIPAGVQL  420
CYP72A565   EIFQVFGNGKVDFDRVQNLKIVPMILYEVLRLYPPVIELTKVTYEEQKLGNLTIPAGVQL  420
            **********:******.****************************************

CYP72A730   MMPSILLHRDKEMWGDDATEFNPGRFAEGVSKAVKSPFFFIPFSRGPRICLGQNFALLQA  475
CYP72A564   MMPSILLHRDKEMWGDDATEFNPGRFAEGVAKAVKSPFFYIPFSWGPRICVGQNFALLQA  480
CYP72A565   MMPSILLHRDQEMWGADSKEFNPGRFADGISKAVKSPFFYIPFSWGPRICVGQNFALLQA  480
            **********:**** *:.********:*::********:**** *****:*********
                                  SRS6
CYP72A730   KMALTMILQRFSFDLSPTYAHAPLTSLTLQPQHGVQVIFRKLKY  519
CYP72A564   KMALAMILQRFSFDLSPTYAHAPFTVLTLQPQHGAQVIFRRLKC  524
CYP72A565   KMALTMILQRFTFDLSPTYAHAPFTVLTLQPQHGAQVVFRKIKC  524
            ****:******:***********:* ********.**:**::*
```

Figure 4.1. Multiple sequence alignment of *Camptotheca* CYP72A564, 565, 730.
Differences in CYP72A730 compared to conserved residues in the *Camptotheca* SLASs CYP72A564 and CYP72A565 are highlighted: three amino acid deletion, green; CYP72A730 amino acids with comparable biochemical properties, cyan; CYP72A730 amino acids with vastly different biochemical properties, yellow. SRS, substrate recognition sequence.

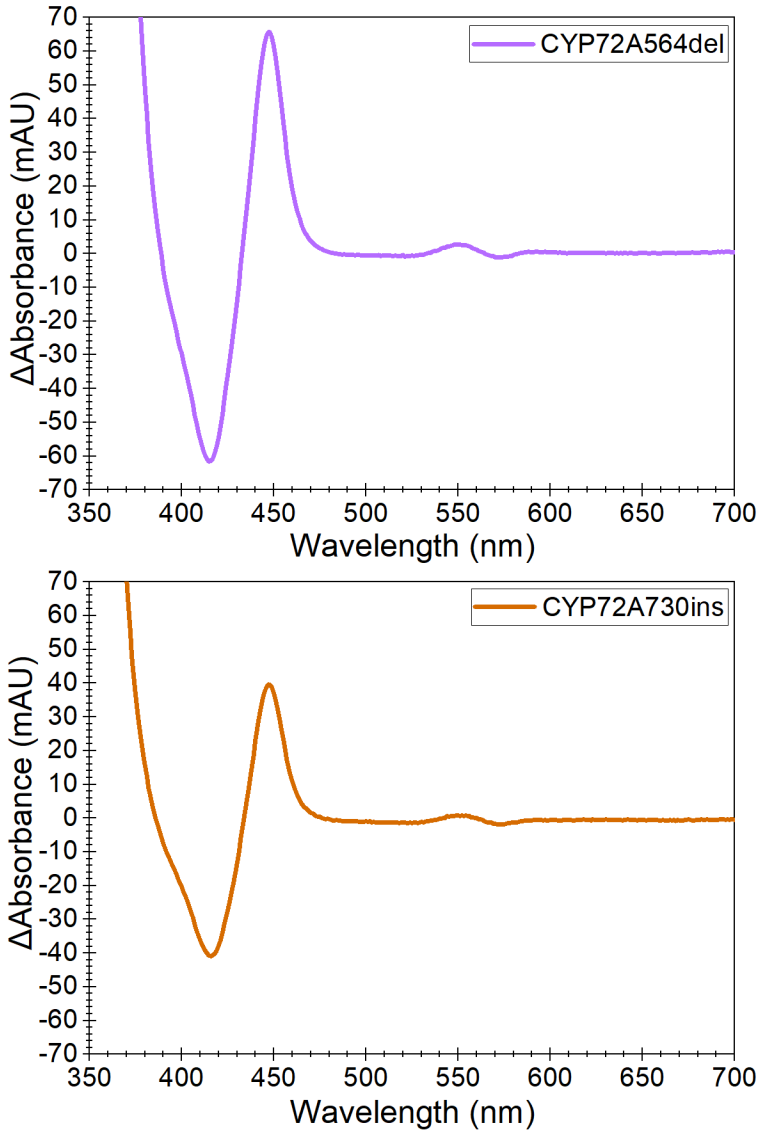


Figure 4.2. Reduced-CO difference spectra.
CO-reduced minus CO-oxidized spectra for 1-to-25 diluted CYP72A564del (top) and 1-to-25 diluted CYP72A730ins (bottom) proteins in 100 mM $NaPO_4$ (pH 7.5) buffer.

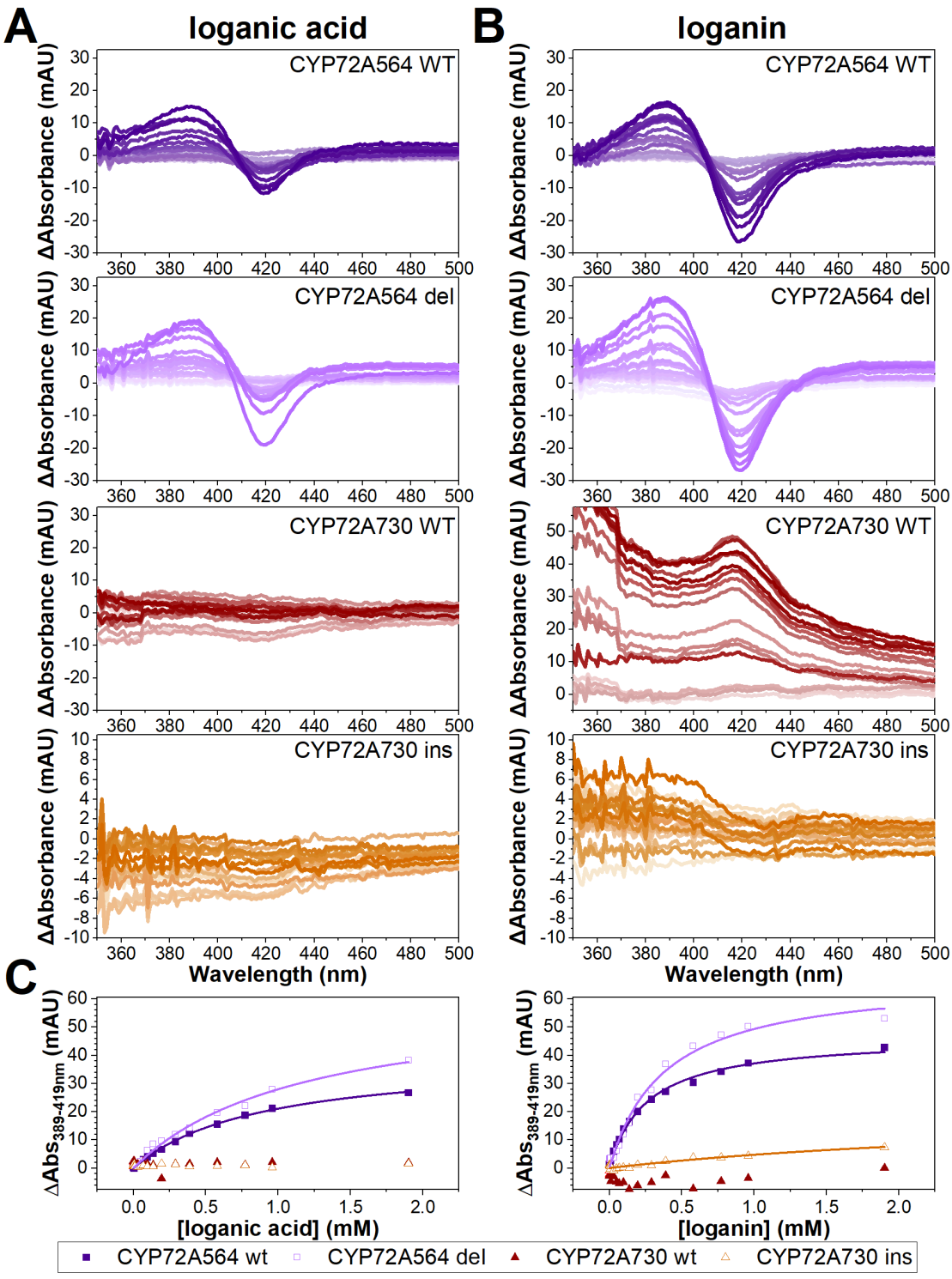


Figure 4.3. Substrate-induced Type I spectral shift titrations.
Difference spectra from 998 nM (lighter color) to 1.90 mM (darker color) loganic acid (A) or loganin (B) for identified CYP. (C) Binding isotherms derived by plotting the difference between the peak (389 nm) and trough (419 nm) of the difference spectra against the substrate concentration using OriginPro 2019. Table 4.2 records the fit parameters.

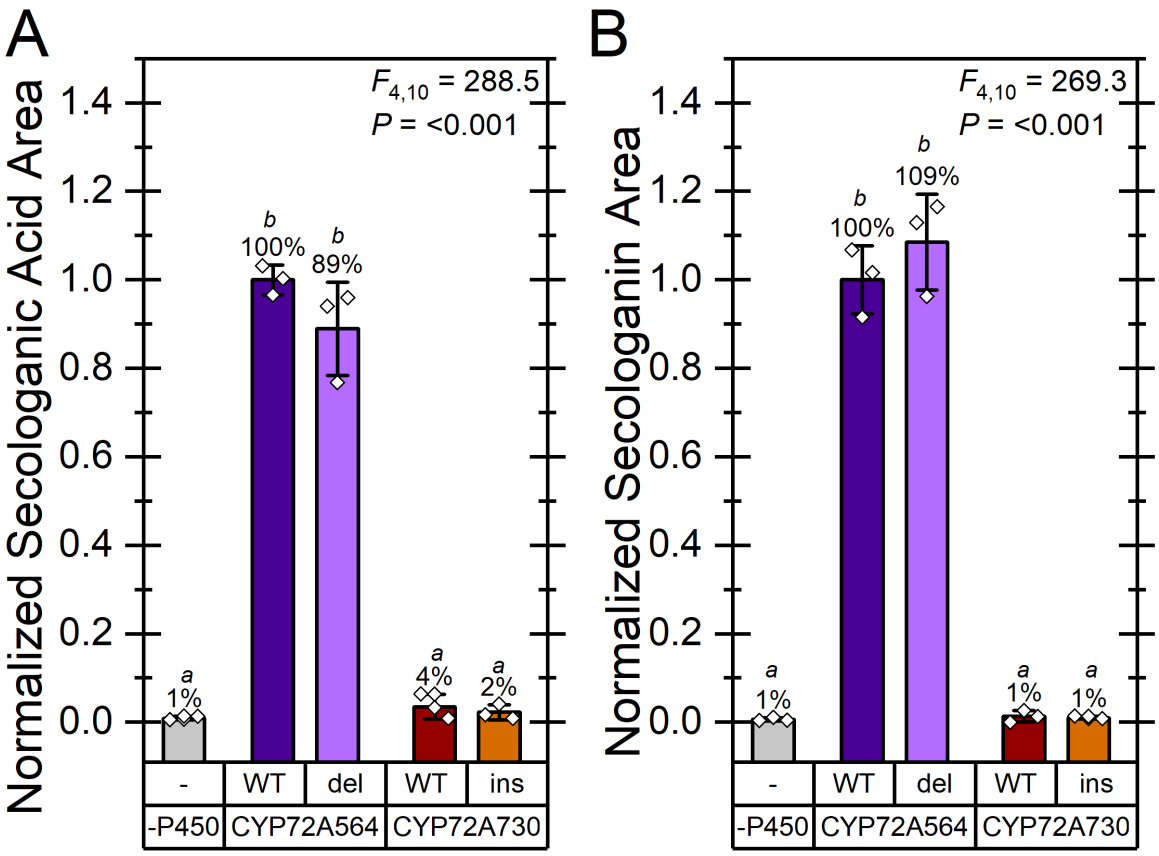


Figure 4.4. Secoiridoid production by CYP72A564, 72A564del, 72A730, 72A730del. Triplicate *in vitro* reactions of the listed CYP mutants with *Camptotheca* CPR1 were assayed for secologanic acid produced from loganic acid (A) and secologanin produced from loganin (B) by LC-MS normalized to an internal standard. The percentages of wild type activity (bars ± standard deviation with ◊ denoting individual replicates) were compared by a one-way ANOVA with Tukey's HSD test used for *post hoc* analysis ($\alpha = 0.05$). Figure 4.5 displays the extracted ion chromatograms used for quantitation. Tables 4.3 and 4.4 contain the output from the pairwise comparisons.

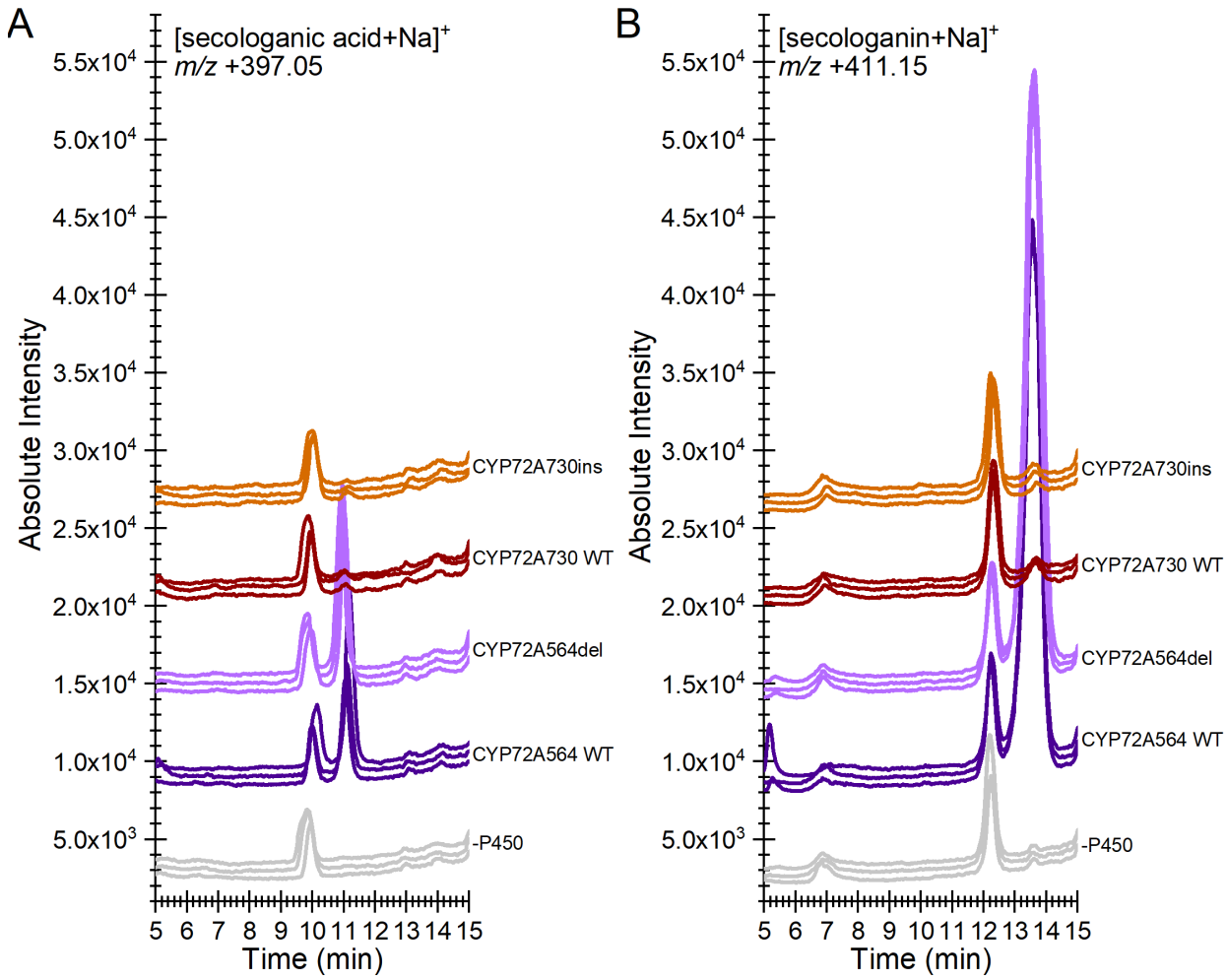


Figure 4.5. Extracted ion chromatograms of secoiridoids from *in vitro* reactions. Selective ion monitoring at *m/z* +397.05 $[secologanic\ acid+Na]^+$ (A) and *m/z* +411.15 $[secologanin+Na]^+$ (B) were used to assay secoiridoid production for CYP72A564, CYP72A564del, CYP72A730, and CYP72A730ins. Reactions included 250. nM P450, 0.5 mU μL^{-1} *Camptotheca* CPR1, 1 mM DLPC, 250. µM substrate, 500 µM NADPH, 1.00 mM glucose-6-phosphate, and 0.1 U μL^{-1} G6PDH in 100. mM $NaPO_4$ (pH 7.5).

4.5. REFERENCES

1. Jeffery, C. J. (2020) Enzymes, pseudoenzymes, and moonlighting proteins: diversity of function in protein superfamilies. *FEBS J.* **287**, 4141–4149
2. Copley, S. D. (2020) Evolution of new enzymes by gene duplication and divergence. *FEBS J.* **287**, 1262–1283
3. Miller, J. C., Hollatz, A. J., and Schuler, M. A. (2021) P450 variations bifurcate the early terpene indole alkaloid pathway in *Catharanthus roseus* and *Camptotheca acuminata*. *Phytochemistry*. **183**, 112626
4. Kang, M., Fu, R., Zhang, P., Lou, S., Yang, X., Chen, Y., Ma, T., Zhang, Y., Xi, Z., and Liu, J. (2021) A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* **12**, 3531
5. Yang, Y., Li, W., Pang, J., Jiang, L., Qu, X., Pu, X., Zhang, G., and Luo, Y. (2019) Bifunctional cytochrome P450 enzymes involved in camptothecin biosynthesis. *ACS Chem. Biol.* **14**, 1091–1096
6. Salim, V., Wiens, B., Masada-Atsumi, S., Yu, F., and De Luca, V. (2014) 7-Deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in Catharanthus roseus iridoid biosynthesis. *Phytochemistry*. **101**, 23–31
7. Salim, V., Yu, F., Altarejos, J., and De Luca, V. (2013) Virus-induced gene silencing identifies *Catharanthus roseus* 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis. *Plant J.* **76**, 754–765
8. Jin, Z., Yan, T., Chang, C., Liu, Z., Wang, Y., Tang, Z., and Yu, F. (2016) Application of virus-induced gene silencing approach in *Camptotheca acuminata*. *Plant Cell Tissue Organ Cult. PCTOC*. **126**, 533–540
9. van Hengel, A. J., Harkes, M. P., Wichers, H. J., Hesselink, P. G. M., and Buitelaar, R. M. (1992) Characterization of callus formation and camptothecin production by cell lines of *Camptotheca acuminata*. *Plant Cell Tissue Organ Cult.* **28**, 11–18
10. Wiedenfeld, H., Furmanowa, M., Roeder, E., Guzewska, J., and Gustowski, W. (1997) Camptothecin and 10-hydroxycamptothecin in callus and plantlets of *Camptotheca acuminata*. *Plant Cell Tissue Organ Cult.* **49**, 213–218
11. Song, and Byun, S. hoon and S. yo (1998) Characterization of cell growth and camptothecin production in cell cultures of *Camptotheca acuminata*. **8**, 631–638
12. Pan, X., Xu, H., liu, X., Gao, X., and lu, Y. (2004) Improvement of growth and camptothecin yield by altering nitrogen source supply in cell suspension cultures of *Camptotheca acuminata*. *Biotechnol. Lett.* **26**, 1745–1748

CHAPTER 5: SYNTHESIS AND APPLICATION OF SUBSTRATE ANALOGUES TO PROBE SLAS ACTIVITY

Despite its importance in creating the iridoid moiety immediately coupled to tryptamine to form terpene indole alkaloids (TIAs), the mechanism by which the *Camptotheca* secologanic acid synthases (SLASs; CYP72A564 and CYP72A565) catalyze C7-C8 scission on loganic acid and loganin to form secologanic acid and secologanin remains ambiguous. Observing that these enzymes facilitate hydroxylation or C-C scission reactions depending on which substrate the proteins are fed, I set out to synthesize a probe substrate to investigate the C-C scission reaction's mechanism and substrate-dependency. From geniposide (a readily available iridoid glucoside), I successfully reached the penultimate stage of my synthetic route and attempted a crucial cross-coupling reaction to install the C7 methyl group of 7-deoxy-7-methyl-loganin. This reaction failed, supporting a need to optimize the cross-coupling across a range of parameters or pursue an alternative route through terminal olefination at C7. Notwithstanding this drawback, the production of 7-methyl-loganin as a mechanistic probe remains an achievable means of assessing the C-C scission mechanism of these SLASs. I discuss the anticipated products from *in vitro* turnovers of this probe compound with CYP72A564 that can be determined by future LC-MS/MS analysis. Together, results obtained with this methylated analogue promise to unveil whether the C-C scission reaction mediated by SLSs and SLASs operates by a heterolytic or a homolytic route and unveil these enzymes' preferred site of hydrogen abstraction.

5.1. INTRODUCTION

The *Camptotheca* secologanic acid synthases (SLASs; CYP72A564 and CYP72A565) are remarkable for their ability to use multiple substrates (7-deoxyloganic acid, loganic acid, loganin) unlike their homologues in other species (1, 2) *and* for their ability to perform the

different reactions on these substrates. These CYPs, when fed 7-deoxyloganic acid (**1**), produce loganic acid (**2**)—a classic, stereospecific hydroxylation by a P450 (Fig. 5.1). In contrast, when fed loganic acid or loganin (**4**), C7-C8 scission coincides with C7 oxidation to an aldehyde and the production of a C8-C10 alkene thereby producing secologanic acid or secologanin (**5**), respectively.

Desiring to investigate the mechanism of this C-C breaking reaction outside of biological systems, Tietze (3) could not synthesize secologanin type compounds (**7**) from tosylated 10-hydroxyloganin aglycone (**6**) with strong base *in vitro.* His work suggested that 10-hydroxylation is not an intermediate step in the production of secologanin from loganin. Radiolabeling experiments with *Lonicera morrowii* (Caprifoliaceae) and *Adina plulifera* (Rubiaceae) in the early 1980s (4) confirmed this hypothesis, observing a route to secologanin *via* 7-deoxyloganic acid and loganin but not through 10-hydroxyloganin (**3**; Fig. 5.2). While investigating secoiridoid biosynthesis in *Olea europaea* (Oleaceae), Rodríguez-López *et al*. (5) observed that the SLS from *Catharanthus roseus* (Apocynaceae) produced 7-ketologanin (**10**) from 7-*epi*-loganin (**9**) *in lieu* of C10 hydroxylation or C7-C8 scission, supporting substrate dependence for C-C scission and an enzymatic preference for C7 hydroxylation. These radiolabeling and *in vitro* substrate feeding experiments alongside the lack of detectable 10-hydroxyloganic acid or 10-hydroxyloganin in *Camptotheca* (6, 7) (amongst other secoiridoid-producing plants) suggest an unusual mechanism at work in the SLAS/SLS (secologanin synthase) C-C scission reaction.

The discoverers of the SLS from *Lonicera japonica* (8) proposed two possible mechanisms for this C-C scission (Fig. 5.3). Starting with proton abstraction by Compound I at C10, the mechanisms diverge as to how the C10 radical decomposes: The homolytic route (red, top) predicts immediate C7-C8 scission to create the C8-C10 alkene and a C7 radical; oxygen

rebound from Compound II to this C7 radical produces a *gem*-diol in equilibrium with the product aldehyde. The heterolytic route from a primary C10 carbocation as drawn by Yamamoto *et al*. (8) (green, middle) is symmetry forbidden—likely explaining why Guengerich and Yoshimoto (9) and Rodríguez-López *et al*. (5) do not include this mechanism. Maintaining the electron transfer from the C10 radical to Compound II to produce the C10 carbocation, a pericyclic, Lewis-base catalyzed heterolytic mechanism with the carbocation (blue, bottom) is feasible.

To investigate the importance of substrate structure, stereochemistry, and inferred positioning in an SLS/SLAS active site, I attempted to synthesize 7-deoxy-7-methyl-loganin (**11**; Fig. 5.3). My earliest attempts were hampered by an inability to remove the C10-alcohol from geniposide (**12**). After overcoming these troubles, my synthetic route led to the successful production of a 7-*epi*-loganin derivative prepared for cross-coupling to yield the methylated probe. However, the crucial Cu(I)-catalyzed cross-coupling reaction consumed my existing 7-*epi*-loganin starting material and thereby prevented me from proceeding with alternative methods, at this time. The failure of this penultimate step highlights to need for an alternative route to 7-methyl-loganin, and the merits of several of these are discussed.

5.2. RESULTS AND DISCUSSION

5.2.1. Synthesis of loganin surrogate

My synthetic route to 7-deoxy-7-methyl-loganin (**11**; Fig. 5.3) drew largely from previous syntheses of iridoids (5, 10, 11) from the inexpensive and commercially available geniposide (**12**). First, acetylation of geniposide with acetic anhydride and pyridine yielded geniposide pentaacetate (**13**) in good yields. These acetates simultaneously served as protecting groups for the glucoside throughout the early part of the synthesis and as a leaving group on C10.

Indeed, C10 reduction catalyzed by $Pd(OH)_2$ on carbon with cyclohexene yielded 10-deoxygeniposide tetraacetate (**14**) in moderate yield.

A lack of diastereoselectivity with the *m*CPBA epoxidation of **14** cut into the yield of the desired epoxide. A series of selective NOESY experiments (Fig. 5.4) confirmed the (7*S*, 8*R*)-configuration of **15** in moderate yield. Epoxide opening with boron trifluoride etherate and water workup yielded 7-ketologanin (**16**), albeit with poor yield. Selective NOESY experiments (Fig. 5.5) also confirmed the (8*R*)-stereochemistry of **16**.

The acetate groups protecting the glucoside, though helpful since their installation by enabling the removal of the C10 hydroxyl of geniposide and protecting the glycoside, were not compatible with the reagents used for the planned cross coupling. Thus, these acetate protecting groups were exchanged for trimethylsilyl ethers by removal of the acetates with sodium methoxide and silation with trimethylsilyl chloride (Fig. 5.3). The combined 41% yield from these steps was both unexpected given the high yields of similar transformations in the literature (5, 12) and troublesome given the loss of mass expected upon deprotection. Notwithstanding these troubles, the approximately 70 mg of material appeared sufficient for the remainder of the synthesis.

Sodium borohydride reduction of **16** yielded a C7 alcohol that was tosylated in pyridine to afford **18** in good yields over two steps, crude (Fig. 5.3). This tosylate facilitated an attempt at the crucial Cu(I)-mediated cross coupling (13) with a methyl Grignard to produce a TMS-protected 7-methyl-loganin (**19)**. This cross-coupling reaction, however, failed to yield **19** after consuming the starting material. Instead, a C_{18}-tosylate was isolated suggesting a loss of the glucoside. An isolated product included a large number of negatively phased carbons in ^{1}H-^{13}C spectra, inconsistent with the presence of an intact iridoid core. Together, these data point to the

degradation of the starting material preventing recovery and attempts with alternative cross-coupling methods.

5.2.2. The need for an alternative route to 7-methyl-loganin.

The apparent incompatibility of the Cu(I)-mediated cross coupling of Yang *et al*. (13) with glycoside **18** necessitates alternative routes to **11**. The installation of the C7-methyl remains the crucial step with several possibilities: modifying the sp^3-sp^3 coupling route discussed earlier, or production of a terminal olefin followed by hydrogenation.

Revisiting and improving the sp^3-sp^3 cross coupling pursued initially is an option to access **11** (Fig. 5.6). Because no glycoside was recovered from the first attempt at cross-coupling, replacing the trimethylsilyl ethers with more stable protecting groups is a starting point (*e.g.*, with *tert*-butyldimethylsilyl ethers, allyl ethers, or methoxymethyl ethers). The Grignard carbanion source remains preferable to alkyl lithium due to the greater ability of the latter to reduce the methyl ester. Likewise, retaining the tosylate of **18** is preferable because it is easily accessible from 7-ketologanin **16**. Thus, the remaining area for consideration is catalyst choice. Cu, Fe, Ni, Zn, and Pd with an assortment of ligands present many options for testing.

Olefination of 7-ketologanin **16** or **17** enables access to the 7-methyl-loganin scaffold by hydrogenation (Fig. 5.6). With the use of a bulky base such as *tert*-butoxide to reduce destruction of the methyl ester, Wittig olefination with a methyl-triphenylphosphine presents one option. The requirement to remove stoichiometric amounts of phosphine and/or phosphine oxide, however, render other options more desirable. A Peterson elimination from reacting **17** with an α-silyl carbanion presents an alternative to the Wittig. Maintaining the ester again remains key, requiring careful selection of the carbanion source and subsequent elimination conditions to produce the C7-methylene. A modified Julia-Kocienski olefination with demonstrated ester

compatibility (14) thus appears to be the preferred path to **20**. This olefination route in general requires the stereo-selective hydrogenation of **20** to yield the desired diastereomer of 7-methyl-loganin. The ability of a Pd/C catalyst and H_2 to produce 7-deoxyloganin from 10-deoxygeniposide (5) support these reagents for stereoselective hydrogenation. Given all the factors to reconfigure in the cross-coupling approach, this sequential olefination and hydrogenation of **16** appears the most advantageous to accessing the 7-methyl-loganin probe.

5.2.3. Experimental approaches to investigating the SLS/SLAS C-C breaking reaction.

The purpose of synthesizing 7-deoxy-7-methyl-loganin (**11**) is to probe the substrate-dependent nature of the C7-C8 scission reaction catalyzed by the SLSs and SLASs. By combining my established *in vitro* reconstitution system with high resolution LC-MS to analyze the metabolites from **11**, the preference of the *Camptotheca* SLASs for C7 or C10 proton abstraction and the mechanism of C-C cleavage can be inferred.

Because the heterolytic mechanisms for C7-C8 scission require abstraction of a proton from the C7 alcohol (Fig. 5.2), the replacement of this alcohol with an immensely less acidic methyl will likely prevent breaking the C7-C8 bond using this mechanism. Instead, hydrogen atom abstraction by Compound I would lead to hydroxylation (Fig. 5.7)—though whether this is from the methyl attached to C7, C8, or both is unclear.

The exchange of the C7 alcohol for a methyl does not obviously preclude homolytic opening of the cylcopentane to produce an olefin and carbon-centered radical (Fig. 5.7). Oxygen rebound would give a terminal olefin and an alcohol. As before, the ability of the SLASs to react at either the C7 (hydroxylating 7-deoxyloganic acid) or C10 (commencing the sequence that cleaves the C7-C8 bond to form secologanic acid or secologanin) (2) prevents confidently predicting from which methyl group the initial abstraction will occur.

Differentiation of the metabolites resulting from single hydroxylation events and C-C bond cleavage with ring opening is not possible based solely upon their (identical) high-resolution ions. Knowing the relative amounts of product produced from loganic acid and loganin with the *in vitro* reconstitution system utilized throughout Chapters 2-4, there will not be sufficient amounts of product to enable NMR analyses. IR spectroscopy (either in line with LC or on fractions from HPLC) would enable differentiation of whether C-C scission occurred by searching for a second alkene C-H stretch. Tandem mass spectroscopy experiments, however, are my preference for their likely ability to not only differentiate the olefins arising from homolytic C-C scission from hydroxylation products supportive of the heterolytic mechanisms. Consideration of the fragmentation patterns can discern whether the preference for abstraction from C7 by the SLASs is general or particular to the 7-deoxyloganic acid scaffold.

Isotope labeling presents an alternative means of differentiating the homolytic and heterolytic C-C scission mechanisms. An ^{18}O from $^{18}O_2$ can only be incorporated into secologanin from loganin if the reaction proceeds *via* the radical rearrangement and formation of a *gem*-diol predicted by the homolytic route (Fig. 5.2). The conceptual straightforwardness of this experiment comes with several technical difficulties: First, the cost of ^{18}O-labled O_2 is prohibitive. Second, the maximum percent labeling is 50% owing to the scrambling of the aldehyde oxygen from the *gem*-diol. Lastly, the ^{18}O label is subject to exchange with unlabeled oxygen from water necessitating expedient data acquisition.

5.2.4. Conclusions

Understanding the mechanism underlying the C-C scission activity of the SLSs (*e.g.*, from *Catharanthus roseus*) and the SLASs (from *Camptotheca acuminata*) presents informed biosynthesis of new-to-nature compounds resembling existing specialized metabolites. Because this C-C breaking reaction is required to form the strictosidinic acid core transformed into all other terpene indole alkaloids (TIAs), structure-activity relationships on the TIA scaffold with a biosynthetic origin will require knowing what reactions each enzyme in the pathway are mechanistically capable of achieving. Although I was unsuccessful in synthesizing 7-methyl-loganin as a probe for the C-C bond breaking mechanism catalyzed by the *Camptotheca* SLASs, insights gleaned from my synthetic route through the penultimate step inform alternative routes to reach this molecule. *In vitro* turnover experiments followed by LC-MS/MS will subsequently provide insight into this important step in TIA biosynthesis.

5.3. EXPERIMENTAL

All reagents and solvents were purchased from commercial suppliers and used without further purification unless otherwise noted. Hexanes were vacuum distilled to remove high boiling impurities. When dry solvents were required, they were dried for two or more days over activated 3Å molecular sieves and stored under dry nitrogen gas. Thin layer chromatography utilized silica gel 60 doped with a fluorophore excited at 254 nm. Spots were visualized by fluorescent quenching and $KMnO_4$ staining. Flash column chromatography was conducted with 40-63 micron silica gel. ^{1}H NMR and ^{13}C NMR spectra were recorded on a 500 MHz spectrometer at the University of Illinois School of Chemical Sciences NMR Spectroscopy Lab. NMR samples were made as solutions in $CDCl_3$ or CD_3OD using the residual solvent peak as the internal standard; δ values are given in ppm and coupling constants (*J*) in hertz (Hz). Mass

spectra were obtained from a high-resolution ESI mass spectrometer at the University of Illinois School of Chemical Sciences Mass Spectrometry Lab.

5.3.1. Acetylation of geniposide.

Geniposide (**12**; 1.008 g, 2.58 mmol) was added to a mixture of acetic anhydride (3.0 mL, 31.1 mmol, 12 eq.) and pyridine (6.0 mL, 74.5 mmol, 29 eq.) and stirred at room temperature for 5 hours. Excess solvent was removed by rotary evaporation before diluting the mixture with 60 mL water and washing thrice with 50 mL EtOAc. The combined organic layers were dried over magnesium sulfate, concentrated under vacuum, and subjected to flash column chromatography (70% EtOAc in hexanes). This yielded geniposide pentaacetate (**13**, 1.279 g, 84% yield; R_f 0.48, 80% EtOAc in hexanes) as a white solid with spectra consistent with previous reports (5, 10).

^{1}H NMR (499 MHz, $CDCl_3$) δ 7.40 (d, *J* = 1.1 Hz, 1H), 5.82 (d, *J* = 3.2 Hz, 1H), 5.22 (td, *J* = 9.5, 0.8 Hz, 1H), 5.14 – 5.05 (m, 2H), 5.00 (ddd, *J* = 9.7, 8.1, 0.8 Hz, 1H), 4.86 (dd, *J* = 8.1, 0.8 Hz, 1H), 4.74 – 4.63 (m, 2H), 4.24 (dd, *J* = 12.3, 4.5 Hz, 1H), 4.15 (dd, *J* = 12.4, 2.5 Hz, 1H), 3.74 – 3.66 (m, 4H), 3.23 – 3.15 (m, 1H), 2.84 (dt, *J* = 19.6, 7.9 Hz, 2H), 2.20 – 2.11 (m, 1H), 2.08 (d, *J* = 0.8 Hz, 3H), 2.07 (d, *J* = 0.8 Hz, 2H), 2.01 (d, *J* = 0.8 Hz, 2H), 2.00 (s, 4H), 1.97 (d, *J* = 0.8 Hz, 2H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 20.52, 20.70, 20.72, 20.79, 20.98, 33.89, 38.63, 46.53, 51.40, 61.69, 61.96, 68.35, 70.81, 72.22, 72.59, 76.91, 77.16, 77.42, 96.04, 96.90, 112.17, 131.27, 137.10, 151.08, 167.43, 169.22, 169.46, 170.33, 170.74.

5.3.2. C10 reduction of geniposide penaacetate.

Pentaacetate **13** (1.008 g, 1.71 mmol) and 10% $Pd(OH)_2$ on carbon (399 mg) were dissolved in 2.85 mL EtOH. The flask was fit with a reflux condenser before replacing the atmosphere with nitrogen gas and heating to 70°C. Cyclohexene (1.13 mL, 17.1 mmol, 10 eq.) was added and the reaction was left to reflux overnight under nitrogen. After cooling, the

reaction was washed through Celite with DCM before concentrating under vacuum. Flash column chromatography (50% EtOAc in hexanes) yielded **14** as a white solid. The spectra were in accordance with literature reports (5, 10).

^{1}H NMR (499 MHz, $CDCl_3$) δ 7.38 (d, *J* = 1.1 Hz, 1H), 5.44 (p, *J* = 2.0 Hz, 1H), 5.22 (t, *J* = 9.5 Hz, 1H), 5.17 (d, *J* = 4.9 Hz, 1H), 5.10 (t, *J* = 9.7 Hz, 1H), 5.01 (dd, *J* = 9.7, 8.1 Hz, 1H), 4.87 (d, *J* = 8.1 Hz, 1H), 4.28 (dd, *J* = 12.3, 4.5 Hz, 1H), 4.18 – 4.07 (m, 1H), 3.72 (ddd, *J* = 10.0, 4.6, 2.6 Hz, 1H), 3.69 (s, 3H), 3.13 (tdd, *J* = 8.0, 4.8, 1.2 Hz, 1H), 2.72 (ttd, *J* = 8.0, 4.3, 2.1 Hz, 2H), 2.15 – 2.01 (m, 7H), 2.00 (s, 3H), 1.96 (s, 3H), 1.75 (dt, *J* = 3.2, 1.9 Hz, 3H), 1.25 (t, *J* = 7.1 Hz, 1H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 15.43, 20.46, 20.72, 20.82, 33.33, 38.49, 49.29, 51.30, 61.85, 68.44, 70.86, 72.19, 72.69, 76.91, 77.16, 77.41, 95.55, 96.51, 112.81, 127.21, 137.68, 150.86, 150.87, 167.62, 169.33, 169.52, 170.32, 170.70. HR-ESI-MS: $[M+H]^+$, calcd for $C_{25}H_{33}O_{13}$, 541.1921; found, 541.1928.

*5.3.3. Epoxidation of **14**.*

To a solution of **14** (581 mg, 1.02 mmol) in 25 mL DCM on ice was added *meta*-chloroperbenzoic acid (251 mg, 1.12 mmol, 1.1 eq.). The ice bath was removed and the reaction left to stir for three hours. After quenching with 20 mL 2N NaOH, the organic layer was removed before washing the aqueous layer twice with 20 mL DCM. The combined organic layers were combined and dried over $MgSO_4$ before concentrating by rotary evaporation. Flash column chromatography with 65% EtOAc in hexanes gave **15** (292 mg, 51% yield; R_f 0.43) as a white solid with spectra in accordance with the literature (5, 10).

^{1}H NMR (499 MHz, $CDCl_3$) δ 7.39 (d, *J* = 1.3 Hz, 1H; C3), 5.38 – 5.17 (m, 1H; C3'), 5.14 (t, *J* = 9.7 Hz, 1H; C4'), 5.07 (dd, *J* = 9.5, 8.0 Hz, 1H; C2'), 5.02 (d, *J* = 8.1 Hz, 1H; C1'), 4.88 (d, *J* = 9.6 Hz, 1H; C1), 4.22 (d, *J* = 3.5 Hz, 2H; C6'), 3.76 – 3.69 (m, 1H; C5'), 3.70 (s, 3H,

OMe), 3.27 (s, 1H; C7), 2.83 – 2.74 (m, 1H; C5), 2.61 (dd, *J* = 14.0, 7.7 Hz, 1H; $C6_{pro\text{-}R}$), 2.18 (dd, *J* = 9.6, 7.3 Hz, 1H; C9), 2.11 – 1.96 (m, 12H; $(OAc)_4$), 1.51 (s, 3H; C10), 1.34 (dd, *J* = 13.9, 10.0 Hz, 1H; $C6_{pro\text{-}S}$). ^{13}C NMR (126 MHz, $CDCl_3$) δ 17.23 (C10), 20.72 ($COO\underline{C}H_3$), 20.81 ($COO\underline{C}H_3$), 30.37 (C5), 34.91 (C6), 43.68 (C9), 51.51 (OMe), 61.55 (C6'), 61.71 (C7), 63.86 (C8), 68.37 (C4'), 70.97 (C2'), 72.25 (C3'), 72.65 (C3'), 95.25 (C1), 96.76 (C1'), 110.12 (C4), 150.96 (C3), 167.23 (C11), 169.44 ($\underline{C}OOCH_3$), 169.52 ($\underline{C}OOCH_3$), 170.30 ($\underline{C}OOCH_3$), 170.61 ($\underline{C}OOCH_3$). HR-ESI-MS: $[M+Na]^+$, calcd for $C_{25}H_{32}O_{14}$,Na 579.1690; found, 579.1693.

5.3.4. Production of 7-ketologanin tetraacetate.

Under dry nitrogen gas, **15** (280 mg, 0.521 mmol) was dissolved in 11.6 mL DCM. $BF_3 \cdot Et_2O$ (1.16 mL) was added dropwise resulting in a light pink solution that darkened over the course of 10 minutes while stirring under nitrogen. The reaction was quenched with 10 mL sat. $NaHCO_3$; the organic layer collected; and the aqueous layer extracted thrice with 20 mL DCM before drying the organic layers of $MgSO_4$. Concentration under vacuum produced a residue that was subjected to flash column chromatography (65% EtOAc in hexanes) to yield **16** (138 mg, 48% yield; R_f 0.45, 65% EtOAc in hexanes). Spectra were consistent with the literature (5, 10)

^{1}H NMR (499 MHz, $CDCl_3$) δ 7.39 (d, *J* = 1.5 Hz, 1H, C3), 5.44 (d, *J* = 2.3 Hz, 1H, C1), 5.23 (t, *J* = 9.6 Hz, 1H, C3'), 5.11 (t, *J* = 9.7 Hz, 1H, C4'), 5.00 (dd, *J* = 9.7, 8.2 Hz, 1H, C2'), 4.89 (d, *J* = 8.1 Hz, 1H, C1'), 4.31 (dd, *J* = 12.5, 4.5 Hz, 2H, C6'), 3.75 (ddd, *J* = 10.0, 4.5, 2.4 Hz, 1H, C5'), 3.70 (s, 3H, OMe), 3.19 – 3.11 (m, 1H, C5), 2.62 (dt, *J* = 19.7, 1.9 Hz, 1H, C6a), 2.55 (dd, *J* = 19.6, 8.2 Hz, 1H, C6b), 2.31 (ddd, *J* = 11.7, 7.2, 2.3 Hz, 1H, C9), 2.09 (s, 3H, OAc), 2.05 – 2.03 (m, 1H, C8), 2.03 (s, 3H, OAc), 2.00 (s, 3H, OAc), 1.90 (s, 3H, OAc), 1.15 (d, *J* = 7.0 Hz, 3H, C10). ^{13}C NMR (126 MHz, $CDCl_3$) δ 13.28 (C10), 20.23 ($COO\underline{C}H_3$), 20.72 ($COO\underline{C}H_3$), 20.87 ($COO\underline{C}H_3$), 21.19 ($COO\underline{C}H_3$), 26.59 (C5), 42.10 (C6), 43.12 (C8), 45.13

(C9), 51.55 (OMe), 60.53 (C5'), 61.78 (C6'), 68.24 (C4'), 70.68 (C2'), 72.45 (C3'), 93.37 (C1), 96.02 (C1'), 111.02 (C4), 150.56 (C3), 166.65 (COOMe), 169.50 ($\underline{C}OOCH_3$), 170.27 ($\underline{C}OOCH_3$), 170.29 ($\underline{C}OOCH_3$), 170.71 ($\underline{C}OOCH_3$), 217.53 (C7). HR-ESI-MS: $[M+H]^+$, calcd for $C_{25}H_{33}O_{14}$, 557.1870; found, 557.1869.

5.3.5. Swapping of acetyl for trimethyl silyl protecting groups.

16 (138 mg, 0.248 mmol) was dissolved in 18 mL MeOH before adding NaOMe (54 μL 4.6 M solution in MeOH, 0.248 mmol, 1.0 eq.) and stirring for three hours at room temperature. Dowex 50W-8X (1.0 g) was added to quench remaining methoxide and trap sodium ions. After stirring with the resin for an hour, the solution was filtered and the solvent removed under vacuum to yield 7-ketologanin (95 mg, 99% yield), which was used without purification.

Under dry nitrogen gas, this sample of 7-ketologanin was dissolved in 1 mL pyridine. Trimethylsilyl chloride (157 μL, 1.24 mmol, 1.25 eq. per alcohol) was added dropwise and the solution stirred under nitrogen overnight. The reaction was quenched with 20 mL water before thrice extracting with 40 mL EtOAc. The combined organic layers were washed twice with 15 mL water before drying over $MgSO_4$. Rotary evaporation gave **17** (69 mg, 41% yield).

^{1}H NMR (500 MHz, $CDCl_3$) δ 7.48 (d, *J* = 1.5 Hz, 1H), 5.42 (d, *J* = 4.1 Hz, 1H), 4.63 (dd, *J* = 7.6, 1.4 Hz, 1H), 3.77 (dd, *J* = 11.4, 2.0 Hz, 1H), 3.71 (s, 2H), 3.71 – 3.65 (m, 1H), 3.50 – 3.38 (m, 2H), 3.33 – 3.20 (m, 2H), 2.65 (dd, *J* = 19.3, 8.4 Hz, 1H), 2.46 (ddd, *J* = 19.3, 4.7, 1.6 Hz, 1H), 2.27 – 2.11 (m, 2H), 1.28 (s, 1H), 1.25 (s, 1H), 1.16 (d, *J* = 7.1 Hz, 3H), 0.95 – 0.85 (m, 1H), 0.21 – 0.03 (m, 36H). ^{13}C NMR (126 MHz, $CDCl_3$) δ -0.33, -0.09, 0.78, 0.94, 1.02, 1.25, 1.44, 1.49, 1.72, 11.11, 14.03, 14.20, 23.13, 23.89, 27.65, 29.07, 29.85, 30.51, 38.88, 43.14, 43.81, 45.40, 51.47, 51.51, 61.93, 68.30, 71.29, 74.84, 76.91, 77.16, 77.27, 77.41, 77.92, 94.18,

99.32, 109.93, 128.95, 131.03, 152.39, 167.09, 218.12. HR-ESI-MS: $[M+Na]^+$, calcd for $C_{29}H_{56}O_{10}Si_4Na$, 699.2843; found, 699.2838.

*5.3.6. Reduction, tosylation of **17**.*

17 (69 mg, 0.102 mmol) was dissolved in 5 mL THF. 5 mL MeOH was added to the solution before adding sodium borohydride (6 mg, 0.159 mmol, 1.6 eq.) and stirring for 90 minutes at room temperature. The reaction was quenched with 10 mL sat. $NaHCO_3$, diluted with 5 mL water, and washed thrice with 15 mL DCM. The combined organic layers were dried over $MgSO_4$ and concentrated under vacuum to give a C7-alcohol (60.5 mg, 88% yield) used without further purification.

This alcohol (0.089 mmol) was dissolved in 1 mL pyridine before adding *para*-toluenesulfonyl chloride (20 mg, 0.107 mmol, 1.2 eq.) and stirring at room temperature for three hours. The reaction was quenched with 10 mL water; extracted thrice with 10 mL DCM; the combined organic layers washed twice with 10 mL 10% $NaHCO_3$; and dried over $MgSO_4$. Concentration under vacuum gave a colorless residue (**18**, 62 mg, 83%, 73% over two steps) used without further purification.

5.3.7. Cu(I)-catalyzed cross-coupling to produce 7-methyloganin structure

Following the method reported by Yang and colleagues (13), the tosylate **18** (62 mg, 0.074 mmol), CuI (1.8 mg, 0.0095 mmol, 0.1 eq.), and LiOMe (3.2 mg, 0.084 mmol, 1.1 eq.) were mixed under air. After replacing the atmosphere with dry nitrogen gas, these solids were dissolved in 0.15 mL THF and cooled in an ice bath before adding TMEDA (2.2 μL, 0.015 mmol, 0.2 eq) then CH_3MgBr (49 μL of 3 M solution in Et_2O, 0.148, 2 eq.). The reaction was left to stir in an ice bath under nitrogen for 12 hours before stirring another 12 hours at room temperature. The reaction was quenched by the addition of 1.5 mL sat. NH_4Cl. The resulting

mixture was washed thrice with 1.5 mL DCM, dried over $MgSO_4$ filtered through a pad of silica gel, washed with copious DCM, and concentrated under vacuum to give a yellow residue. Flash column chromatography with 5% EtOAc in hexanes gave a colorless residue, the spectra of which were consistent with a tosylate and *not* **19**.

5.4. FIGURES

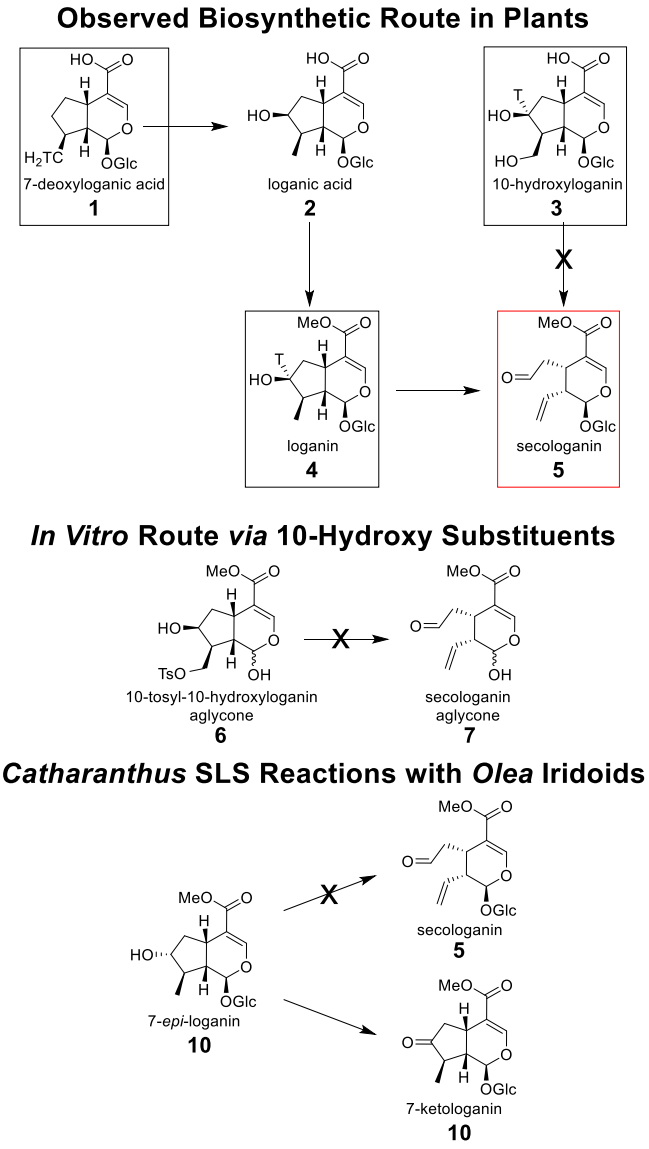


Figure 5.1. Iridoid to secoiridoid conversion *en route* to terpene indole alkaloids derived from strictosidine or strictosidinic acid.
Compounds shown with black boxes denote radiolabeled substrates fed to plants (4) with a readout of incorporation into secologanin (red box). "X" denotes transformations not observed by Inoue *et al*. (4), Tietze (3), and Rodríguez-López *et al*. (5) in their respective experiments.

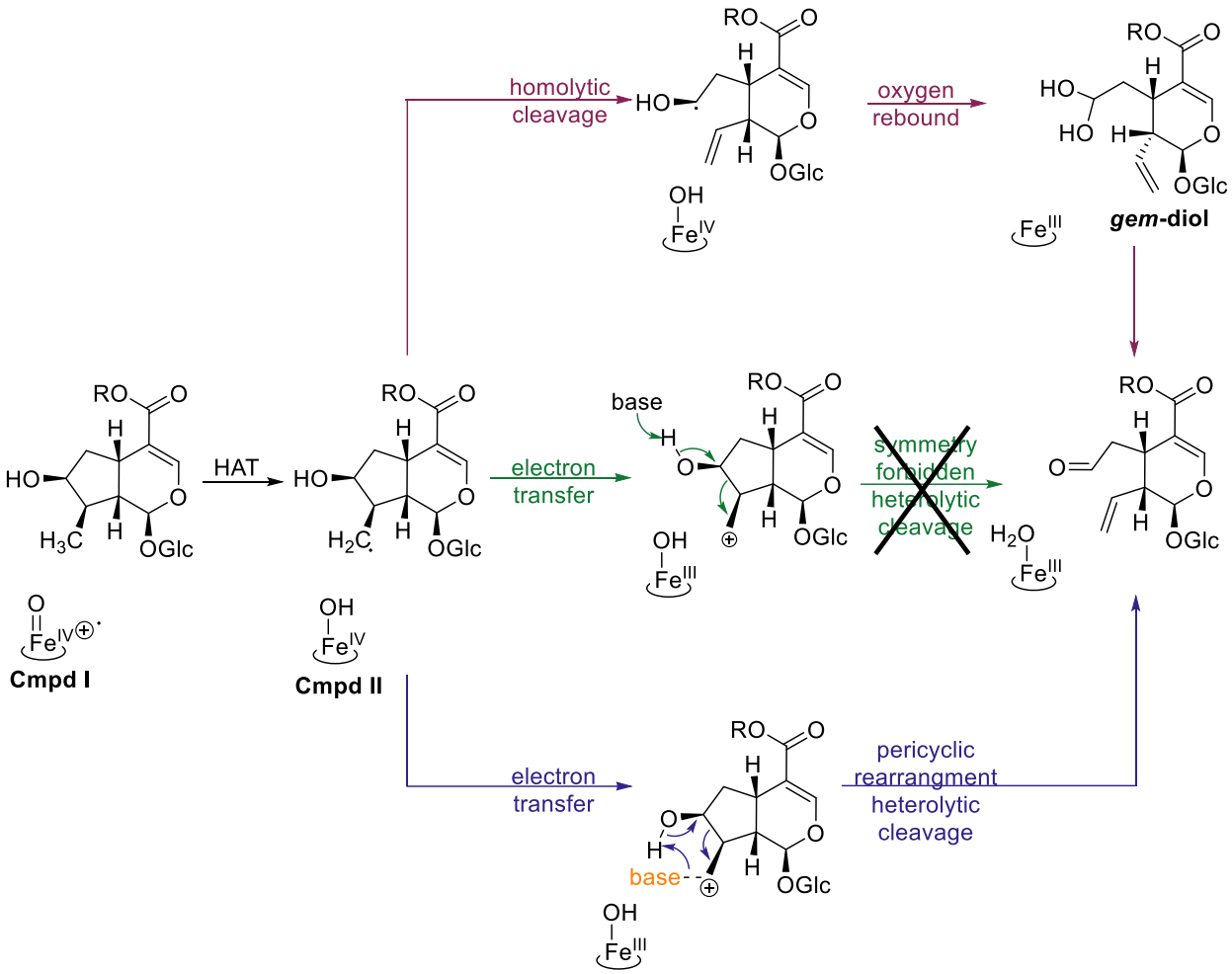


Figure 5.2. Mechanistic proposals for production of secoiridoids by secologanin synthases (SLSs) and secologanic acid synthases (SLASs).

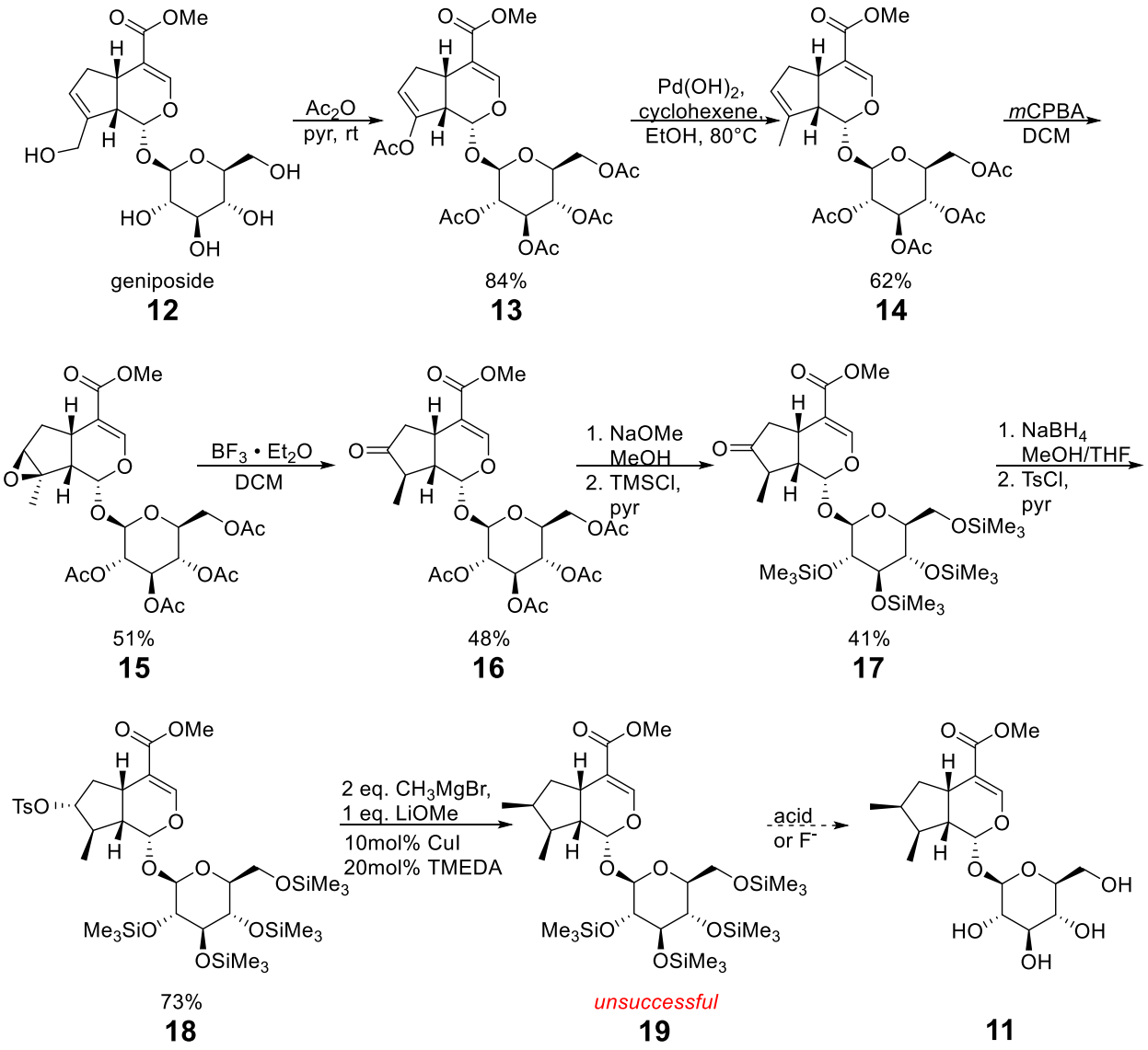


Figure 5.3. Synthetic route to 7-deoxy-7-methyl-loganin from geniposide.

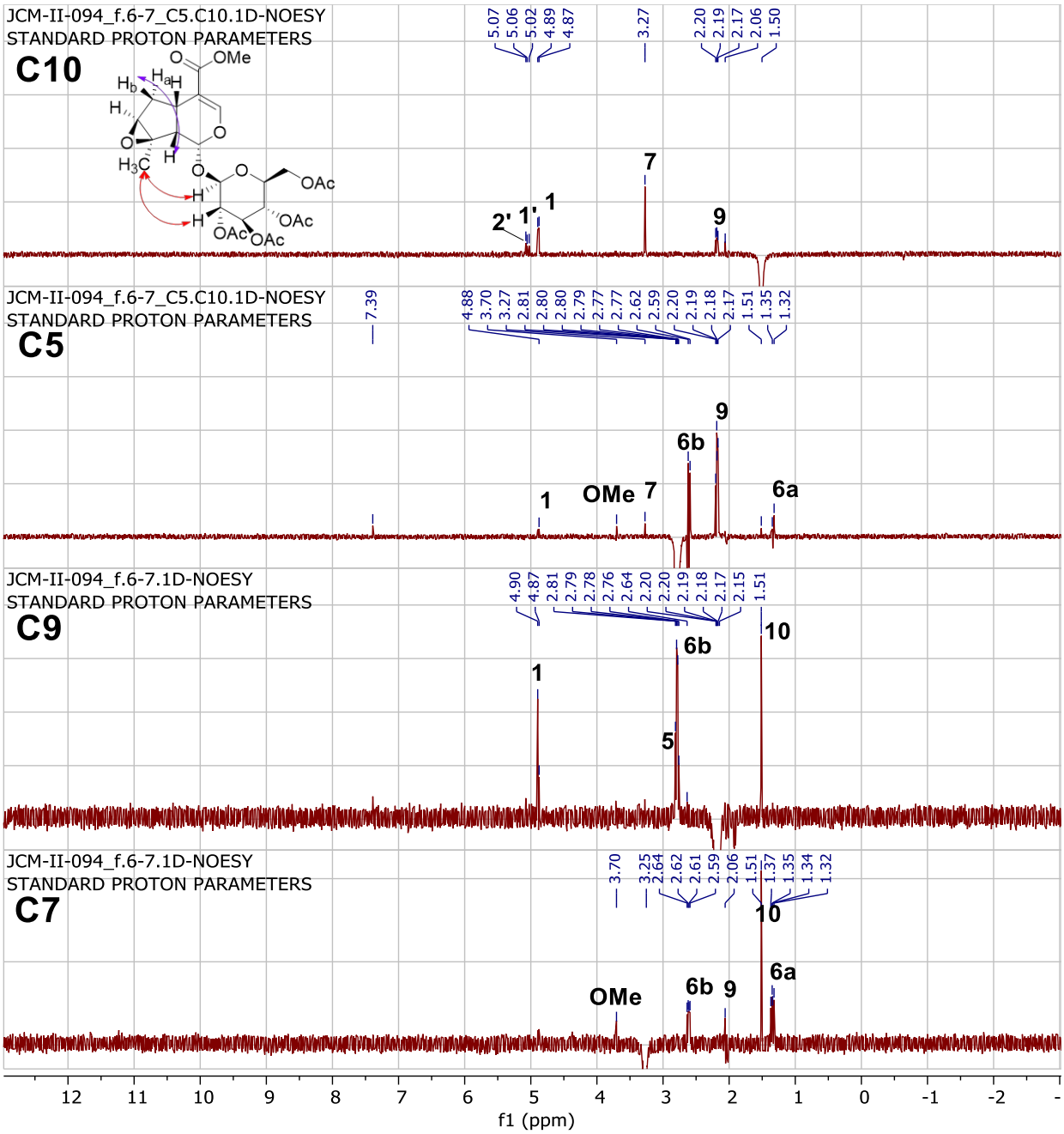


Figure 5.4. Selective NOESY confirms the stereochemistry of epoxide 15.
Selective NOESY irradiating the proton(s) at C10 (top), C5 (middle top), C9 (middle bottom), and C7 (bottom) collected with a 0.5 s. mixing time. Numbers over peaks indicate the position of the proton(s) on **15**.

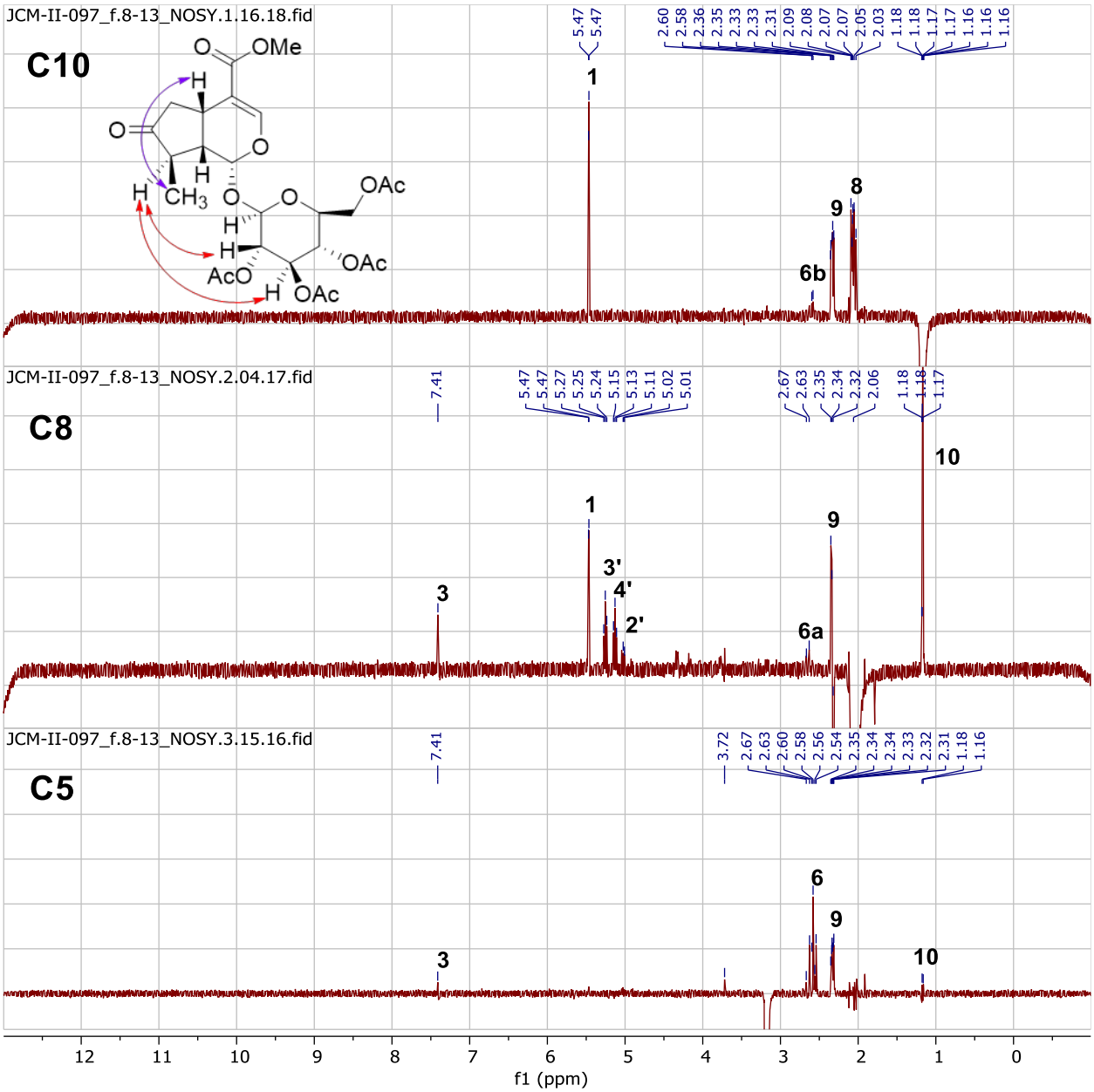
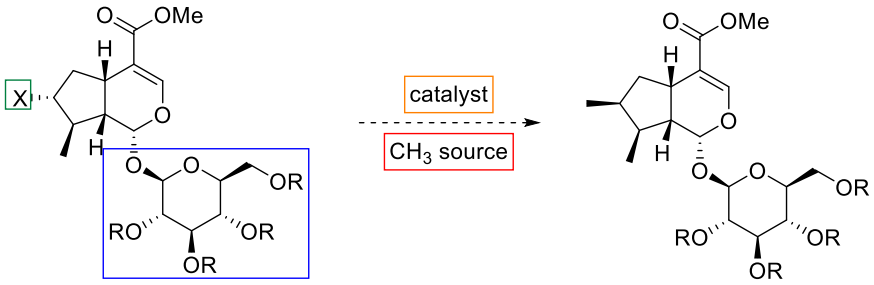
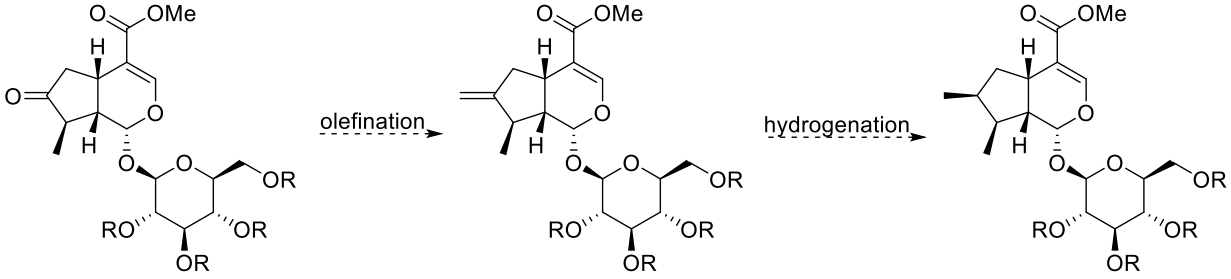


Figure 5.5. Selective NOESY confirms the stereochemistry of 7-ketologanin 16. Selective NOESY irradiating the proton(s) at C10 (top), C8 (middle), and C5 (bottom) collected with a 0.5 s. mixing time. Numbers over peaks indicate the position of the proton(s) on **16**.

Modifications to Cross Coupling Route



C7 Terminal Olefination



R = Ac, **16**
R= $SiMe_3$, **17**

R = Ac, **20a**
R= $SiMe_3$, **20b**

R = $SiMe_3$, **19**
R= H, **11**

Figure 5.6. Alternative synthetic approaches to 7-methyl-loganin

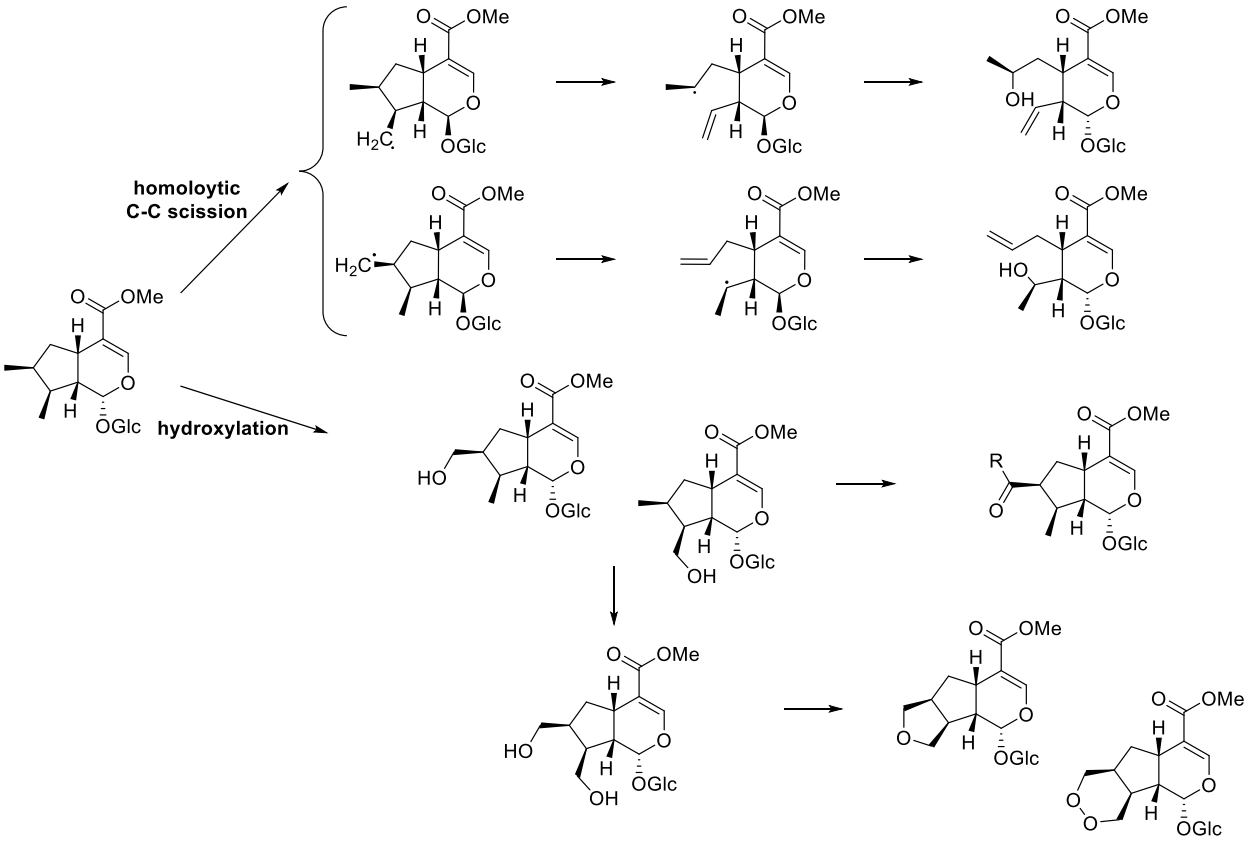


Figure 5.7. Potential metabolites from reacting SLAS CYP72A564 with 11.

5.5. REFERENCES

1. Miller, J. C., Hollatz, A. J., and Schuler, M. A. (2021) P450 variations bifurcate the early terpene indole alkaloid pathway in *Catharanthus roseus* and *Camptotheca acuminata*. *Phytochemistry*. **183**, 112626
2. Yang, Y., Li, W., Pang, J., Jiang, L., Qu, X., Pu, X., Zhang, G., and Luo, Y. (2019) Bifunctional cytochrome P450 enzymes involved in camptothecin biosynthesis. *ACS Chem. Biol.* **14**, 1091–1096
3. Tietze, L. F. (1974) Fragmentation of hydroxyloganin derivatives. Easy access to secologanin type compounds. *J. Am. Chem. Soc.* **96**, 946–947
4. Inoue, K., Takeda, Y., Tanahashi, T., and Inouye, H. (1981) Studies on monoterpene glucosides and related natural products. XLII. On the possibility of the intermediacy of 10-hydroxyloganin in the biosynthesis of secologanin. *Chemical & Pharmaceutical Bulletin*. **29**, 981–990
5. Rodríguez-López, C. E., Hong, B., Paetz, C., Nakamura, Y., Koudounas, K., Passeri, V., Baldoni, L., Alagna, F., Calderini, O., and O'Connor, S. E. (2021) Two bi-functional cytochrome P450 CYP72 enzymes from olive (*Olea europaea*) catalyze the oxidative C-C bond cleavage in the biosynthesis of secoxy-iridoids – flavor and quality determinants in olive oil. *New Phytol.* **229**, 2288–2301
6. Sadre, R., Magallanes-Lundback, M., Pradhan, S., Salim, V., Mesberg, A., Jones, A. D., and DellaPenna, D. (2016) Metabolite diversity in alkaloid biosynthesis: A multilane (diastereomer) highway for camptothecin synthesis in *Camptotheca acuminata*. *The Plant Cell*. **28**, 1926–1944
7. Jin, Z., Wan, R., Yan, R., Su, Y., Huang, H., Zi, L., and Yu, F. (2019) Microwave-Assisted Extraction of Multiple Trace Levels of Intermediate Metabolites for Camptothecin Biosynthesis in Camptotheca acuminata and Their Simultaneous Determination by HPLC-LTQ-Orbitrap-MS/MS and HPLC-TSQ-MS. *Molecules*. **24**, 815
8. Yamamoto, H., Katano, N., Ooi, A., and Inoue, K. (2000) Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. *Phytochemistry*. **53**, 7–12
9. Guengerich, F. P., and Yoshimoto, F. K. (2018) Formation and cleavage of C–C bonds by enzymatic oxidation–reduction reactions. *Chem. Rev.* **118**, 6573–6655
10. Inouye, H., Yoshida, T., Tobita, S., and Okigawa, M. (1970) Studies on monoterpene glucosides—IX: Chemical correlation between asperuloside and loganin. *Tetrahedron*. **26**, 3905–3915
11. Fujita, T., Inoue, K., Ono, M., Nakajima, H., Fijie, I., and Inouye, H. (1992) Radioimmunoassay of iridoid glucosides: Part 1. General method for preparation of the haptens and the conjugates with a protein of this series of glucosides. *Heterocycles*. **33**, 673
12. Gouasmat, A., Lemétais, A., Solles, J., Bourdreux, Y., and Beau, J.-M. (2017) Catalytic iron(III) chloride mediated site-selective protection of mono- and disaccharides and one trisaccharide. *European Journal of Organic Chemistry*. **2017**, 3355–3361
13. Yang, C.-T., Zhang, Z.-Q., Liang, J., Liu, J.-H., Lu, X.-Y., Chen, H.-H., and Liu, L. (2012) Copper-catalyzed cross-coupling of nonactivated secondary alkyl halides and tosylates with secondary alkyl grignard reagents. *J. Am. Chem. Soc.* **134**, 11124–11127

14. Aïssa, C. (2006) Improved Julia−Kocienski Conditions for the Methylenation of Aldehydes and Ketones. *J. Org. Chem.* **71**, 360–363

CHAPTER 6: INVESTIGATING C-C LYASE REACTIONS OF α-HYDROXY KETONES USING BACTERIAL P450s: PART 1, (WOES) USING $P450_{CAM}$ AND $P450_{BM3}$

The C-C lyase reaction of human, steroid metabolizing CYP17A1 has been of longstanding interest to biochemists for its unusual reaction and to medicinal chemists as a target for prostate cancer chemotherapeutics. A long series of experiments have narrowed the mechanistic possibilities for the C-C breaking step to two possible mechanisms: One utilizing the nucleophilic peroxoanion to simultaneously break the C17-C20 and dioxygen bond while the C17 ketone, water, and acetate bonds form; the other utilizes Compound I to form a C17-hydroperoxo steroid that initially forms a dioxetane before decomposing into the C17 steroid and acetate products. To facilitate further mechanistic studies to differentiate these two possibilities, I attempted to coax two well-studied bacterial P450s—CYP101A1 ($P450_{cam}$) and CYP102A1 ($P450_{BM3}$)—into performing the C-C lyase reaction on several α-hydroxyketones. Initial experiments used commercially available cyclopentyl- and cyclohexyl- α-hydroxyketones, but the predicted ketone products were not observed. Molecular dynamics (MD) simulations suggested that bulkier compounds more closely resembling the terpene substrates of CYP101A1 were more likely to undergo the C-C lyase reaction. After synthesizing a [2.2.1]-bicyclo α-hydroxyketone from camphor and an adamantyl α-hydroxyketone from 2-adamantanone, *in vitro* turnovers with CYP101A1 again yielded no evident of the C-C lyase reaction. Returning to MD, these simulations supported mutating Tyr96Phe and Thr101Val to prevent hydrogen-bonding interactions that were orienting the α-hydroxyketone moiety away from the oxygen-heme complex. These CYP101A1 mutants likewise did not catalyze C-C lyase chemistry on any of the α-hydroxyketone chemistry. Another look at the MD simulations suggested that the substrates, though frequently in position for the Compound I-mediated mechanism, were rarely in a

conformation favoring the C-C lyase reaction by the peroxyanion mechanism and that such confirmations were too short-lived for catalysis.

6.1. INTRODUCTION

Cytochrome P450-catalyzed carbon-carbon bond (C-C) breaking reactions extend well beyond the CYP72A subfamily, the plant kingdom, and specialized metabolism. These unusual transformations are predominantly performed by P450s to reshape steroids, fatty acids, furanocoumarins, and a wide array of isoprenoid-derived natural products (1). The conversion of cholesterol into the array of steroids mammals employ requires numerous P450 C-C cleavage reactions (2): cholesterol to pregnenolone (C20-C22; CYP11A1); 17-α-hydroxy-pregnenolone and 17-α-hydroxy-progesterone to dehydroepiandrosterone and androstenedione, respectively (C-17-C20; CYP17A1); testosterone to estradiol (C10-C19; CYP19A1).

The essential role of CYP17A1 in producing androgen-type steroids has made it a major target for treating prostate cancers dependent on these steroids (3, 4). The multifunctionality of this P450 (Fig. 6.1A), however, has substantially complicated these efforts. The inhibition of 17-hydroxylation of pregnenolone and progesterone by CYP17A1 in particular prevents the biosynthesis of cortisols and cortisones. These corticosteroids are essential to health, thus presenting a narrow therapeutic window for drugs that inhibit both the hydroxylation *and* lyase activities of CYP17A1.

Differentiation of the hydroxylation and lyase activities by mechanistic studies have emerged with a goal of designing C-17-C20 lyase-selective inhibitors of CYP17A1 that do not affect the C17 hydroxylation activity. From a combination of experiments including ^{18}O-isotopic labeling (5–7), resonance Raman (rRaman) (8–12), kinetic solvent isotope effect (KSIE) (13), and feeding 17α-peroxo- (14) and 17- and/or 21-fluorinated (15) steroids as well as combined

density functional theory (DFT) and molecular dynamics (MD) modeling (16), two proposed mechanisms have emerged (Fig. 6.1B) for the lyase reaction (1): The first of these corresponds to heterolytic C-C cleavage of the α-hydroxyketone with the peroxoanion-iron complex; the second corresponds to the production of a peroxide with the 17-hydroxyl to form a 1,2-dioxetane that subsequently decomposes. A high degree of uncoupling between CYP17A1 and cytochrome P450 reductase (CPR), the lyase reaction's sluggish rate of product formation, and the membrane-bound nature of this system continue to present challenges for studying this important chemical transformation.

Desiring to clarify the mechanism behind P450-mediated C-C scission of α-hydroxyketones in CYP17A1, two well-characterized bacterial CYPs $P450_{cam}$ (CYP101A1) and $P450_{BM3}$ (CYP102A1) were chosen for modeling the lyase activity of CYP17A1. With cyclopentane and cyclohexane based α-hydroxyketones and ones synthesized on camphor ([2.2.1]-bicyclic) and adamamantane scaffolds, neither of these bacterial enzymes performed the C-C lyase reaction. Molecular dynamics simulations instead suggested that maintaining the α-hydroxyketone close to the oxygen-heme complex is important for C-C scission.

6.2. RESULTS

6.2.1. Attempts with cyclopentyl, cyclohexyl α-hydroxyketones[1]

Cyclopentyl and cyclohexyl α-hydroxyketones **1** and **2**, respectively, were commercially available and resembled known substrates for $P450_{cam}$ (17) and $P450_{BM3}$ when a polyfluorinated decoy molecule is used (18). **2** induced Type I binding spectra when added to $P450_{cam}$ (Fig. 6.2), corroborating it as a likely substrate for this CYP. The rate of NADH oxidation increased upon

[1] These *in vitro* assays are a combination of work from Yelena Grinkova, Dr. Mark McLean, and JCM.

addition of this compound; however, the cyclopentanone (**3**) and cyclohexanone (**4**) products expected for the C-C lyase reaction were not observed (Fig. 6.3).

Because wild type $P450_{cam}$ very quickly converts the peroxoanion intermediate into Compound I (19, 20), the expectation was that, if the peroxyanion mechanism is responsible for the C-C lyase reactivity (Fig. 6.1), quick conversion of the peroxy/hydroperoxy species to Compound I would hinder catalysis. The Thr252Ala mutant known to diminish Compound I production (19, 20) was used to evaluate this possibility. **1** and **2** both induced Type I spectral shifts to Thr252Ala (Fig. 6.4). As with the wild type, no significant production of cyclopentanone or cyclohexanone, respectively, was observed (Fig. 6.5).

Concerned that the α-hydroxyketone portion of these molecules was locked in a H-bond interaction with Tyr96 of $P450_{cam}$ as observed for the ketone of camphor (21), $P450_{BM3,}$ where cyclohexane is a known substrate in the presence of decoy molecules (18), was used. Neither **1** nor **2** yielded significant amounts of cycloketone **3** or **4** with $P450_{BM3}$ (Fig. 6.6).

The lack of success using the monocyclic α-hydroxyketones **1** and **2** suggested that compounds more closely resembling the bicylic structure of camphor (**14**; Fig. 6.7) were better for investigating the C-C lyase reaction. Because adamantane and adamantanone are known substrates of $P450_{cam}$ (22), their tricyclic scaffold was also targeted for the synthesis of probe α-hydroxyketones. While these syntheses were underway, docking and molecular dynamics with $P450_{cam}$ were used to identify possible mutants that might increase the positioning required for a C-C lyase reaction.

6.2.2. Synthesizing [2.2.1]-bicyclo, adamantyl α-hydroxyketones

My initial goal was the synthesis of **8**, an α-hydroxyketone at the C5 position of camphor (**14**) from 5-ketoborneol (**5**; Fig. 6.7). The first transformation, a Wittig olefination at C5, was

consistently troublesome. A combination of low (<40%) yields of what was presumed to be **6** still contaminated with triphenylphosphine and/or triphenylphosphine oxide led to attempting the oxidation of crude alkene **6** to produce *vicinal*-diol **7**.[2] Whether conducted with OsO_4 or $NaIO_4$/LiBr (23), the resulting product mixture was difficult to separate, with no fewer than five compounds present. Given the troubles faced in this synthetic route, I turned my attention to synthesizing other [2.2.1]-bicylco scaffolds and an adamantane scaffold.

As with 5-ketoborneol (**5**), the olefination of norbornanone **9** (Fig. 6.7) was cumbersome with added complications from this smaller compound's greater volatility. Reports of obtaining **10** by reducing ethylidenenorbornene (**11**) using a solid-supported nickel boride catalyst (24) prompted pursuit of this alternative route. Unlike Choi and Yoon, however, I was unable to isolate **10** and instead obtained an insoluble white solid likely resulting from polymerization.

Observing Hong and coworker's (25) successful conversion of trimethylsilyl-protect cyanohydrins to α-hydroxyketones *en route* to bisorbicillinoids using alkyl Grignards, I switched synthetic routes once more (Fig. 6.7). A camphor cyanohydrin silyl ether (**12**) was produced in high yields with a zinc-catalyzed cyanosilylation (26). Reduction of the nitrile in **15** with methyl Grignard and conversion to α-hydroxyketone **16** required refluxing for 20 hours in toluene and was low yielding, but successful.

An α-hydroxyketone (**19**) was also synthesized from 2-adamantanone (**15**; Fig. 6.7). Reduction of the nitrile with methyl Grignard was unsuccessful, necessitating an alternative route. First, 2-adamantanone (**15**) was reduced with ethynyl magnesium bromide giving 2-ethynyl-2-adamantanol (**18**) in good yield. The oxidation of **18** with catalytic mercury (II) to

[2] The earliest attempts at Wittig olefination as well as OsO_4-oxidation were performed by Dr. Indika Bandara and supplied by Prof. John Dawson (University of South Carolina).

produce the desired α-hydroxyketone **18** was low yielding; however, the remaining starting material (**18**) was also recovered.

6.2.3. Molecular dynamics supports active site mobility as foremost problem[3]

To investigate how various α-hydroxyketones were oriented in the active site, **1**, **16** (an α-hydroxyketone derived from *R*-(+)-camphor), and **12** (a norbornane α-hydroxyketone) were docked into the oxy-ferrous, wild type $P450_{cam}$ crystal structure (27). The percentage of orientations meeting the constraints predicted by Bonomo *et al*. (16) for C-C lyase by the peroxoanion mechanism in CYP17A1 (Fig. 6.1) were used to predict the likelihood of C-C scission: distance between the proximal, heme-bound oxygen atom and the hydrogen of the substrate alcohol < 3.62 Å; distance from the distal oxygen atom to the carbonyl carbon < 3.47 Å. Because this predicted geometry from DFT calculations were established in the literature, this method for identifying catalytically competent orientations was selected. Moreover, the likelihood for a positioning conducive to C-C lyase by the Compound I-mediated mechanism was inferred from these same simulations by applying the distance constraint from the oxygen atom proximal to the heme to the substrate alcohol (< 3.62 Å to the hydrogen atom).

After 400-3000 ns of molecular dynamics simulations, looking at the distribution of the docked molecules orientations revealed that none of these substrates yielded populations that matched the geometric constraints for the peroxyanion-mediated C-C lyase reaction (Fig. 6.8; Table 6.1). Instead, the most populated distribution of each molecule had the α-hydroxyketone H-bonding to Tyr96 and Thr101 and thus far away from the oxygen-bound heme. Intriguingly,

[3] Dr. Mark McLean (University of Illinois) performed the docking and molecular dynamics work.

the smaller cyclopentyl and norbornyl α-hydroxyketones has sizable populations with the alcohol positioned for reacting with Compound I.

Before generating, expressing, and purifying Tyr96 and Thr101 mutants of $P450_{cam}$, the same docking and molecular dynamics regimen was applied after generating Tyr96Phe, Thr101Val, and Thr252Ala replacements of $P450_{cam}$ in the computer models to evaluate the helpfulness of these mutations. The subsequent distributions for **1**, **3**, and **4** (Fig. 6.8) showed populations with these α-hydroxyketones nearer the geometry required for the peroxyanion mechanism as well as the Compound I mechanism. Examination of snapshots from the MD simulations presented orientations in which the distal oxygen was within 3.5 Å of the ketone with the alcohol in position to H-bond to the proximal oxygen as in the proposed transition state (Fig. 6.1).

Site-directed mutagenesis of Tyr96Phe and Thr101Val in $P450_{cam}$ reduced enzyme stability in the presence of oxygen. The decreased stability of such mutants limited the assays employed. Regardless, when given the camphor- (**8**) and adamantanone- (**16**) derived α-hydroxyketones did not appear to undergo C-C lyase reactions (Fig. 6.9).

6.3. DISCUSSION

Whether attached to a cyclopentyl, cyclohexyl, camphor ([2.2.1]-bicyclo), or adamantyl structure, $P450_{cam}$ and $P450_{BM3}$ were unable to transform these α-hydroxyketones by C-C scission to a ketone and carboxylic acid. Although molecular dynamics simulations suggested mutating Tyr96Phe and Thr101Val in $P450_{cam}$ would better position these compounds for reacting with the peroxyanion, *in vitro* assays with the mutant P450s likewise yielded no evidence of lyase reactivity.

The molecular dynamics simulations supply several reasons why coaxing $P450_{cam}$ into C-C scission of α-hydroxyketones was unsuccessful: First, although the α-hydroxyketone moiety is closer to the dioxygen species more frequently, the distribution of small molecule orientations within the active site features multiple conformations that frequently occur besides those geometrically oriented for catalysis. Second, the substrates are more frequently in catalytically suggestive orientations in the triple-mutant simulations, but the lifetimes of such orientations are unclear. Following the peroxyanion-mediated mechanism, formation of the ordered, six-membered transition state likely requires considerable residence time for the α-hydroxyketone near the heme. Third, although the distance between the proximal oxygen bound to the heme and the hydrogen of the substrate alcohol shifted towards the 3.62 Å cutoff for catalysis, the distribution of distal oxygen-carbonyl carbon distances shifted only slightly towards small distances and infrequently dropped below the 3.47 Å cutoff. These results together suggest that, although these compounds bind to $P450_{cam}$, the α-hydroxyketone moiety is not positioned sufficiently close to the reactive oxygen species bound to heme or does not remain in favorable orientations long enough to facilitate C-C scission.

The closer positioning of the alcohol of each docked molecule to the proximal oxygen throughout the Tyr96Phe, Thr101Val, Thr252Ala triple-mutant simulations hints that the Compound I-mediated mechanism is not at work. This distribution presumably would remain were Compound I or some similar monooxygen species modeled instead of the oxy-complex. If this assumption holds true, hydrogen abstraction from the alcohol by this ferryl species is feasible for a large proportion of the observed substrate orientations. That these α-hydroxyketones appeared to be better positioned for C-C lyase reactivity *via* the Compound I-

mediated mechanism than by the peroxyanion-mediated one and that no such activity was observed portends poorly for the Compound I-based mechanism.

The difficulties and general lack of success in coaxing $P450_{cam}$ and $P450_{BM3}$ into C-C scission of α-hydroxyketones highlights the uniqueness of CYP17A1. Whereas 17-α-hydroxy-pregnenolone and 17-α-hydroxy-progesterone are held in place by Asn202 and Arg239 in CYP17A1 (16), the compounds supplied to the bacterial P450s studied here had no comparable interactions to hold the α-hydroxyketone in position for catalysis. By switching to another bacterial P450 able to hold an α-hydroxyketone over the heme in the active site (CYP199A4), this hypothesis is being examined—the subject of Chapter 7.

6.4. EXPERIMENTAL

6.4.1. Protein expression and purification

CYP101A1, its redox partners putidaredoxin (Pd) and putidaredoxin reductase (PdR), and the Thr252Ala mutant were expressed and purified according to reported protocols (28–30). Yelena Grinkova (University of Illinois) created the Tyr96Phe, Thr101Val mutant using site directed mutagenesis, expressed, and purified the protein. CYP102A1 was likewise expressed and purified according to established protocols (31).

6.4.2. Substrate-induced spectral shift assays

Camphor in the storage buffer was separated from CYP101A1 by size exclusion chromatography using Superfine G25 Sephadex and 50 mM Tris-HCl (pH 7.4). The P450 was diluted to 2 μM in 50 mM Tris-HCl (pH 7.4), 150 mM KCl, and the UV-Vis spectra recorded. Aliquots of the compound being assays were added, spectra recorded, and the process repeated for the whole dilution series. Difference spectra of the substrate-bound minus substrate-free spectra were used to calculate K_s by fitting the spin-shift (ΔAbs_{389nm}-ΔAbs_{417nm}) versus

compound concentration to a single-binding model (ΔAbs_{389nm}-ΔAbs_{417nm} = (ΔAbs_{max} * [compound]) / K_s + [compound]) using OriginPro2020.

6.4.3. Substrate, NADH turnover with CYPs

Substrate turnover and NADH consumption assays were conducted as previously reported (30). After quenching reactions with equivolume 2 M KOH and adding an internal standard (5-*endo*-bromocamphor), the reaction mixture was extracted with with DCM or EtOAc and analyzed by GC-FID using a DB-17 capillary column (30 m x 0.320 mm x 0.25 μm). Reactions from **1** or **2** utilized the following oven program: 40°C, 5 min.; ramp 5°C min^{-1} to 110°C; ramp 10°C min^{-1} to 230°C; 230°C, 3 min. Reactions with [2.2.1]-bicyclo- or adamantyl-substrates were separated using the following oven program: 70°C, 2 min; ramp 10°C min^{-1} to 230°C; 230°C, 5 min.

6.4.4. Synthesis of α-hydroxyketones

All reagents and solvents were purchased from commercial suppliers and used without further purification unless otherwise noted. Hexanes were vacuum distilled to remove high boiling impurities. When dry solvents were required, they were dried for two or more days over activated 3Å molecular sieves and stored under dry nitrogen gas. Thin layer chromatography utilized silica gel 60 doped with a fluorophore excited at 254 nm. Spots were visualized by fluorescent quenching and $KMnO_4$ staining. Flash column chromatography was conducted with 40-63 micron silica gel. ^1H NMR and ^{13}C NMR spectra were recorded on a 500 MHz spectrometer at the University of Illinois School of Chemical Sciences NMR Spectroscopy Lab. NMR samples were made as solutions in $CDCl_3$ using the residual solvent peak as the internal standard; δ values are given in ppm and coupling constants (*J*) in hertz (Hz). Mass spectra were

obtained from a high-resolution ESI mass spectrometer at the University of Illinois School of Chemical Sciences Mass Spectrometry Lab.

6.4.4.1. Cyanosilylation of camphor. Under dry nitrogen gas, 2.2 mL pentane was added to (*R*)-(+)-camphor (**14**; 405 mg, 2.63 mmol) and ZnI_2 (47 mg, 0.15 mmol, 0.05 eq.). The resulting solution was cooled in an ice bath before trimethylsilyl cyanide (TMSCN, 344 mg, 3.47 mmol, 1.3 eq.) was added dropwise. The reaction was then removed from ice and left to stir overnight under nitrogen atmosphere. Water was added to quench the reaction before extracting with diethyl ether. The combined organic layers were dried over $MgSO_4$, and the solvent removed by rotary evaporation to yield 669 mg **15** (99% yield).

6.4.4.2. Reduction and deprotection of 15. In an oven-dried round bottom flask fitted with a condenser under nitrogen atmosphere, 350 mg **15** (1.39 mmol) was dissolved in 0.5 mL THF before adding 1.8 mL toluene. This solution was cooled in an ice before adding 1.1 mL 3 M CH_3MgBr in diethyl ether (3.2 mmol, 2.3 eq.) dropwise. The ice bath was replaced with a sand bath heated to 80°C for refluxing over 9 hours. The sand bath was replaced with an ice bath, and the reaction quenched with 11 mL 3 N HCl added dropwise. The resulting mixture was heated to 40°C overnight to remove the silyl ether. After dilution with 20 mL water, the sample was washed thrice with 20 mL diethyl ether. The combined organic layers were washed once each with 20mL sat. $NaHCO_3$ and brine, dried over $MgSO_4$, and the solvent removed under vacuum. Flash column chromatography with 10-20% EtOAc in hexanes gave **16** (25 mg, 0.13 mmol, 10% yield; R_f 0.19, 10% EtOAc in hexanes). ^{1}H NMR (499 MHz, $CDCl_3$) δ 2.68 (s, 1H), 2.27 (s, 3H), 1.89 – 1.81 (m, 2H), 1.77 – 1.68 (m, 1H), 1.44 – 1.38 (m, 1H), 1.23 (ddd, J = 12.6, 9.1, 5.0 Hz, 1H), 1.11 (s, 3H), 1.03 – 0.96 (m, 2H), 0.94 (s, 3H), 0.85 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ

10.58, 20.49, 20.92, 26.66, 27.82, 30.32, 41.21, 45.26, 50.43, 52.47, 87.67, 212.32. Spectra were in accord with previous reports (32–35).

6.4.4.3. Reduction of 2-adamantanone with ethynyl Grignard. 502 mg 2-adamantanone (**17**; 3.33 mmol) were dissolved in 3.35 mL dry THF in an oven-dried round bottom flask under dry nitrogen atmosphere and cooled in an ice bath. Ethynyl magnesium bromide (0.5 M in THF; 10 mL, 5.0 mmol, 1.5 eq.) was added dropwise before removing the ice bath and stirring at room temperature for 4 hours. The reaction was quenched by dropwise addition of 13 mL 0.5 M HCl (6.5 mmol, 2.0 eq.). The reaction mixture was washed thrice with 50 mL of ethyl acetate, the combined organic layers dried over $MgSO_4$, and the solvent removed by rotary evaporation. Flash column chromatography with 20% EtOAc in hexanes to yield **18** (496 mg, 2.81 mmol, 84% yield). ^{1}H NMR (499 MHz, $CDCl_3$) δ 2.53 (s, 1H), 2.20 – 2.11 (m, 5H), 1.96 (t, *J* = 3.1 Hz, 3H), 1.93 (s, 1H), 1.82 (dq, *J* = 13.0, 2.9 Hz, 3H), 1.77 (q, *J* = 3.3 Hz, 3H), 1.70 (d, *J* = 3.3 Hz, 3H), 1.60 – 1.53 (m, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 26.85, 26.96, 31.64, 35.47, 37.69, 38.85, 72.63, 72.79, 72.82, 76.90, 77.16, 77.41, 88.59. HR-ESI-MS *m/z* $[M+H]^+$, calcd for $C_{12}H_{18}O$, 177.1279; found, 177.1276. Spectra were in accord with previous reports (36–39).

6.4.4.4. Oxymercuration of 18 yields α-hydroxyketone 19. Alkyne **18** (385 mg, 2.18 mmol) was dissolved in 1.4 mL THF. $HgSO_4$ (87 mg, 0.29 mmol, 0.13 eq.) was suspended in 0.68 mL water before adding 0.34 mL conc. H_2SO_4 and 1.4 mL THF slowly. This sulfate solution was added dropwise to the solution containing **19** and stirred at room temperature for 20 hours. The reaction mixture was diluted with 40 mL water and neutralized with 50 mL sat. $NaHCO_3$. The aqueous solution was washed thrice with 50 ml diethyl ether before the combined organic layers were washed once with 100 mL brine, dried over $MgSO_4$, and the solvent

removed under vacuum. Flash column chromatography with 30% EtOAc in hexanes yielded α-hydroxyketone **19** (109 mg, 0.56 mmol, 26% yield). Starting material **18** (220 mg, 1.25 mmol) was also recovered (83% net recovery). ^{1}H NMR (499 MHz, $CDCl_3$) δ 2.22 (s, 3H), 2.19 – 2.14 (m, 5H), 1.85 – 1.76 (m, 5H), 1.75 (s, 1H), 1.72 – 1.60 (m, 8H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 23.86, 26.81, 26.94, 32.52, 33.74, 34.78, 37.29, 76.91, 77.16, 77.42, 79.75, 209.46. HR-ESI-MS *m/z* $[M+H]^+$, calcd for $C_{12}H_{19}O_2$, 195.1385; found, 195.1385. Spectra were in accord with a previous report (37).

6.4.5. Docking and molecular dynamics simulations[4]

The ferrous-oxy heme structure of CYP101A1 (PDB: 2A1M) (27) was selected for docking. **1**, **12**, and **16** were independently docked into the active site of this equilibrated structure using MOE (40). After equilibration of the docked α-hydroxyketones using the Amber force field ff19SB (41) to 0.01 kcal mol^{-1} A^{-2}, the resulting molecular models were subjected to MD simulations.

The P450 MD simulations were set up using AmberTools 20 (see ref at bottom) using Amber force field ff19SB (41). Heme parameters were obtained from Shahrokh *et al*. (42). Substrate parameters were generated from the respective mol2 files using the General Amber Force Field (43). Point mutations were generated using AmberTools (44) prior to minimization and equilibration without changing the docking mode of the α-hydroxyketone. Systems were solvated using an octahedral periodic boundary with a 12 Å spacing. Potassium and chloride ions were added to neutralize the system, bringing the KCl concentration to 0.15 M.

[4] I performed the initial docking and screening of the α-hydroxyketones into CYP101A1. Dr. Mark McLean (University of Illinois) performed the molecular dynamics simulations.

MD simulations were performed using GROMACS 2020 (45). The solvated systems were first energy minimized using a steepest decent algorithm, followed by sequential equilibration runs, constraining the protein backbone before slowly releasing the constraints on the sidechains and substrates. Equilibration runs employed a modified Berendsen thermostat and a Berendsen barostat for temperature and pressure coupling. For production runs, temperature and pressure couplings were achieved using a modified Berendsen thermostat and a Parrinello-Rahman barostat, respectively. Electrostatics were calculated using Particle Mesh Ewald. For production runs the time step was 2 femtoseconds with coordinates saved every 10 ps. Runs continued until the accumulated simulation time reached ~0.4 μs (for **1**), ~2 μs (for **12**) or ~1.5 μs (for **16**).

6.5. TABLE AND FIGURES

Table 6.1. Frequency with which α-hydroxyketones are positioned as necessary for catalysis by the given mechanism in molecular dynamics simulations of CYP101A1.

CYP101A1	Compound	Peroxyanion[a]		Compound I[b]	
		N	Percentage	*N*	Percentage
Wild Type	**1**[c]	37	0.08%	4,969	10.3%
	12[d]	1,067	0.53%	57,451	28.5%
	16[e]	262	0.18%	4,716	3.15%
Tyr96Phe, Thr101Val, Thr252Ala	**1**[f]	35	0.11%	14,117	44.3%
	12[d]	2,497	1.23%	72,387	35.9%
	16[g]	31	0.02%	17,905	12.8%

[a] Proximal oxygen within 3.62 Å of alcohol hydrogen ***and*** distal oxygen within 3.47 Å of carbonyl carbon.
[b] Proximal oxygen within 3.62 Å of alcohol hydrogen.
[c] 48,367 frames sampled at 10 ps per frame; ~0.5 μs total
[d] 201,889 frames samples at 10 ps per frame; ~2 μs total
[e] 149,295 frames samples at 10 ps per frame; ~1.5 μs total
[f] 31,875 frames samples at 10 ps per frame; ~0.3 μs total
[g] 139,804 frames samples at 10 ps per frame; ~1.4 μs total

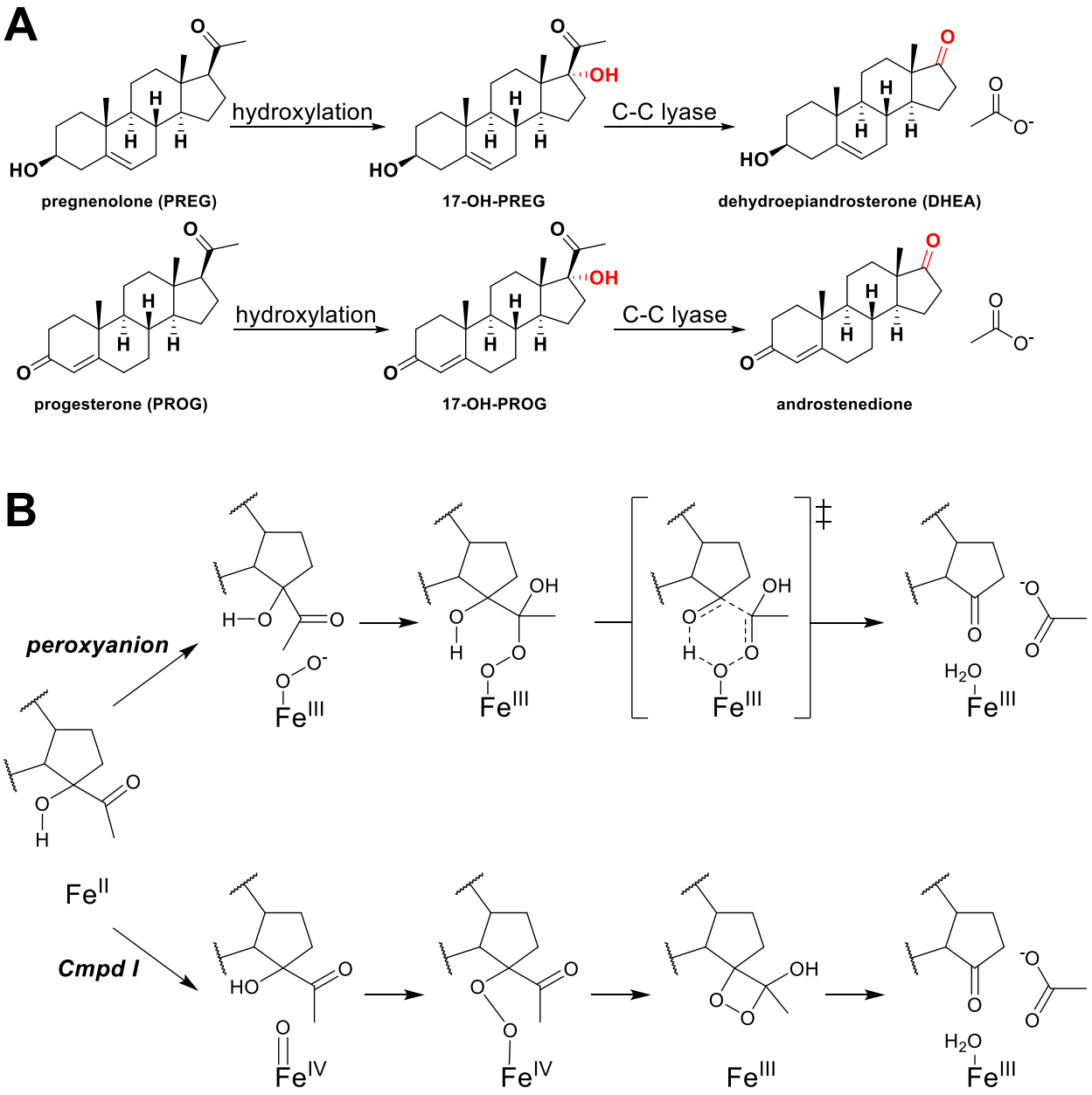


Figure 6.1. Multifunctional CYP17A1 catalyzes C17-hydroxylation and C17-C20 scission. (A) CYP17A1 first hydroxylates at C17 pregnenolone or progesterone before breaking the C17-C20 bond to form dehydroandrosterone and androstenedione, respectively. (B) Two mechanisms remain with experimental support after an array of experiments into the C-C lyase reaction catalyzed by CYP17A1. One is facilitated by the peroxyanion intermediate (top); the other through the oxo ferryl species known as Compound I (bottom).

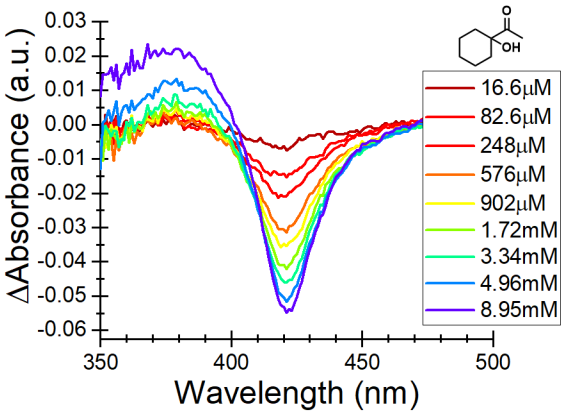


Figure 6.2. P450$_{cam}$ Type I binding spectra induced by cyclohexyl α-hydroxyketone. Successive aliquots of **2** were added at the concentrations shown (16.6 μM-8.95 mM) to 2 μM CYP101A1. The spectrum with no substrate was subtracted from the spectra in the presence of **2** to produce the difference spectra shown.

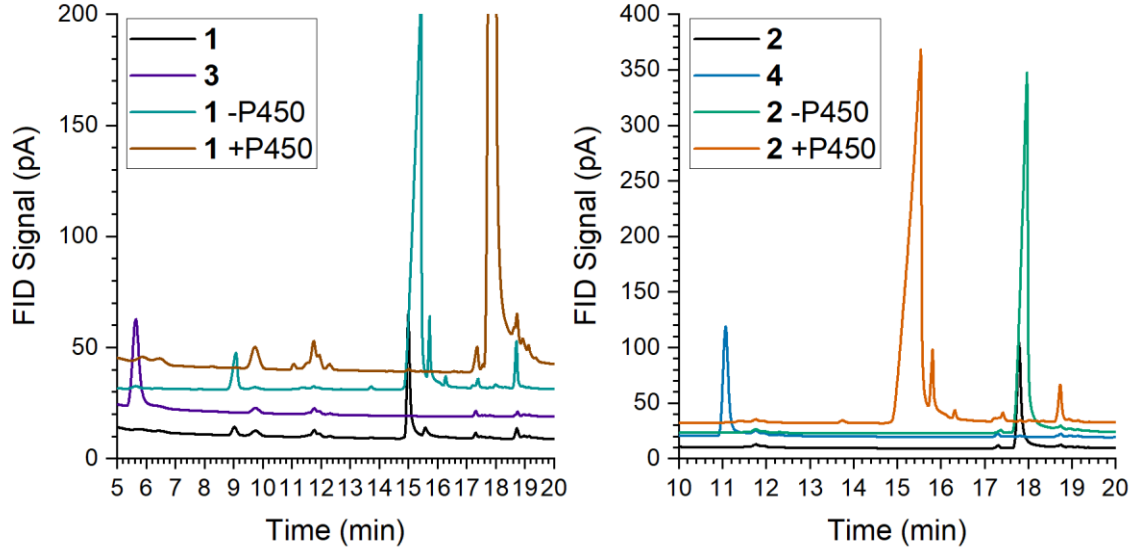


Figure 6.3. Gas chromatographs of P450$_{cam}$ reactions with cyclopentyl, cyclohexyl α-hydroxyketones.
0.5 μM CYP101A1, 1.1 μM PdR, and 5 μM Pd were incubated in 50 mM KPO_4 (pH 7.95), 100 mM KCl in the presence of 500 μM **1** and **2**, respectively. Reactions commenced with the addition of NADH to ~1 mM and were quenched with equivolume 2 M KOH before extraction with EtOAc and GC-FID analysis.

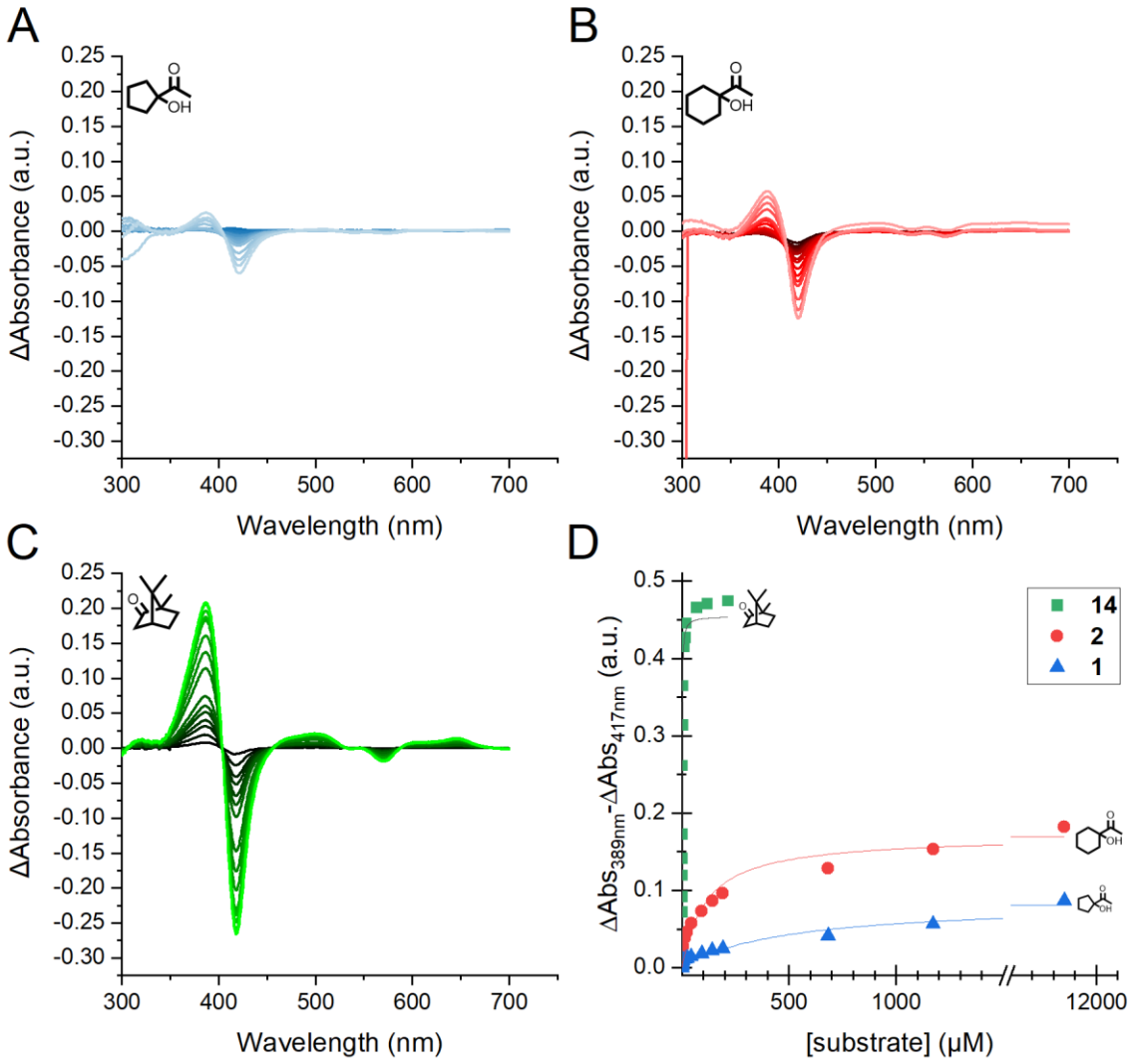


Figure 6.4. P450$_{cam}$ Thr252Ala Type I binding spectra induced by cyclopentyl, cyclohexyl α-hydroxyketones.
Successive aliquots of **1** (A), **2** (B), or **14** (C) were added over a range of concentrations (0.1 μM-12 mM for **1** and **2**; 10 nm – 220 μM for **14**) to 2 μM CYP101A1. The spectrum with no substrate was subtracted from the spectra in the presence of these compounds to produce the difference spectra shown. (D) Spin-shift (ΔAbs_{389nm} - ΔAbs_{417nm}) versus substrate concentration plot used to determine binding constant (K_s) by non-linear curve fitting in Origin2020.

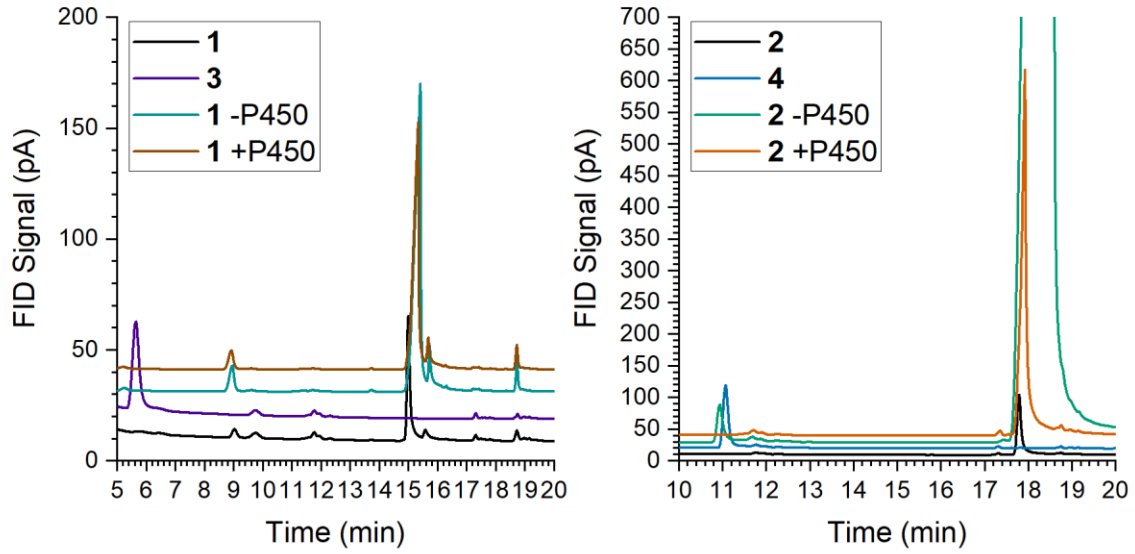


Figure 6.5. Gas chromatographs of P450$_{cam}$ Thr252Ala reactions with cyclopentyl, cyclohexyl α-hydroxyketones.
0.5 μM CYP101A1, 1.1 μM PdR, and 5 μM Pd were incubated in 50 mM KPO_4 (pH 7.95), 100 mM KCl in the presence of 500 μM **1** and **2**, respectively. Reactions commenced with the addition of NADH to ~1 mM and were quenched with equivolume 2 M KOH before extraction with EtOAc and GC-FID analysis.

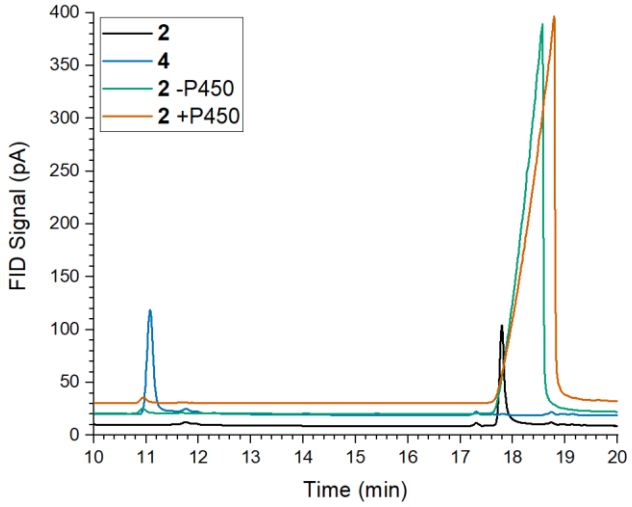


Figure 6.6. Gas chromatographs of P450$_{BM3}$ reactions with cyclohexyl α-hydroxyketone.
0.5 μM CYP102A1 was incubated in 50 mM KPO_4 (pH 7.95), 100 mM KCl in the presence of 500 μM **2**. Reactions commenced with the addition of NADPH to ~1 mM and were quenched with equivolume 2 M KOH before extraction with EtOAc and GC-FID analysis.

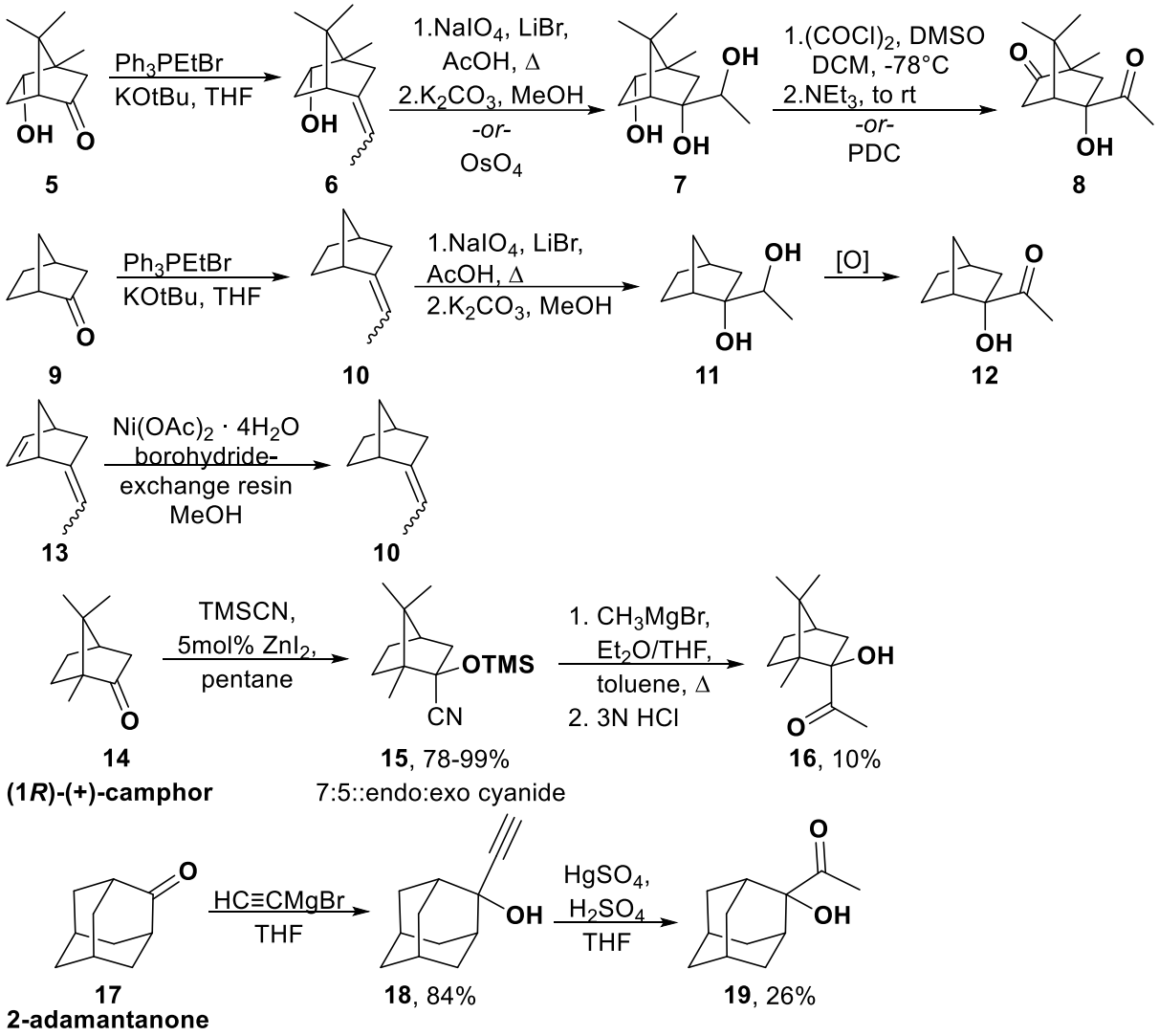


Figure 6.7. Synthetic routes to [2.2.1]-bicyclo and adamantyl α-hydroxyketones.

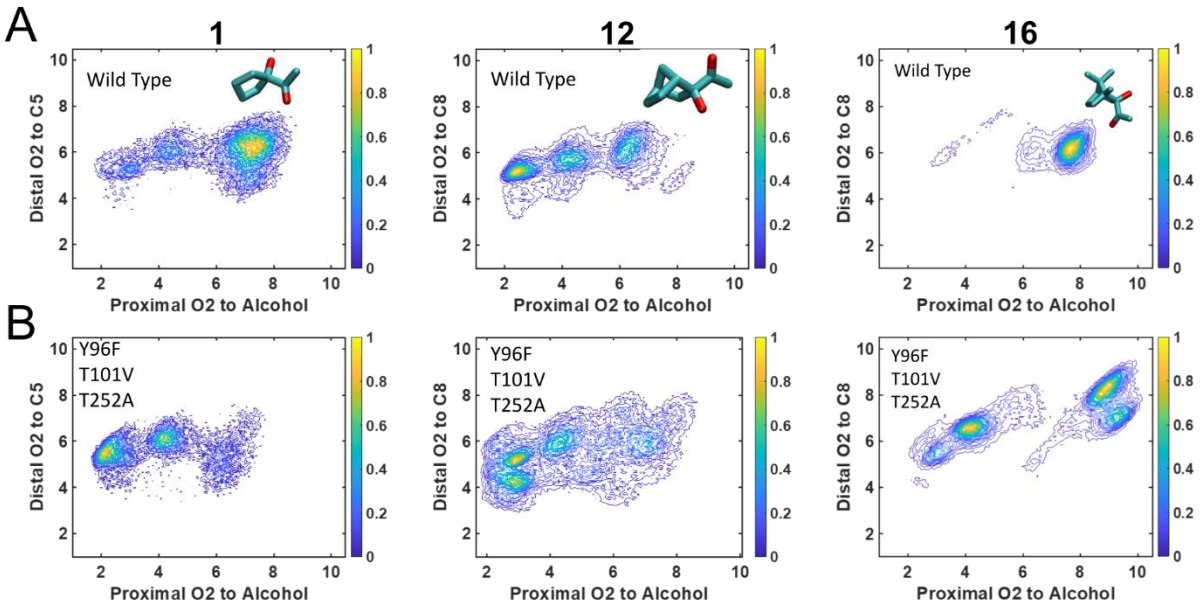


Figure 6.8. Proximity of α-hydroxyketone moiety to oxygen-bound heme from molecular dynamics simulations.
Histograms showing the distribution of α-hydroxyketone orientations throughout MD simulations by measuring the distance from the proximal oxygen bound to the heme to the alcohol hydrogen and distal oxygen to the carbonyl carbon for wild type (A) and Tyr96Phe, Thr101Val, Thr252Ala triple-mutant (B) CYP101A1.

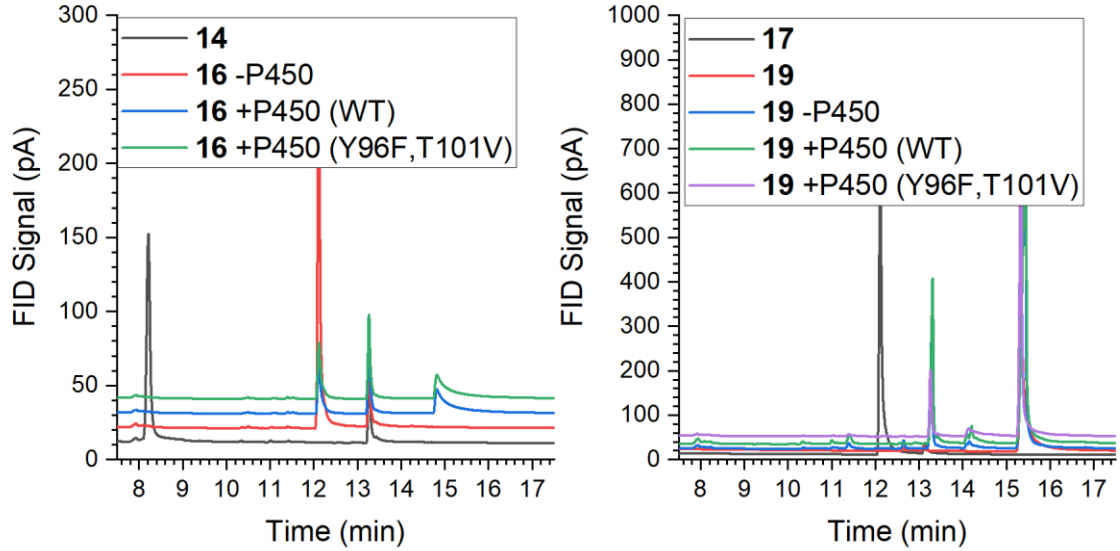


Figure 6.9. Gas chromatographs of P450$_{cam}$ reactions with camphor- and adamantanone-derived α-hydroxyketones.
0.5 μM CYP101A1, 1.1 μM PdR, and 5 μM Pd were incubated in 50 mM KPO_4 (pH 7.95), 100 mM KCl in the presence of 500 μM **16** and **19**, respectively. Reactions commenced with the addition of NADH to ~1 mM and were quenched with equivolume 2 M KOH before extraction with EtOAc and GC-FID analysis.

6.6. REFERENCES

1. Guengerich, F. P., and Yoshimoto, F. K. (2018) Formation and cleavage of C–C bonds by enzymatic oxidation–reduction reactions. *Chem. Rev.* **118**, 6573–6655
2. Auchus, R. J., and Miller, W. L. (2015) P450 enzymes in steroid processing. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (Ortiz de Montellano, P. R. ed), pp. 851–879, Springer International Publishing, Cham, 10.1007/978-3-319-12108-6_12
3. Gomez, L., Kovac, J. R., and Lamb, D. J. (2015) CYP17A1 inhibitors in castration-resistant prostate cancer. *Steroids*. **95**, 80–87
4. Njar, V. C. O., and Brodie, A. M. H. (2015) Discovery and development of galeterone (TOK-001 or VN/124-1) for the treatment of all stages of prostate cancer. *J. Med. Chem.* **58**, 2077–2087
5. Akhtar, M., Corina, D., Miller, S., Shyadehi, A. Z., and Wright, J. N. (1994) Mechanism of the acyl-carbon cleavage and related reactions catalyzed by multifunctional P-450s: Studies on cytochrome P-$450_{17\alpha}$. *Biochemistry*. **33**, 4410–4418
6. Lee-Robichaud, P., Shyadehi, A. Z., Wright, J. N., Akhtar, M., and Akhtar, M. (1995) Mechanistic kinship between hydroxylation and desaturation reactions: Acyl-carbon bond cleavage promoted by pig and human CYP17 (P-$450_{17\alpha}$; 17α-hydroxylase-17,20-lyase). *Biochemistry*. **34**, 14104–14113
7. Yoshimoto, F. K., Gonzalez, E., Auchus, R. J., and Guengerich, F. P. (2016) Mechanism of 17α,20-lyase and new hydroxylation reactions of human cytochrome P450 17A1: ^{18}O labeling and oxygen surrogate evidence for a role of a perferryl oxygen. *J. Biol. Chem.* **291**, 17143–17164
8. Gregory, M., Mak, P. J., Sligar, S. G., and Kincaid, J. R. (2013) Differential hydrogen bonding in human CYP17 dictates hydroxylation versus lyase chemistry. *Angew. Chem. Int. Ed.* **52**, 5342–5345
9. Mak, P. J., Gregory, M. C., Denisov, I. G., Sligar, S. G., and Kincaid, J. R. (2015) Unveiling the crucial intermediates in androgen production. *Proc. Natl. Acad. Sci.* **112**, 15856–15861
10. Mak, P. J., Duggal, R., Denisov, I. G., Gregory, M. C., Sligar, S. G., and Kincaid, J. R. (2018) Human cytochrome CYP17A1: The structural basis for compromised lyase activity with 17-hydroxyprogesterone. *J. Am. Chem. Soc.* **140**, 7324–7331
11. Liu, Y., Denisov, I. G., Grinkova, Y. V., Sligar, S. G., and Kincaid, J. R. (2020) P450 CYP17A1 variant with a disordered proton shuttle assembly retains peroxo-mediated lyase efficiency. *Chem. – Eur. J.* **26**, 16846–16852
12. Liu, Y., Denisov, I. G., Sligar, S. G., and Kincaid, J. R. (2021) Substrate-specific allosteric effects on the enhancement of CYP17A1 lyase efficiency by cytochrome b_5. *J. Am. Chem. Soc.* **143**, 3729–3733
13. Gregory, M. C., Denisov, I. G., Grinkova, Y. V., Khatri, Y., and Sligar, S. G. (2013) Kinetic solvent isotope effect in human P450 CYP17A1-mediated androgen formation: Evidence for a reactive peroxoanion intermediate. *J. Am. Chem. Soc.* **135**, 16245–16247
14. Gonzalez, E., Johnson, K. M., Pallan, P. S., Phan, T. T. N., Zhang, W., Lei, L., Wawrzak, Z., Yoshimoto, F. K., Egli, M., and Guengerich, F. P. (2018) Inherent steroid 17α,20-lyase activity in defunct cytochrome P450 17A enzymes. *J. Biol. Chem.* **293**, 541–556

15. Vogt, C. D., Bart, A. G., Yadav, R., Scott, E. E., and Aubé, J. (2021) Effects of fluorine substitution on substrate conversion by cytochromes P450 17A1 and 21A2. *Org. Biomol. Chem.* **19**, 7664–7669
16. Bonomo, S., Jørgensen, F. S., and Olsen, L. (2017) Mechanism of cytochrome P450 17A1-catalyzed hydroxylase and lyase reactions. *J. Chem. Inf. Model.* **57**, 1123–1133
17. Mueller, E. J., Loida, P. J., and Sligar, S. G. (1995) Twenty-five years of P450cam research. in *Cytochrome P450: Structure, Mechanism, and Biochemistry* (de Montellano, P. R. O. ed), pp. 83–124, Springer US, Boston, MA, 10.1007/978-1-4757-2391-5_3
18. Shoji, O., Aiba, Y., and Watanabe, Y. (2019) Hoodwinking cytochrome $P450_{BM3}$ into hydroxylating non-native substrates by exploiting its substrate misrecognition. *Acc. Chem. Res.* **52**, 925–934
19. Imai, M., Shimada, H., Watanabe, Y., Matsushima-Hibiya, Y., Makino, R., Koga, H., Horiuchi, T., and Ishimura, Y. (1989) Uncoupling of the cytochrome P-450cam monooxygenase reaction by a single mutation, threonine-252 to alanine or valine: possible role of the hydroxy amino acid in oxygen activation. *Proc. Natl. Acad. Sci.* **86**, 7823–7827
20. Martinis, S. A., Atkins, W. M., Stayton, P. S., and Sligar, S. G. (1989) A conserved residue of cytochrome P-450 is involved in heme-oxygen stability and activation. *J. Am. Chem. Soc.* **111**, 9252–9253
21. Poulos, T. L., Finzel, B. C., and Howard, A. J. (1987) High-resolution crystal structure of cytochrome P450cam. *J. Mol. Biol.* **195**, 687–700
22. White, R. E., McCarthy, M.-B., Egeberg, K. D., and Sligar, S. G. (1984) Regioselectivity in the cytochromes P-450: Control by protein constraints and by chemical reactivities. *Arch. Biochem. Biophys.* **228**, 493–502
23. Emmanuvel, L., Shaikh, T. M. A., and Sudalai, A. (2005) $NaIO_4$/LiBr-mediated diastereoselective dihydroxylation of olefins: A catalytic approach to the Prevost−Woodward reaction. *Org. Lett.* **7**, 5071–5074
24. Choi, J., and Yoon, N. M. (1996) An excellent nickel boride catalyst for the selective hydrogenation of olefins. *Synthesis*. **1996**, 597–599
25. Hong, R., Chen, Y., and Deng, L. (2005) Catalytic enantioselective total syntheses of bisorbicillinolide, bisorbicillinol, and bisorbibutenolide. *Angew. Chem. Int. Ed.* **44**, 3478–3481
26. Martínez, A. G., Vilar, E. T., Fraile, A. G., de la Moya Cerero, S., de Diego, J. M. G.-F., and Subramanian, L. R. (1994) A facile synthesis of homochiral 1-norbornanecarboxylic acids and 1-norbornanecarbonitriles. *Tetrahedron Asymmetry*. **5**, 1599–1603
27. Nagano, S., and Poulos, T. L. (2005) Crystallographic study on the dioxygen complex of wild-type and mutant cytochrome P450cam. Implications for the dioxygen activation mechanism. *J. Biol. Chem.* **280**, 31659–31663
28. Gunsalus, I. C., and Wagner, G. C. (1978) Bacterial P-450cam methylene monooxygenase components: Cytochrome m, putidaredoxin, and putidaredoxin reductase. in *Methods in Enzymology* (Fleischer, S., and Packer, L. eds), pp. 166–188, Biomembranes - Part C: Biological Oxidations, Academic Press, **52**, 166–188
29. Raag, R., Martinis, S. A., Sligar, S. G., and Poulos, T. L. (1991) Crystal structure of the cytochrome $P\text{-}450_{CAM}$ active site mutant Thr252Ala. *Biochemistry*. **30**, 11420–11429

30. Makris, T. M., von Koenig, K., Schlichting, I., and Sligar, S. G. (2007) Alteration of P450 distal pocket solvent leads to impaired proton delivery and changes in heme geometry. *Biochemistry*. **46**, 14129–14140
31. Yeom, H., Sligar, S. G., Li, H., Poulos, T. L., and Fulco, A. J. (1995) The role of Thr268 in oxygen activation of cytochrome $P450_{BM-3}$. *Biochemistry*. **34**, 14733–14740
32. Palomo, C., Oiarbide, M., Aizpurua, J. M., González, A., García, J. M., Landa, C., Odriozola, I., and Linden, A. (1999) Highly diastereoselective aldol reactions with camphor-based acetate enolate equivalents. *J. Org. Chem.* **64**, 8193–8200
33. Palomo, C., Oiarbide, M., Landa, A., González-Rego, M. C., García, J. M., González, A., Odriozola, J. M., Martín-Pastor, M., and Linden, A. (2002) Design and synthesis of a novel class of sugar-peptide hybrids: C-linked glyco β-amino acids through a stereoselective "acetate" Mannich reaction as the key strategic element. *J. Am. Chem. Soc.* **124**, 8637–8643
34. García, J. M., Odriozola, J. M., Lecumberri, A., Razkin, J., and González, A. (2008) A concise and efficient route to the Alzheimer's therapeutic agent (R)-arundic acid. *Tetrahedron*. **64**, 10664–10669
35. Koval'skaya, S. S., Kozlov, N. G., and Dikusar, E. A. (2010) Rearrangements and hydration of 2-ethynylisoborneol and 2-ethynylisocamphanol. *Russ. J. Org. Chem.* **46**, 1493–1502
36. Bovonsombat, P., and Nelis, E. M. (1995) Formation of a dihaloenone of homoadamantanone. *Synth. Commun.* **25**, 1223–1229
37. Koval'skaya, S. S., Kozlov, N. G., and Dikusar, E. A. (2002) Synthesis of 1-Acetamido-2-acetyladamantane. *Russ. J. Org. Chem.* **38**, 494–498
38. Battilocchio, C., Baxendale, I. R., Biava, M., Kitching, M. O., and Ley, S. V. (2012) A flow-based synthesis of 2-aminoadamantane-2-carboxylic acid. *Org. Process Res. Dev.* **16**, 798–810
39. Sum, Y. N., Yu, D., and Zhang, Y. (2013) Synthesis of acetylenic alcohols with calcium carbide as the acetylene source. *Green Chem.* **15**, 2718–2721
40. Rupasinghe, S., and Schuler, M. A. (2006) Homology modeling of plant cytochrome P450s. *Phytochem. Rev.* **5**, 473–505
41. Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., and Simmerling, C. (2020) ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552
42. Shahrokh, K., Orendt, A., Yost, G. S., and Cheatham III, T. E. (2012) Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **33**, 119–133
43. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174
44. Case, D. A., Belfon, K. A. A., Ben-Shalom, I., Brozell, S. R., Cerutti, D., Cheatham III, T. E., Cruzeiro, V. W. D., Darden, T., Duke, R. E., Giambasu, G., Gilson, M. K., Gohlke, H., Götz, A. W., Harris, R., Izadi, S., Izmailov, S. A., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Man, V., Merz, K. M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., Pan, F., Pantano, S., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J., Simmerling, C., Skrynnikov, N. R., Smith, J., Swails, J., Walker, R. C., Wang, J., Wang, J.,

Wilson, L., Wolf, R. M., Wu, X., Xiong, Y., Xue, Y., York, D. M., and Kollman, P. A. (2020) *AMBER 2020*, University of California, San Francisco

45. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25

CHAPTER 7: INVESTIGATING C-C LYASE REACTIONS OF α-HYDROXY KETONES USING BACTERIAL P450s: PART 2, ONGOING WORK WITH CYP199A4

Investigating the structural and mechanistic underpinnings of the C-C lyase reactivity of CYP17A1 by coaxing bacterial P450s to do the same reaction presents the ability to deploy biochemical and biophysical experiments not currently possible with CYP17A1. Earlier work with bacterial CYP101A1 and CYP102A1 failed to produce this reactivity, however, even with α-hydroxyketones constructed on the scaffold of known substrates. Molecular dynamics simulations of CYP101A1 suggested that substrate mobility was too great within the active site and that the α-hydroxyketone moiety was poorly positioned, preventing the desired C-C scission reactivity. Desiring a P450 having available crystal structure information, easy expression and purification, known and varied substrate scope, and rigid binding of these substrates with a single region held over the heme, I proposed CYP199A4 as a better candidate for studying the C-C lyase reaction. After docking several possible substrates for the 4-substituted benzoate common to substrates of CYP199A4, molecular dynamics simulations supported several as good candidates for probe substrates. I subsequently developed a two-step synthetic method to access a racemic mixture of one of these α-hydroxyketones. The De Voss and Bell laboratories in Australia have worked with this particular P450 system for many years, and hence we formed a collaboration with their groups to investigate the metabolism of my substrates with CYP199A4. Early experiments determined that a low abundance contaminant was a better substrate for the enzyme, complicating analyses. These early assays also suggested that two CYP199A4 mutants (Phe182Leu and Phe298Val) are better able to bind and turnover this α-hydroxyketone. After purifying the *para*-substituted α-hydroxyketone benzoic acid to 99+%, my collaborators in Australia are currently investigating whether this highly pure compound is turned over by

CYP199A4, generating the product predicted for the C-C lyase reaction, and are attempting to co-crystallize the protein with the substrate. With this substrate and a P450 able to mediate the C-C scission reaction now available, multiple avenues for exploring the structural and mechanistic determinants of this activity in CYP199A4 (and by implication CYP17A1) are presently available.

7.1. INTRODUCTION

Molecular dynamics simulations described in Chapter 6 attributed the failure of $P450_{cam}$ (CYP101A1) and $P450_{BM3}$ (CYP102A1) to catalyze carbon-carbon bond (C-C) scission of α-hydroxyketones to rare and transient positioning of the reactive moiety near the catalytically active oxygen-heme complex. In order to evaluate this explanation and its implications for the C17-C21 lyase reactivity of mammalian CYP17A1 (Fig. 6.1), another P450 needed to be selected as a model system. Criteria for choosing another paradigmatic enzyme included having available crystal structures, easy expression and purification, known and varied substrate scope, and rigid binding of these substrates with a single region held over the heme.

Although these requirements were stringent, a benzoic acid-metabolizing bacterial P450 (CYP199A4) met each of these criteria. This enzyme and its partner proteins express well in *E. coli* (1). In addition, CYP199A4 has an array of high-resolution crystal structures available both substrate free and with a variety of substrates bound (2–6). This P450 oxidizes a wide array of *para*-substituted benzoic acids as substrates (3, 7, 8), including a comparatively large 4-cyclohexyl-benzoate (9). Lastly and most importantly, interactions between the protein and the benzoate anion (Fig. 7.1) firmly orient the position *para*- to the carboxylic acid over the heme for oxidation (6, 10, 11).

After screening potential α-hydroxyketone substrates for favorable orientations in CYP199A4 using molecular dynamics (MD), two enantiomers from these were selected for synthesis. With the help of the De Voss and Bell collaborators in Australia, *in vitro* assays have evidenced P450-dependent C-C lyase reactivity with these model substrates. From these preliminary results, ongoing work includes co-crystallization of the α-hydroxyketones with CYP199A4 as well as a battery of spectroscopic and biochemical assays meant to investigate the mechanism of C-C scission.

7.2. RESULTS

7.2.1. Molecular dynamics (MD) simulations show favorable orientations for three compounds[1]

Because no crystal structures of the heme-oxy complex were available for CYP199A4, the first step towards MD was the creation of a CYP199A4 oxy-complex model. Aligning the hemes from the structures of the $P450_{cam}$ oxy-complex (12) and a high-resolution substrate-bound CYP199A4 (PDB: 5UVB) enabled a reasonable placement of the oxy-complex into CYP199A4. The 4-cyclopropyl benzoate substrate from the structure was removed before separately docking three different α-hydroxyketones—(*R*)-**1**, (*S*)-**1**, and **2** (Fig. 7.2)—into the active site using MOE. The pose for each molecule with the alcohol and ketone moieties best in position for C-C lyase by the peroxyanion-mediated mechanism (see Fig. 6.1) was selected as the starting point for MD.

After equilibrating the respective α-hydroxyketone-bound structures, MD over 5 μs (for both enantiomers of **1**) and 1 μs (for **2**) demonstrated a distribution of orientations (Fig. 7.2) meeting the geometric restraints established for C17-C21 lyase in CYP17A1 (13). (*S*)-**1** was

[1] I performed the initial docking and screening of the α-hydroxyketones into CYP199A4. Dr. Mark McLean (University of Illinois) performed the molecular dynamics simulations.

better positioned for catalysis than (*R*)-**1** for either mechanism (Table 7.1), although both of these compounds were positioned for the peroxyanion-mediated mechanism less than 1% of the time over 5 μs. Achiral α-hydroxyketone **2** presented a sizable population in position for the peroxyanion-mediated (3.0%) and Compound I-mediated (13.2%) mechanisms (Table 7.1).

7.2.2. Synthesis of α-hydroxyketones on a 4-benzoic acid scaffold

The achiral α-hydroxyketone **2** demonstrated a much larger population oriented for peroxyanion-mediated catalysis than either enantiomer of **1** (Fig. 7.2). Synthetic routes to **2** were longer, however, and the quickly accessible **1** (Fig. 7.3) showed catalytically competent orientations in the MD simulations. Thus, *rac*-**1** was my first target for synthesis.

A two-step route to *rac*-**1** began with metal-halogen exchange using an isopropyl magnesium chloride-lithium chloride reagent (14) and 4-bromobenzonitrile (**5**; Fig. 7.3) to produce an aryl Grignard *in situ*. This nucleophile, when reacted with neat 2,3-butadione, yielded α-hydroxyketone **6**. The subsequent attempt to hydrolyze the nitrile to carboxylic acid **1** with base successfully transformed the nitrile into a carboxylic acid, as desired, at the cost of eliminating the α-alcohol to form **7**.

The success of the nucleophilic addition to 2,3-butadione to create the α-hydroxyketone moiety of **6** led to its retention in another route to **1** (Fig. 7.3). Initial attempts with methyl 4-bromobenzoate did not yield α-hydroxyketone **9.** Swapping iodine for bromine (**8**) improved the magnesium-halogen exchange and led to the isolation of **9**. Ester hydrolysis with lithium hydroxide produced the target **1** as a racemic mixture at 95% purity. Later *in vitro* work with CYP199A4 described below revealed that the major contaminant was 4-acetylbenzoic acid (**3**), a better substrate for the enzyme, necessitating further purification. Repeating the synthesis and adding reverse-phase flash chromatography as the last step yielded *rac*-**1** at 99+% purity.

7.2.3. First pass in vitro assays with CYP199A4 suggest C-C scission reaction[2]

Although *rac*-**1** did not appear to yield the Type I spectral shifts with wild type CYP199A4 (Fig. 7.4) often observed for substrates of CYPs, NADH consumption increased ~100-fold upon addition of the compound (188 ± 5.1 μM_{NADH} μM_{P450}^{-1} min^{-1}). Analysis of the reaction products by HPLC revealed that the peak corresponding to *rac*-**1** was largely unchanged whereas a contaminant was readily consumed (Fig. 7.5). Co-elution of standards (Fig. 7.5) revealed the contaminant to be 4-acetylbenzoic acid (**3**; Fig. 7.2), the predicted product of the C-C lyase reaction. CYP199A4 appeared to preferentially oxidize **3** and the compounds derived from it compared to α-hydroxyketone **1**, rendering definitive assessment of whether the enzyme was catalyzing C-C scission difficult.

While I repeated the synthesis and purification of **1** to eliminate contaminating **3**, additional assays with wild type CYP199A4 suggested that **1** was not a strong substrate. Competitive binding experiments with the weak-binding 4-methoxybenzamide (**10**; Fig. 7.4) did not show evidence the α-hydroxyketone bound to the P450. When given large excesses of reducing equivalents as large amounts of NADH or a regenerating system, wild type CYP199A4 did not consume much of **1** (Fig. 7.5) further suggesting the enzyme does not readily turnover this compound.

Two mutants of CYP199A4 (Phe182Leu and Phe298Val; Fig. 7.1) known by my collaborators to accommodate larger substrates have demonstrated more promising interactions with **1** than did the wild type enzyme. In the presence of **1**, the Phe298Val mutant gave a small but perceptible red shift in the Soret band whereas the Phe182Leu mutant gave a Type I spectral

[2] Collaborators Joel H.Z. Lee (Ph.D. candidate; University of Adelaide) and Dr. Stephen G. Bell (University of Adelaide) conducted all *in vitro* experiments with CYP199A4 reported herein.

shift (Fig. 7.6). Both mutant enzymes consumed considerable amounts of **1** in the presence of a NADH regenerating system (Fig. 7.7) to yield predominantly 4-hydroxyacetylbenzoic acid, presumably by hydroxylation of **3** from the initial sample *and* from **3** generated by C-C scission. The identification of other, less polar, and more oxidized products from these reactions by GC-MS is currently underway.

7.3. DISCUSSION

Starting from MD simulations to evaluate possible α-hydroxyketone substrates for CYP199A4, successful synthesis of one of the promising compounds (**1**), and early tests with CYP199A4 provide tentative evidence for C-C lyase activity. The wild type enzyme does not readily turnover **1**; however, both the Phe182Leu and the Phe298Val mutants do turn over this compound with the former consuming large amounts of the α-hydroxyketone. Equipped with a substrate-enzyme pair to model the C17-C21 lyase activity of CYP17A1, numerous avenues of study are opening for investigating the mechanism of this important C-C breaking reaction, as outlined below.

Having a substrate-(mutant) CYP199A4 pair, co-crystallization of **1** with one or both of the Phe mutants is a foremost goal. Successful co-crystallization of **1** with CYP199A4 opens the possibility of co-crystallizing an α-hydroxyketone and the oxy-complex to attempt turnover *in crystallo*. From the substrate-bound oxy complex, reducing equivalents to the iron from the X-ray beam can supply the electrons required for catalysis. Ideally, this CYP and the reaction it facilitates can be captured in the act of catalysis (15) to differentiate between the peroxyanion- and Compound I-mediated mechanisms.

MD simulations suggested that (*S*)-**1** is likely better positioned for the C-C lyase reaction. Thus, the presence of (*R*)-**1** in the racemic mixture supplied to CYP199A4 to date may be

inhibiting the turnover of (*S*)-**1**. Chiral HPLC of reactions with rac-**1** by the De Voss group in Australia will unveil whether (*S*)-**1** is preferentially consumed by CYP199A4. Should the enzyme preferentially consume one of the enantiomers, preparative chiral chromatography to obtain the preferred enantiomer may be required for successful co-crystallization with CYP199A4.

Cryoradiolysis and resonance Raman spectroscopy of **1** and CYP199A4, as done with CYP17A1 (16–21), are also important to establish that a comparable O-O intermediate is observed with the bacterial system as in the human P450. Observing an inverse kinetic solvent isotope effect for the C-C breaking reaction (22) is another important check that the mechanism by which CYP199A4 mediates α-hydroxyketone C-C scission is with that by which CYP17A1 operates.

The benzoic acid scaffold of CYP199A4 substrates also provides unique opportunities to investigate C-C scission. Substitution on the aryl ring with electron-withdrawing and electron-donating groups can probe the electronics of the C-C breaking step of the reaction. For example, modified versions of **1** can evaluate the electron density in the transition state around the carbon that becomes the carbonyl during C-C scission. Likewise, modified versions of **2** can evaluate the electron density on the carbon that becomes the carboxylic acid during this reaction. Such studies, though insightful, require the synthesis of each benzoic acid derivative to be assayed. In addition, carefully selecting which kinetic parameter to compare (k_{cat}, k_{cat}/K_M, non-steady state parameters *etc.*) is required lest differences in binding affinity amongst the substrates unduly influence the results.

The identification of a paradigmatic, bacterial P450 (CYP199A4) that can catalyze C-C breaking of α-hydroxyketones is a key step in further understanding such reactions in CYP17A1.

Although 4-benzoate α-hydroxyketone substrates are not commercially available, the synthesis I developed can produce, in moderate yields, sufficient amounts of such compounds for an array of biochemical analyses. The identification of a substrate-enzyme pair capable of C-C scission has laid the groundwork for further work to determine definitively the structural and chemical mechanisms at play enabling C-C lyase in this CYP199A4 model system and, by implication, human CYP17A1.

7.4. EXPERIMENTAL

7.4.1. Substrate docking and molecular dynamics simulations[3]

The 4-cyclopropyl benzoate-bound structure of CYP199A4 (PDB: 5UVB) was selected for docking because it was among the highest in resolution (1.54 Å) and included a substrate with a branched hydrocarbon *para* to the carboxylate. After importing the structure into MOE, the CYP199A4 structure was aligned with the backbone of the oxygen-bound Thr252Ala CYP101A1 (PDB: 2A1O) (23). This alignment was manually adjusted to align the heme of CYP199A4 and the oxy-ferrous heme of the CYP101A1 structure. The CYP101A1 backbone and waters, CYP199A4 heme, and 4-cyclopropyl benzoate were all removed from the structure before equilibrating the oxy-ferrous CYP199A4 model to 0.01 kcal mol^{-1} A^{-2} using the Amber force field ff19SB (24). (*R*)-**1**, (*S*)-**1**, and **2** were independently docked into the active site of this equilibrated structure using MOE (25). After equilibration of the docked α-hydroxyketones conducted as stated above, the resulting molecular models were subjected to MD simulations.

The P450 MD simulations were set up using AmberTools 20 (26) using Amber force field ff19SB (24). Heme parameters were obtained from Shahrokh *et al*. (27). Substrate

[3] I performed the initial docking and screening of the α-hydroxyketones into CYP199A4. Dr. Mark McLean (University of Illinois) performed the molecular dynamics simulations.

parameters were generated from the respective mol2 files using the General Amber Force Field (28). Point mutations were generated using AmberTools (26) prior to minimization and equilibration. Systems were solvated using an octahedral periodic boundary with a 12 Å spacing. Potassium and chloride ions were added to neutralize the system, bringing the KCl concentration to 0.15 M.

MD simulations were performed using GROMACS 2020 (29). The solvated systems were first energy minimized using a steepest decent algorithm, followed by sequential equilibration runs, constraining the protein backbone before slowly releasing the constraints on the sidechains and substrates. Equilibration runs employed a modified Berendsen thermostat and a Berendsen barostat for temperature and pressure coupling. For production runs, temperature and pressure couplings were achieved using a modified Berendsen thermostat and a Parrinello-Rahman barostat, respectively. Electrostatics were calculated using Particle Mesh Ewald. For production runs the time step was 2 femtoseconds with coordinates saved every 10 ps for **1** and 100 ps for **2**. Runs continued until the accumulated simulation time reached ~5 μs (for both enantiomers of **1**) or ~3 μs (for **2**).

*7.4.2. Synthesis of α-hydroxyketone **1***

All reagents and solvents were purchased from commercial suppliers and used without further purification unless otherwise noted. Hexanes were vacuum distilled to remove high boiling impurities. When dry solvents were required, they were dried for two or more days over activated 3Å molecular sieves and stored under dry nitrogen gas. Thin layer chromatography utilized silica gel 60 doped with a fluorophore excited at 254 nm. Spots were visualized by fluorescent quenching and $KMnO_4$ staining. Flash column chromatography was conducted with 40-63 micron silica gel. ^{1}H NMR and ^{13}C NMR spectra were recorded on a 500 MHz

spectrometer at the University of Illinois School of Chemical Sciences NMR Spectroscopy Lab. NMR samples were made as solutions in $CDCl_3$ or CD_3OD using the residual solvent peak as the internal standard; δ values are given in ppm and coupling constants (*J*) in hertz (Hz). Mass spectra were obtained from a high-resolution ESI mass spectrometer at the University of Illinois School of Chemical Sciences Mass Spectrometry Lab.

7.4.2.1. Synthesis of nitrile α-hydroxyketone 6. In an oven dried round bottom flask under dry nitrogen gas, 502 mg (2.75 mmol, 1.1 eq.) 4-bromobenzonitrile (**5**) was dissolved in dry THF (to 1.2 M). This flask was cooled to 0°C in an ice bath before isopropyl magnesium chloride-lithium chloride (1.3 M in THF, 2.75 mmol, 1.1 eq.) was added dropwise and left to stir for 2 hours. Neat 2,3-butadione (215 mg, 2.50 mmol, 1.0 eq.) was added dropwise before removing the ice bath and warming the reaction mixture to room temperature. The reaction was stirred overnight under dry nitrogen gas, quenched with 5 mL sat. NH_4Cl, diluted with 20 mL water, extracted thrice with 15 mL DCM. The combined organic layers were washed once with 15 mL brine and dried over $MgSO_4$ before removing the solvents by rotary evaporation. Flash chromatography with 60% EtOAc in hexanes gave **6** (190. mg, 0.900 mmol, 36% yield; R_f, 0.38). ^{1}H NMR (499 MHz, $CDCl_3$) δ 7.71 – 7.64 (m, 2H), 7.62 – 7.56 (m, 2H), 4.49 (s, 1H), 2.10 (s, 3H), 1.78 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 23.63, 24.49, 76.90, 77.16, 77.41, 80.03, 112.19, 118.53, 127.02, 132.59, 146.85, 208.32. HR-ESI-MS *m/z* $[M+H]^+$, calcd for $C_{11}H_{10}NO_2$, 188.0712; found, 188.0714.

7.4.2.2. Hydrolysis of 6. A solution of KOH (134 mmol, 180 eq., 14 M) in EtOH (2.2 mL) and water (7.4 mL) was added to **6** (140. mg, 0.740 mmol). The resulting mixture was fitted with a reflux condenser, heated to 75°C, and stirred at that temperature until all the starting material disappeared (48 hours). The reaction mixture was cooled to room temperature, diluted

with 10 mL water, washed thrice with 15 mL DCM, acidified to a pH of ~2 with conc. HCl, and extracted thrice with 15 mL DCM. The combined organic layers from these later DCM extractions were washed with 15 mL brine, dried over $MgSO_4$, and the solvent removed by rotary evaporation to yield **7** (46 mg, 30% yield). ^{1}H NMR (499 MHz, $CDCl_3$) δ 8.14 – 8.10 (m, 1H), 8.10 – 8.07 (m, 2H), 7.50 – 7.46 (m, 2H), 4.99 (q, *J* = 6.5 Hz, 1H), 1.52 (d, *J* = 6.5 Hz, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 25.45, 70.19, 76.91, 77.16, 77.42, 125.56, 128.50, 130.67, 151.93, 171.70.

7.4.2.3. Synthesis of methyl ester α-hydroxyketone 9. In an oven dried round bottom flask under dry nitrogen gas, methyl 4-iodobenzoate (**8**; 399 mg, 1.53 mmol; 3.51 g, 13.4 mmol; 1.1 eq.) was dissolved in dry THF (to 1.2 M). This flask was cooled to -20°C in an ice bath before isopropyl magnesium chloride-lithium chloride (1.3 M in THF; 1.2 mL, 1.53 mmol; 10.3 mL, 13.4 mmol; 1.1 eq.) was added dropwise and left to stir for 2 hours. Neat 2,3-butadione (119 mg, 1.39 mmol; 1.05 g, 12.1 mmol; 1.0 eq.) was added dropwise before removing the ice bath and warming the reaction mixture to room temperature. The reaction was stirred for 1 hour under dry nitrogen gas, quenched with 0.1 M HCl (3 mL; 25 mL), diluted with water (15 mL; 100 mL), extracted with EtOAc (4x15 mL; 4x60 mL). The combined organic layers were washed once with brine (20 mL; 100 mL) and dried over $MgSO_4$ before removing the solvents by rotary evaporation. Flash chromatography with 40% EtOAc in hexanes gave **9** (R_f, 0.36; 148 mg, 44% yield; 565 mg, 29% yield). ^{1}H NMR (499 MHz, $CDCl_3$) δ 8.06 – 8.00 (m, 2H), 7.56 – 7.50 (m, 2H), 4.56 (s br, 1H), 3.91 (s, 3H), 2.08 (s, 3H), 1.79 (s, 3H). ^{13}C NMR (126 MHz, $CDCl_3$) δ 23.60, 24.34, 52.36, 76.90, 77.16, 77.41, 80.08, 126.25, 129.99, 130.07, 146.50, 166.79, 208.96. HR-ESI-MS *m/z* $[M+H]^+$, calcd for $C_{12}H_{15}O_4$, 223.0970; found, 223.0971.

7.4.2.4. Hydrolysis of methyl ester 9 to make *rac*-1. Methyl ester **9** (128 mg, 0.578 mmol) was dissolved in THF (to 0.11 M). A 0.5 M solution of LiOH (2.89 mmol, 5.0 eq) in water was added all at once and the solution stirred until all the starting material disappeared (2 hours). The solution was diluted with 15 mL water and washed thrice with 15 mL EtOAc. The aqueous layer was acidified to a pH of ~1.5 with 3 N HCl before extracting thrice with 15 mL EtOAc. The resulting orange organic layers were combined, dried over $MgSO_4$, and the solvent removed by rotary evaporation to yield 95% pure *rac*-**1** (91 mg, 0.34 mmol, 59% yield; λ_{max} 238 nm). A sample was subjected to reverse-phase flash column chromatography over C_{18}-functionalized silica to remove contaminating 4-acetyl benzoic acid (**3**; λ_{max} 248 nm). After activating the column with 4 CV 0.2% TFA in MeOH and equilibrating with 4 CV 0.2% TFA, 10% MeOH in water, a 10-60% MeOH in water with 0.2% TFA gradient over 12 CV separated these compounds as determined by the UV spectra of the fractions. Fractions containing only *rac*-**1** were combined and lyophilized to yield 15 mg (0.072 mmol; 21% yield) of this α-hydroxyketone as a white solid of 99+% purity. ^{1}H NMR (500 MHz, CD_3OD) δ 8.04 – 7.98 (m, 2H), 7.63 – 7.57 (m, 2H), 2.09 (s, 3H), 1.65 (s, 3H). ^{13}C NMR (126 MHz, MeOD) δ 24.44, 26.16, 48.49, 48.66, 48.83, 49.00, 49.17, 49.34, 49.51, 81.61, 126.54, 130.82, 131.12, 149.44, 169.50, 211.58. HR-ESI-MS *m/z* $[M-H]^-$, calcd for $C_{11}H_{11}O_4$, 207.0657; found, 223.0661.

7.4.3. Protein expression and purification[4]

CYP199A4 and its redox partners (HaPux and HaPuR) were expressed in BL21(DE3) *E. coli* and purified according to previously reported methods (2).

[4] All experiments with CYP199A4 (Sections 7.4.3-7.4.6) were performed by Joel H.Z. Lee, Ph.D. candidate (University of Adelaide).

7.4.4. Spectral binding assays

To measure the spin-state shift, 0.1-5 μL aliquots of a 100 mM substrate stock solution in EtOH or DMSO were successively added to 500 μL of ~1-2 μM P450 in Tris-HCl buffer (50 mM, pH 7.4). The UV-Vis spectrum was recorded after each addition, and additional substrate was added until there was no further shift.

7.4.5. NADH consumption assays

In vitro NADH turnovers were performed at 30˚C and contained CYP199A4 (1 μM), HaPuX (5 μM), HaPuR (0.5 μM) and 100 ng μL^{-1} bovine liver catalase in oxygenated Tris-HCl buffer (50 mM, pH 7.4) in a total volume of 1.2 mL. The absorbance at 340 nm was set to zero and the mixture was incubated at 30 ˚C for 2 min before NADH was added to a concentration of ~320 μM (an absorbance of ~2.0). The rate of NADH background oxidation was measured before initiating the reaction by addition of substrate. To start the reaction, 500 μM substrate was added from a 100 mM stock in EtOH/DMSO, and NADH depletion was monitored at 340 nm. The rate of NADH consumption by the P450 enzyme in units of $\mu M_{NADH}\ \mu M_{P450}^{-1}\ min^{-1}$ was calculated from the slope of the graph of $Abs_{340\ nm}$ versus time using an extinction coefficient of $\varepsilon_{340\ nm} = 6.22\ mM^{-1}\ cm^{-1}$ and reported as min^{-1}.

All experiments were performed three times with the mean and standard deviation reported. Control reactions were also performed in which either the P450 or NADH was omitted from the turnover mixture (replaced with the same volume of buffer).

*7.4.6. Turnover of **1** with CYP199A4*

Reactions were carried out in a total volume of 600 μL in Tris-HCl buffer (50 mM, pH 7.4) in 1.5 mL Eppendorf tubes. The reactions contained the following components: P450 (1 μM), HaPuR (0.5 μM), HaPuX (5 μM), Alcohol Dehydrogenase (ADH, 4.5 μL of 0.029 g mL^{-1} suspension, Roche), bovine liver catalase (100 ng μL^{-1}, Sigma-Aldrich), EtOH (12 μL) and substrate (150 or 50 μM). NADH (320 μM) was added last to start the reaction. Reactions were incubated at 25°C and 250 rpm. Alternative conditions included excess NADH, 1 mM NADH; for mutant F182L and F298V CYP199A4 turnovers, only 50 μM of substrate was used. Negative controls included no P450/redox partners and no ADH. The remaining volume was made up with Tris-HCl buffer.

7.5. TABLE AND FIGURES

Table 7.1. Frequency with which α-hydroxyketones are positioned as necessary for catalysis by the given mechanism in molecular dynamics simulations of CYP199A4.

	Peroxyanion[a]		Compound I[b]	
Compound	*N*	Percentage	*N*	Percentage
(*R*)-**1**[c]	555	0.11%	46,343	9.10%
(*S*)-**1**[c]	1,591	0.31%	270,430	53.0%
2[d]	946	2.96%	4219	13.2%

[a] Proximal oxygen within 3.62 Å of alcohol hydrogen ***and*** distal oxygen within 3.47 Å of carbonyl carbon.
[b] Proximal oxygen within 3.62 Å of alcohol hydrogen.
[c] 510,002 frames sampled at 10 ps per frame; ~5 μs total
[d] 32,001 frames samples at 100 ps per frame; ~3 μs total

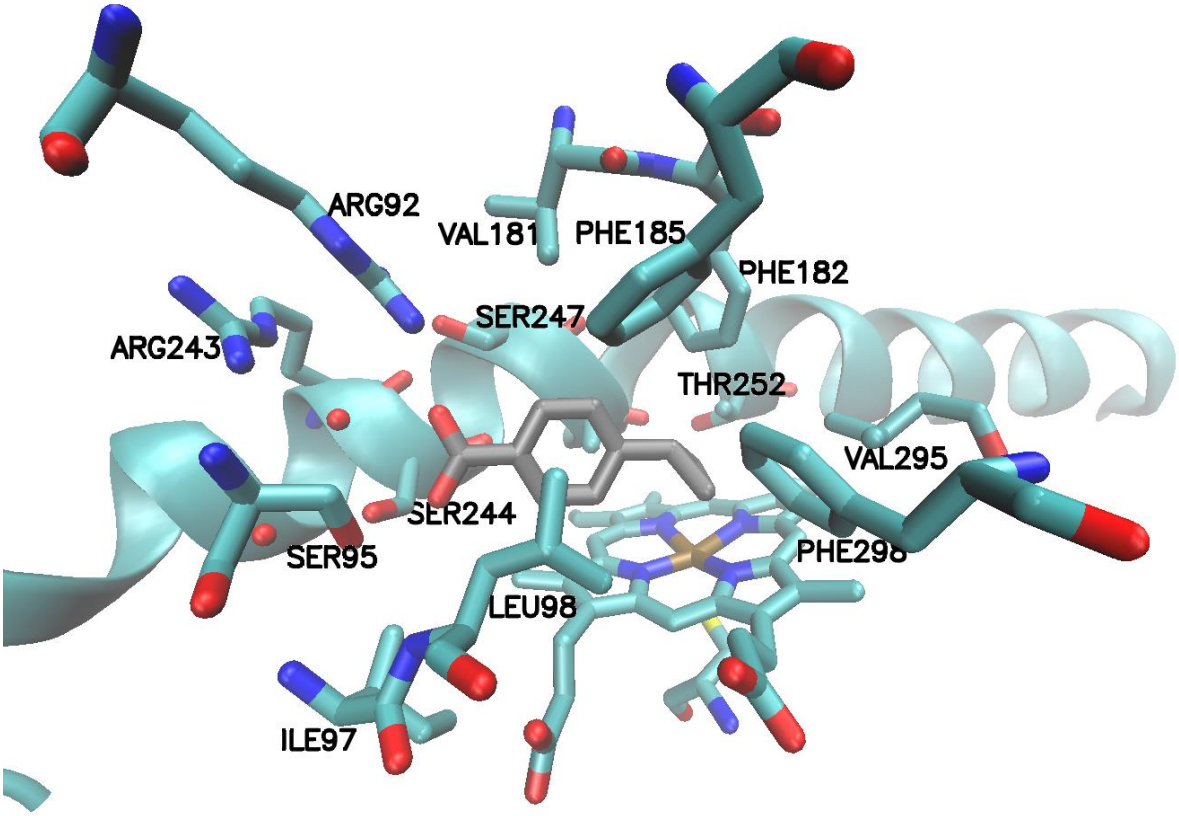


Figure 7.1. CYP199A4 key substrate binding interactions.
4-Cylopropyl benzoate-bound CYP199A4 (PDB: 5UVB) with all residues within 4.5 Å of the benzoic acid and the heme (cyan tubes) and the I-helix (cyan ribbon) shown.

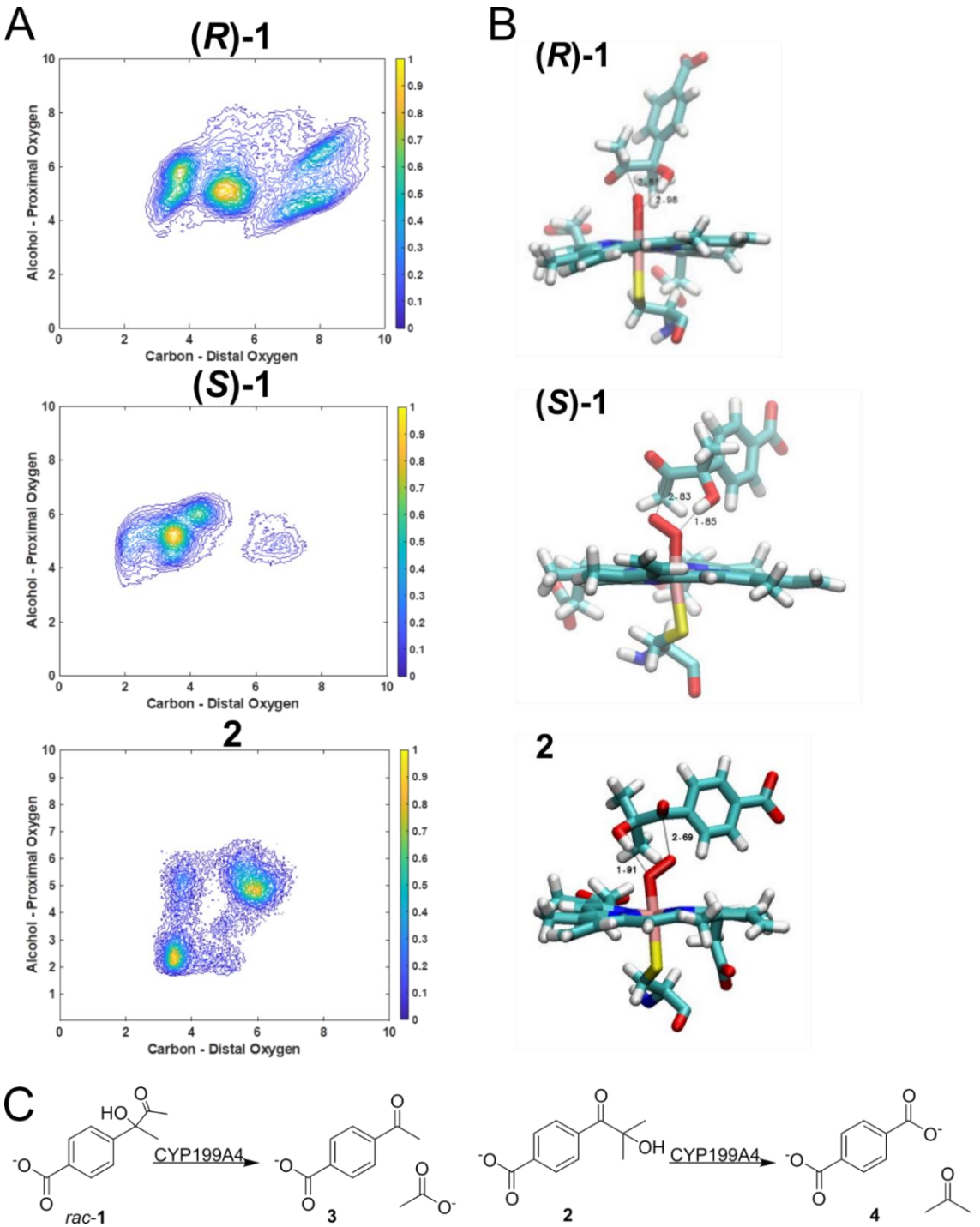


Figure 7.2. Molecular dynamics simulations predict catalytically favorable conformations for C-C lyase reaction with several compounds.

(A) Histograms showing the distribution of α-hydroxyketone orientations throughout MD simulations. (B) Snapshot of the simulation where the α-hydroxyketone is in position for catalysis by the peroxoanion-mediated mechanism of C-C scission. (C) Scheme displaying the α-hydroxyketones docked (**1** and **2**) and the predicted products (**3** and **4**) of a C-C lyase reaction catalyzed by CYP199A4.

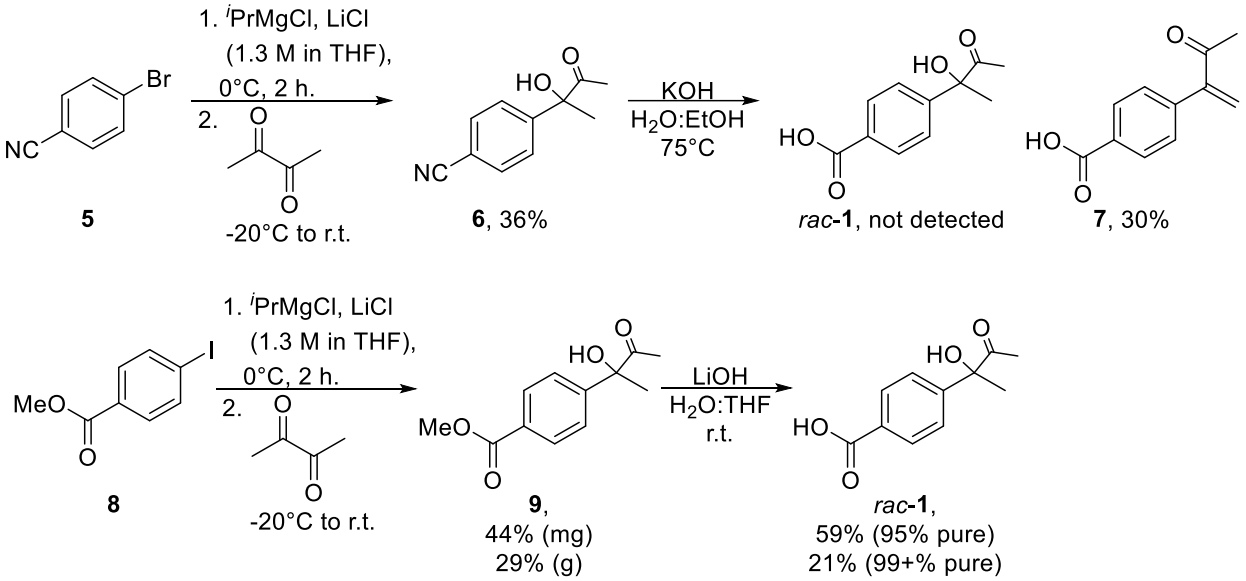


Figure 7.3. Synthesis of α-hydroxyketones onto the *para*-benzoic acid scaffold.

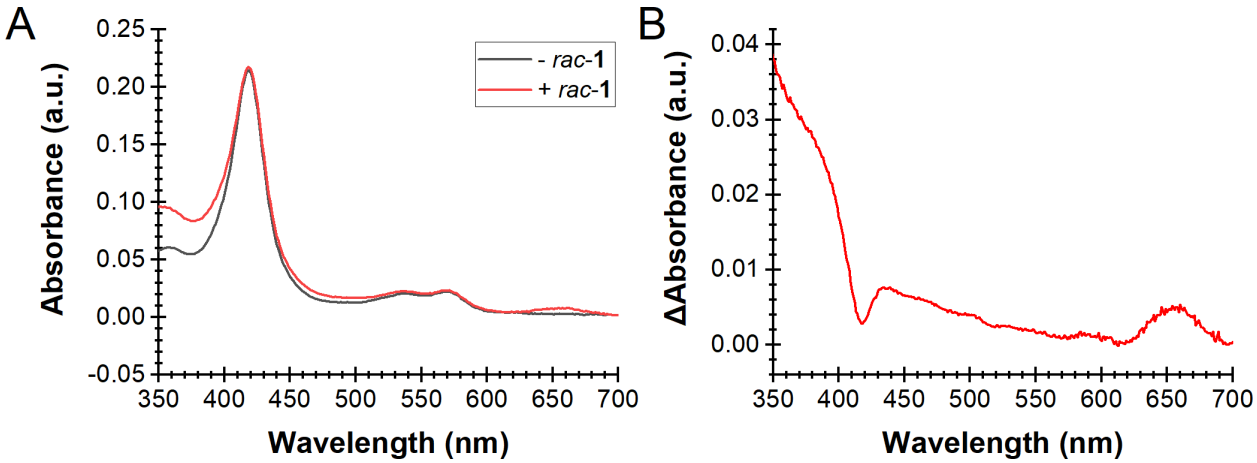


Figure 7.4. Lack of spectral shifts suggest wild type CYP199A4 does not bind *rac*-1 well. ~1 µM CYP199A4 was incubated without (minus) or with (plus) *rac*-**1**. Shown are the resulting absolute (A) and difference (B) spectra. Data collected by Joel H.Z. Lee, Ph.D. candidate (University of Adelaide).

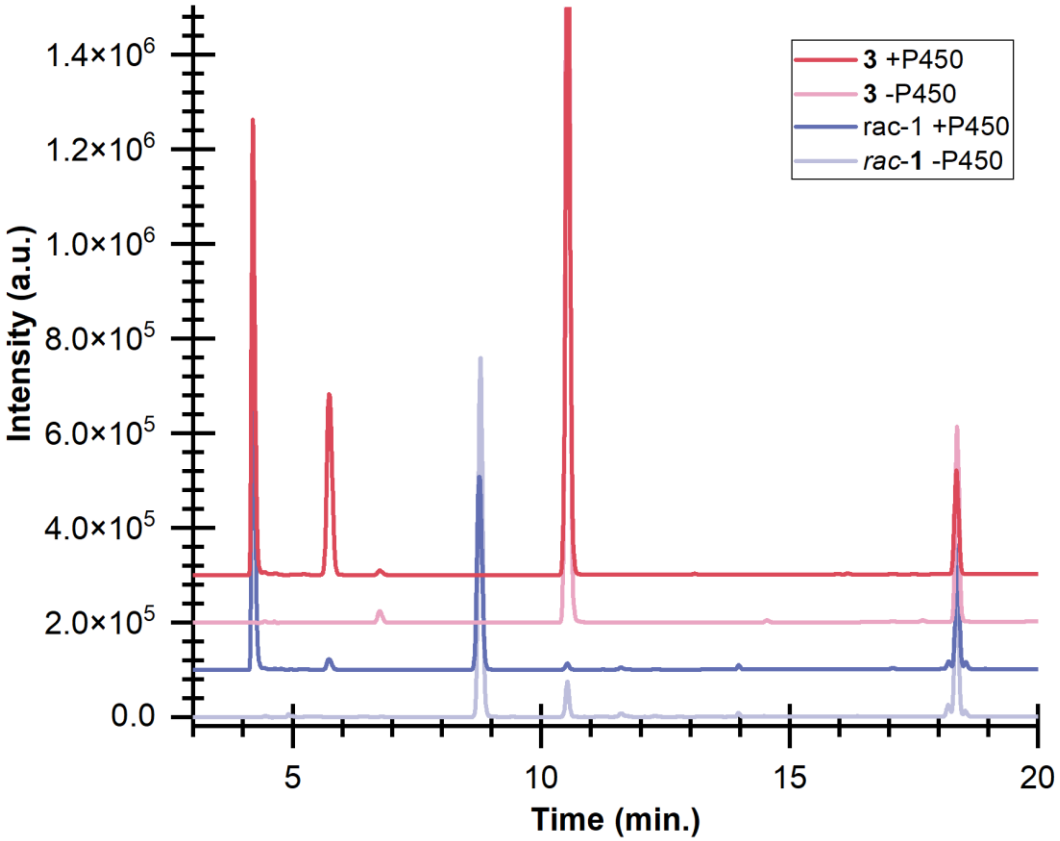


Figure 7.5. Wild type CYP199A4 reactions with *rac*-1. Reactions contained 1 µM CYP199A4, 0.5 µM HaPuR, 5 µM HaPux, 320 µM NADH, and 50 µM substrate. Data collected by Joel H.Z. Lee, Ph.D. candidate (University of Adelaide).

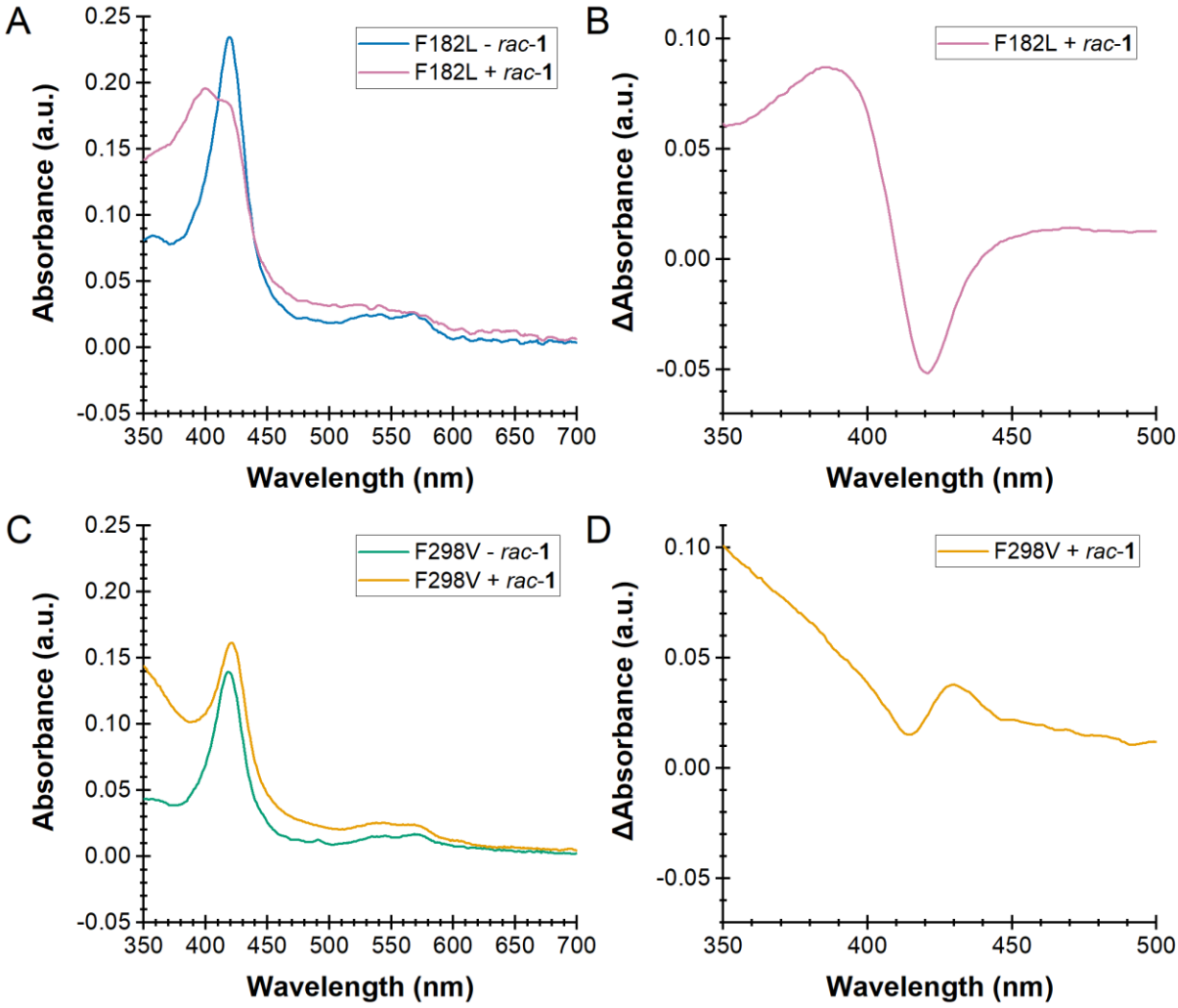


Figure 7.6. Spectral shifts observed in presence of *rac*-1 for two CYP199A4 mutants. ~1 µM CYP199A4 was incubated without (minus) or with (plus) *rac*-**1**. Shown are the resulting absolute (A, C) and difference (B, D) spectra with the F182L (A, B) and F298V (C, D) CYP199A4. Data collected by Joel H.Z. Lee, Ph.D. candidate (University of Adelaide).

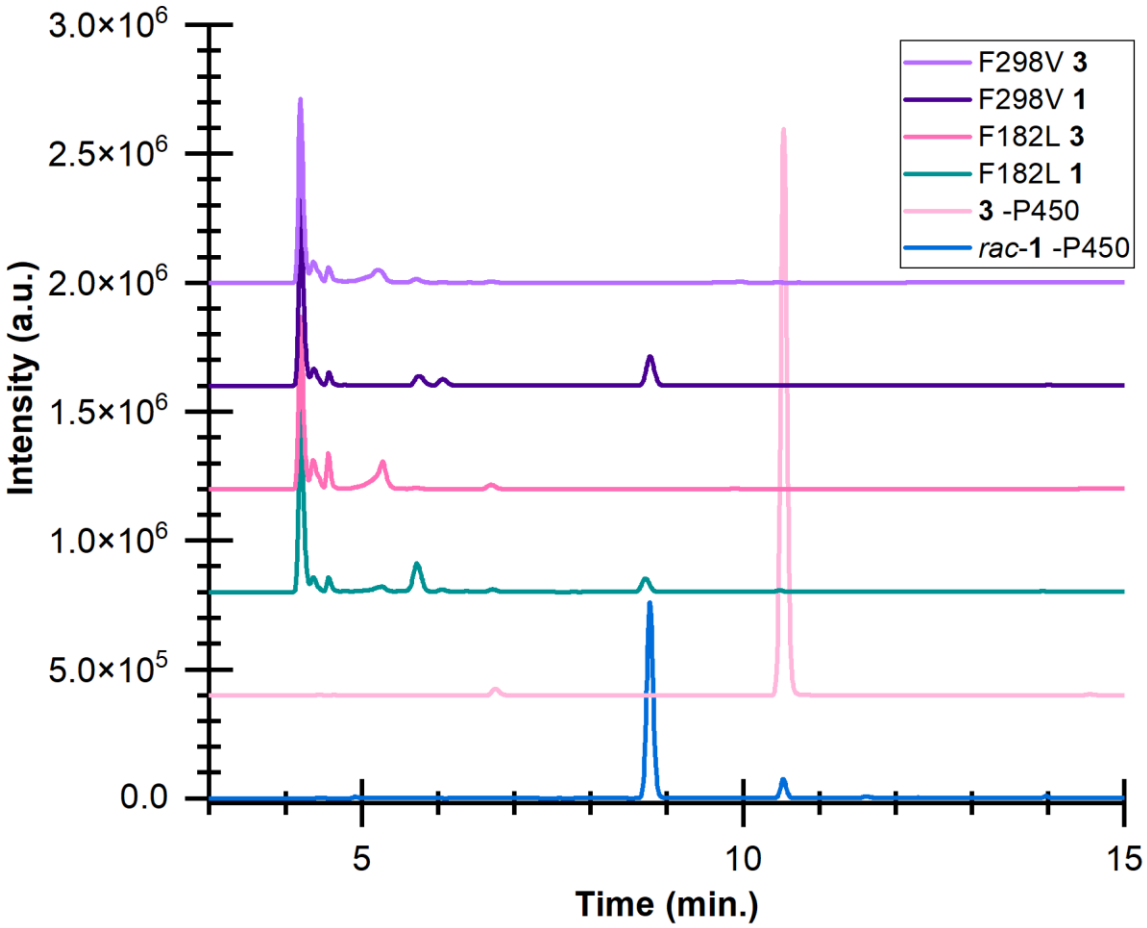


Figure 7.7. Phe182Leu, Phe298Val CYP199A4 reactions with *rac*-1.
Reactions contained 1 µM CYP199A4, 0.5 µM HaPuR, 5 µM HaPux, 320 µM NADH, and 50 µM substrate in the presence of alcohol dehydrogenase, catalase, and EtOH. Data collected by Joel H.Z. Lee, Ph.D. candidate (University of Adelaide).

7.6. REFERENCES

1. Bell, S. G., Tan, A. B. H., Johnson, E. O. D., and Wong, L.-L. (2009) Selective oxidative demethylation of veratric acid to vanillic acid by CYP199A4 from *Rhodopseudomonas palustris* HaA2. *Mol. Biosyst.* **6**, 206–214
2. Bell, S. G., Yang, W., Tan, A. B. H., Zhou, R., Johnson, E. O. D., Zhang, A., Zhou, W., Rao, Z., and Wong, L.-L. (2012) The crystal structures of 4-methoxybenzoate bound CYP199A2 and CYP199A4: Structural changes on substrate binding and the identification of an anion binding site. *Dalton Trans.* **41**, 8703–8714
3. Podgorski, M. N., Coleman, T., Chao, R. R., De Voss, J. J., Bruning, J. B., and Bell, S. G. (2020) Investigation of the requirements for efficient and selective cytochrome P450 monooxygenase catalysis across different reactions. *J. Inorg. Biochem.* **203**, 110913
4. Coleman, T., Stok, J. E., Podgorski, M. N., Bruning, J. B., De Voss, J. J., and Bell, S. G. (2020) Structural insights into the role of the acid-alcohol pair of residues required for dioxygen activation in cytochrome P450 enzymes. *J. Biol. Inorg. Chem.* **25**, 583–596
5. Coleman, T., Kirk, A. M., Chao, R. R., Podgorski, M. N., Harbort, J. S., Churchman, L. R., Bruning, J. B., Bernhardt, P. V., Harmer, J. R., Krenske, E. H., De Voss, J. J., and Bell, S. G. (2021) Understanding the mechanistic requirements for efficient and stereoselective alkene epoxidation by a cytochrome P450 enzyme. *ACS Catal.* **11**, 1995–2010
6. Chao, R. R., Lau, I. C.-K., Coleman, T., Churchman, L. R., Child, S. A., Lee, J. H. Z., Bruning, J. B., De Voss, J. J., and Bell, S. G. (2021) The stereoselective oxidation of para-substituted benzenes by a cytochrome P450 biocatalyst. *Chem. – Eur. J.* **27**, 14765–14777
7. Bell, S. G., Zhou, R., Yang, W., Tan, A. B. H., Gentleman, A. S., Wong, L.-L., and Zhou, W. (2012) Investigation of the substrate range of CYP199A4: Modification of the partition between hydroxylation and desaturation activities by substrate and protein engineering. *Chem. – Eur. J.* **18**, 16677–16688
8. Coleman, T., Wong, S. H., Podgorski, M. N., Bruning, J. B., De Voss, J. J., and Bell, S. G. (2018) Cytochrome P450 CYP199A4 from *Rhodopseudomonas palustris* catalyzes heteroatom dealkylations, sulfoxidation, and amide and cyclic hemiacetal formation. *ACS Catal.* **8**, 5915–5927
9. Coleman, T., Kirk, A. M., Lee, J. H. Z., Doherty, D. Z., Bruning, J. B., Krenske, E. H., De Voss, J. J., and Bell, S. G. (2022) Different geometric requirements for cytochrome P450-catalyzed aliphatic versus aromatic hydroxylation results in chemoselective oxidation. *ACS Catal.* **12**, 1258–1267
10. Coleman, T., Chao, R. R., Bruning, J. B., Voss, J. J. D., and Bell, S. G. (2015) CYP199A4 catalyses the efficient demethylation and demethenylation of para-substituted benzoic acid derivatives. *RSC Adv.* **5**, 52007–52018
11. Coleman, T., Chao, R. R., De Voss, J. J., and Bell, S. G. (2016) The importance of the benzoic acid carboxylate moiety for substrate recognition by CYP199A4 from *Rhodopseudomonas palustris* HaA2. *Biochim. Biophys. Acta BBA - Proteins Proteomics.* **1864**, 667–675
12. Schlichting, I., Berendzen, J., Chu, K., Stock, A. M., Maves, S. A., Benson, D. E., Sweet, R. M., Ringe, D., Petsko, G. A., and Sligar, S. G. (2000) The catalytic pathway of cytochrome P450cam at atomic resolution. *Science.* **287**, 1615–1622

13. Bonomo, S., Jørgensen, F. S., and Olsen, L. (2017) Mechanism of cytochrome P450 17A1-catalyzed hydroxylase and lyase reactions. *J. Chem. Inf. Model.* **57**, 1123–1133
14. Kopp, F., Wunderlich, S., and Knochel, P. (2007) Halogen–magnesium exchange on unprotected aromatic and heteroaromatic carboxylic acids. *Chem. Commun.* 10.1039/B618923G
15. Weber, I. T., Agniswamy, J., Fu, G., Shen, C.-H., and Harrison, R. W. (2012) Reaction intermediates discovered in crystal structures of enzymes. in *Advances in Protein Chemistry and Structural Biology* (Christov, C., and Karabencheva-Christova, T. eds), pp. 57–86, Structural and Mechanistic Enzymology, Academic Press, **87**, 57–86
16. Gregory, M., Mak, P. J., Sligar, S. G., and Kincaid, J. R. (2013) Differential hydrogen bonding in human CYP17 dictates hydroxylation versus lyase chemistry. *Angew. Chem. Int. Ed.* **52**, 5342–5345
17. Mak, P. J., Gregory, M. C., Denisov, I. G., Sligar, S. G., and Kincaid, J. R. (2015) Unveiling the crucial intermediates in androgen production. *Proc. Natl. Acad. Sci.* **112**, 15856–15861
18. Duggal, R., Liu, Y., Gregory, M. C., Denisov, I. G., Kincaid, J. R., and Sligar, S. G. (2016) Evidence that cytochrome b_5 acts as a redox donor in CYP17A1 mediated androgen synthesis. *Biochem. Biophys. Res. Commun.* **477**, 202–208
19. Mak, P. J., Duggal, R., Denisov, I. G., Gregory, M. C., Sligar, S. G., and Kincaid, J. R. (2018) Human cytochrome CYP17A1: The structural basis for compromised lyase activity with 17-hydroxyprogesterone. *J. Am. Chem. Soc.* **140**, 7324–7331
20. Liu, Y., Denisov, I. G., Sligar, S. G., and Kincaid, J. R. (2021) Substrate-specific allosteric effects on the enhancement of CYP17A1 lyase efficiency by cytochrome b_5. *J. Am. Chem. Soc.* **143**, 3729–3733
21. Liu, Y., Grinkova, Y., Gregory, M. C., Denisov, I. G., Kincaid, J. R., and Sligar, S. G. (2021) Mechanism of the clinically relevant E305G mutation in human P450 CYP17A1. *Biochemistry*. **60**, 3262–3271
22. Gregory, M. C., Denisov, I. G., Grinkova, Y. V., Khatri, Y., and Sligar, S. G. (2013) Kinetic solvent isotope effect in human P450 CYP17A1-mediated androgen formation: Evidence for a reactive peroxoanion intermediate. *J. Am. Chem. Soc.* **135**, 16245–16247
23. Nagano, S., and Poulos, T. L. (2005) Crystallographic study on the dioxygen complex of wild-type and mutant cytochrome P450cam. Implications for the dioxygen activation mechanism. *J. Biol. Chem.* **280**, 31659–31663
24. Tian, C., Kasavajhala, K., Belfon, K. A. A., Raguette, L., Huang, H., Migues, A. N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., and Simmerling, C. (2020) ff19SB: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. *J. Chem. Theory Comput.* **16**, 528–552
25. Rupasinghe, S., and Schuler, M. A. (2006) Homology modeling of plant cytochrome P450s. *Phytochem. Rev.* **5**, 473–505
26. Case, D. A., Belfon, K. A. A., Ben-Shalom, I., Brozell, S. R., Cerutti, D., Cheatham III, T. E., Cruzeiro, V. W. D., Darden, T., Duke, R. E., Giambasu, G., Gilson, M. K., Gohlke, H., Götz, A. W., Harris, R., Izadi, S., Izmailov, S. A., Kasavajhala, K., Kovalenko, A., Krasny, R., Kurtzman, T., Lee, T. S., LeGrand, S., Li, P., Lin, C., Liu, J., Luchko, T., Luo, R., Man, V., Merz, K. M., Miao, Y., Mikhailovskii, O., Monard, G., Nguyen, H., Onufriev, A., Pan, F., Pantano, S., Qi, R., Roe, D. R., Roitberg, A., Sagui, C., Schott-Verdugo, S., Shen, J.,

Simmerling, C., Skrynnikov, N. R., Smith, J., Swails, J., Walker, R. C., Wang, J., Wang, J., Wilson, L., Wolf, R. M., Wu, X., Xiong, Y., Xue, Y., York, D. M., and Kollman, P. A. (2020) *AMBER 2020*, University of California, San Francisco

27. Shahrokh, K., Orendt, A., Yost, G. S., and Cheatham III, T. E. (2012) Quantum mechanically derived AMBER-compatible heme parameters for various states of the cytochrome P450 catalytic cycle. *J. Comput. Chem.* **33**, 119–133
28. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., and Case, D. A. (2004) Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174
29. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. **1–2**, 19–25

CHAPTER 8: BREAKING CARBON-CARBON BONDS WITH CYTOCHROME P450s FROM PLANTS, ANIMALS, AND BACTERIA— A SUMMARY FOR THE GENERAL PUBLIC

8.1. THOSE ARE WORDS, BUT WHAT DO THEY MEAN?

First, I apologize for placing this chapter (the one I hope to be most intelligible to all audiences) at the end of the dissertation. If you have read thus far, I earnestly commend you for your devotion; however, you are unlikely to be my target audience for this chapter—the general public with little background in experimental natural science. Should that description fit you and you have stumbled upon this part of the tome that I have written, stay here. This chapter (more than any of the others) is the one that you will want to read.

My aim in these few pages is to summarize the motivations behind, the results from, and the societal import of my research endeavors over the past five years. Sadly, there will be no tales of woe or intrigue, no escapes from near-death experiences, no narrative storytelling of my day-to-day activities as a diary or journal might yield. Instead, the seven other chapters in this dissertation are meant to be a sort of masterpiece to allow me entry into the guild of research scientists (more specifically Ph.D. chemists). In this chapter I seek to distill the major findings of the work constituting my scientific apprenticeship so that others need take only a few hours reading to learn what has taken me several years' time.

As with all trades and crafts, there is a particular jargon used within the guild to discuss the craft. Biochemists like me freely talk of enzymes, bonds, proteins, cytochrome P450s, *etc.* as an automotive mechanic might speak of pistons, gaskets, seals, gears, motors, and wires. Here

are some terms and processes with which you will want to familiarize yourself to better understand my work:

- **Proteins**: These large molecules are one of several important classes of compounds found in biological systems. Proteins come in all kinds of shapes and sizes owing to their modular nature. Like constructions made from Lego blocks, proteins take all sorts of shapes and sizes from 20 (and occasionally a few more) building blocks called amino acids. The order in which these 20 amino acids are compiled determine the shape and thus the function of the protein. This order is encoded in a genetic code (DNA) shared by all life on earth with another molecule (RNA) serving as a messenger telling a cell to make a particular protein.
- **Enzymes**: These are the microscopic (actually, nanoscopic) machines of cells. Almost always made of proteins, these machines work by making and breaking chemical bonds to make life happen on the molecular level. When strung together, enzymes become factory-like in their ability to assemble and disassemble molecules in processes collectively called metabolism. As proteins, the sequence of an enzyme determines its structure, and (like any machine) its structure determines its function. Occasionally, enzymes require abilities not achievable with just 20 amino acids. Other, smaller molecules called cofactors are then incorporated into the protein to create the active enzyme.
- **Heme**: This is a particular cofactor that includes an iron atom held in place by a ring in another molecule called a porphyrin. Because of iron's abundance throughout the earth and its ability to interact with molecular oxygen, heme is commonly found in proteins that utilize oxygen in some way. Examples of such proteins include the oxygen-carrier proteins hemoglobin (in the blood) and myoglobin (in the muscles) of animals as well as the cytochrome P450s that I have studied.
- ***Escherichia coli* (*E. coli*)**: This species of bacteria is largely found in the lower-intestinal tracts of warm-blooded animals. Because they are fast growing, well studied, and *very* easy to work with and manipulate, *E. coli* are used as microbial factories to do biochemical work. Most importantly for my work, a shared genetic architecture across life on earth (DNA encodes RNA that encodes protein) enables the production of proteins found in one organism's genome by *E. coli* for closer study. This concept is called heterologous expression.
- **Spectroscopy**: This is the study of how light and energy interact with matter. Different kinds of energy (*e.g.*, wavelengths of light) interact variably with different kinds of chemicals. By observing which interactions are occurring and the strength of these interactions, scientists are able to infer the chemical composition of a sample. When wavelengths of light corresponding to ultraviolet (UV) or those of lesser energy (visible, infrared, microwave, *etc.*) are used, the sample is unlikely to be degraded and a lot of data can be collected in a short amount of time. From these data, the presence of certain chemicals and the amount present can be deduced.
- **Synthesis**: This is how chemists speak of making new chemicals. When using organic chemicals in round-bottomed flasks, we speak of organic synthesis. When performed by a

biological system, we speak of biosynthesis and strings of reactions constituting a biosynthetic pathway.

- **Catalyst**: A chemical (large or small) that increases the rate of a reaction. Enzymes are amongst the most prolific catalysts in the world with some enhancing the rate of reaction over one-million-fold. When discussing how catalysts function without being consumed within the reaction, chemists will describe a *catalytic cycle* describing the different steps these machines take to facilitate the reaction of interest.

Having supplied these explanations, I now turn my attention to dissecting the jargon found in the title of my dissertation and this chapter. From here, I discuss the motivations (religious as well as scientific and societal) for undertaking the research endeavors summarized herein. Condensing five years' worth of results in the span of a few pages is the next portion of the chapter. I conclude with a discussion of the questions my work leaves unanswered, new hypotheses and questions that emerged from it, and its small but discernable contributions to society.

8.1.1. What is so special about breaking carbon-carbon bonds?

Living organisms are physically made of molecules. Organisms must either acquire such compounds from the environment (by eating, drinking, and breathing in humans) or produce such compounds from these acquired resources. In plants, for example, all the molecules that enable a seed to grow into a tall tree are synthesized from a few gases (carbon dioxide, CO_2; oxygen, O_2), salts (ammonium, NH_4^+; nitrate, NO_3^-; phosphate, PO_4^{3-}; sulfate, SO_4^{2-}), water (H_2O), and energy sourced from sunlight. Along the way, these "inorganic" chemicals are converted into "organic" and "bio" molecules that meet the physical requirements to sustain life.

The enzymes that make the millions of molecules of life of numerous, diverse, and (when strung together in biosynthetic pathways) able to make these molecules in all sorts of shapes and sizes (Fig. 8.1). Because carbon fixed from CO_2 serves as the backbone or scaffold of these biomolecules, chemists and biologists organize these structures by the structure of their carbon backbones. This taxonomic approach also helpfully maps onto the biological functions of these

molecules. The repertoire of enzymes that make the carbon-carbon (C-C) bonds building these scaffolds and the shapes that they can form, however, are finite. The limitations of raw materials and ways to connect them together prevent the direct synthesis of certain scaffolds and biomolecules needed by living organisms

Such a problem is not unique to the atomic world. To construct an arch of great size, one must first construct a scaffold to hold the structure in place until the arch is built and self-sustaining. Then and only then is the scaffolding removed. Likewise, cutting out many letters (*e.g.*, A, B, and D) from paper using stencils first requires drawing the shapes with portions still connected to the outside of the letter. Often, the letter's overall shape is then excised with this attachment still in place. Removal of this portion of the cutout occurs only when one is prepared to paste the letters in place and form their final shape. Biological systems take a similar approach: The standard repertoire of enzymes constructs a core scaffold; once the proper time in the biosynthetic sequence arrives, another enzyme facilitates the removal of the appropriate C-C bond to achieve the required three-dimensional shape. Breaking C-C bonds is difficult, however, because the bonds are generally strong and resistant to transformation due to their energetic stability. These observations together raise an important question—which kinds of enzymes are capable of cutting C-C bonds like molecular scissors?

8.1.2. What is a cytochrome P450? What do they do?

Proteins, like the biomolecules discussed above, are organized by their three-dimensional shape. This taxonomy conveniently corroborates much of the function as well because a protein's structure determines its function. To date there are no fewer than 24,000 protein families organized on these bases within the online database InterPro (1). Cytochrome P450s (CYPs, P450s) are just one of these families (InterPro family identifier IPR001128), but there are

more than 446,000 CYPs found in under 30,000 species (2). Although more common in eukaryotic organisms like plants, animals, and fungi, P450s occur in each domain of the microbial world as well (Tables 8.1 and 8.2; Fig. 8.2). CYPs are among a small selection of enzymes capable of breaking C-C bonds (3). The numbers given here describe an important point: Among tens of thousands of other kinds of nanoscopic proteins, CYPs fulfill a specific yet vital role of difficult bond transformations for tens of thousands of species.

This ability of some P450s to break C-C bonds is related to the broader ability of these enzymes to react with another kind of inert bond, C-H bonds. Using O_2, some electrons (e^-) supplied from a partner protein, some protons (H^+) from solution, and a heme cofactor containing iron (Fe), CYPs utilize a catalytic cycle (Fig. 8.3) able to oxidize (*i.e.*, a specific kind of chemical transformation) even these very inert bond types. Although not every C-C bond-breaking reaction appears to occur in the same way (in the language of chemists, by the same mechanism), this series of reactive intermediates derived from O_2 are what enable CYPs to perform these difficult transformations and what interest chemists like me.

8.2. WHY UNDERTAKE THESE ENDEAVORS?

Fascination with P450s and their ability to mediate these difficult C-C bond-breaking reactions is not in itself sufficient justification to investigate these matters. Before outlining the anticipated benefits to society from my work, I will describe the religious importance of my labors. This ordering signifies the relative importance of the sacred and secular motivations for undertaking these research endeavors; furthermore, my religious motivations are the very foundation that has enabled me to conduct the work summarized herein.

8.2.1. Religious motivations[1]

Natural science depends upon the assumption that there are fundamental physical or natural laws that govern how the natural world operates. If this were not true, observations made in one place in time could differ from observations made elsewhere under the same circumstances. Indeed, because God the great Creator of all things "still upholds heaven and earth and all creatures, and so governs them that […] all things, come not by chance but by His fatherly hand" (HC 27 see also WCF 5.1, 5.2; WSC 11; WLC 18; BC 12, 13), the physical circumstances that produce a phenomenon in one place and time must upon their coalescence cause the same event, apart from an extraordinary act of God. These bare facts allow for me to confidently eat food, knowing it will supply nourishment to my body; to take an antibiotic, knowing that (unless bacteria are resistant) it will kill them; and to report experimental results, knowing that the same enzyme will do the same reaction on the same substrate in the same way under the same circumstances. Thus, I can depend on thousands of years of human knowledge about the natural world passed on to me throughout my education to order the world in which I live. Successive generations can likewise take the findings of my work, insofar as they accurately describe the natural order, and learn still more about the natural world.

My Christian faith supplies more than assurance that natural science is possible. Just as nature is not governed by change but by providence, so too humanity is not aimless in the world.

[1] For clarity and ease of reading, the parenthetical references given in this section do not correspond to items cited in this chapter's reference list. Instead, they correspond to Biblical references from the English Standard Version (Crossway: Wheaton, Illinois, 2001) and summaries of Biblical teachings found in two sets of confessional documents: the Westminster Confession of Faith (WCF), Shorter Catechism (WSC), and Larger Catechism (WLC); and the Three Forms of Unity—the Heidelberg Catechism (HC), the Belgic Confession (BC), and the Canons of Dort (CD). For those unfamiliar with these confessional documents, their purpose is to summarize the teachings of the Bible for declaring what is believed by the church and instruction of her members. My use of them is analogous to citing a review article for reference instead of the primary literature. The WCF, WSC, HC, BC, and CD are easily accessible from Westminster Theological Seminary; each of the Westminster documents (WCF, WSC, and WLC) are available at The Westminster Standard.

God created humanity to glorify Him and enjoy communion with Him forever (WSC 1, WLC 1; see also HC 6). God supplied us with duties to guide us creatures according to His infinite wisdom: to love the LORD with our whole being and our neighbors as ourselves (Matthew 22:37-40, HC 4; see also WSC 39-42; WLC 91-98); at the time of creation in particular, to fill, subdue, and have dominion over all creation (Gen. 1:28) unto the prospering of humanity and creation.

Clearly, the current state of humanity is not filled with universal prosperity and the earth suffers under the misdeeds of humankind. The transgression of our first parents and every person since then have brought a curse upon the human race and the created world alike (WCF 6, WSC 13-19, WLC 21-28, and BC 14). God has thankfully not left us to the full desserts of our transgressions, but of His grace supplies a way of return to communion with Him through His Son (WCF 7, 8; WSC 20, 21; WLC 30, 31; HC 18; BC 17)—a matter beyond the scope of this section, chapter or dissertation.

Among the gracious gifts supplied from God to all humanity is an ability to fight against sickness, disease, and death as well as to cultivate creation unto its and humanity's prospering. One of my particular callings is to contribute (ever so slightly) to these ends. As discussed in more detail in the following section, my research seeks out greater understanding of the natural world to increase access to pharmaceuticals by increased availability and lower costs, reduce the burden of sourcing such compounds from species threatened with extinction, allow for the development of better drugs, and improve how humanity synthesizes chemicals by reducing the use of hazardous chemicals. These pursuits are unto the blessings of my neighbors near and far off, fulfilling the law of God. These pursuits, though mixed with much weakness and sin (WCF 16.5), empowered by the Spirit of God (WCF 16.3) are a pleasing offering to God in Christ

(WCF 16.6). These pursuits, thus, have been and continue to be worth the great time, effort, and travail that I have devoted to them.

8.2.2. Scientific and societal motivations

Although the breaking of C-C bonds by P450s is featured in the projects on which I have been working, the reasons for pursuing these research projects are not the same. The project on which I have spent the most time and effort relates to the medicinal plant *Camptotheca acuminata*. A relative of dogwoods (they share the family Cornales) native to the Himalayan foothills of southern China and Tibet, *Camptotheca* produces a terpene indole alkaloid (TIA) camptothecin used as a cancer chemotherapeutic (4). Derivatives of camptothecin had an estimated worldwide market size exceeding $1 billion in 2003, $2.2 billion in 2008, and a drop to $600 million in 2018 owing to the widespread release of generic suppliers (5). Whatever the market share, camptothecin and its derivatives remain in use for multiple cancer types including colon, non-small cell lung, pancreatic, gastric, breast, and cervical cancers (5–7).

The demand for camptothecin is not without drawbacks. The structure of the compound is not accessible by organic syntheses that are economically competitive. In *Camptotheca* however, camptothecin accumulates to no more than 4 mg kg^{-1} dry weight in young leaves (4). Let me put this in perspective: a leaf usually has a density around 0.2 kg L^{-1} when dry (8) equating to 23 kg dried leaves in a 30 gallon bag of yard waste. Thus, a bag full of dried, young *Camptotheca* leaves might contain 92 mg camptothecin. Patients are often dosed at 1.5 mg m^{-2} or approximately 3 mg per dose for a typical adult. A treatment cycle includes five doses every 21 days resulting in 15 mg per treatment cycle per patient. In sum, ideal conditions produce six patients one cycle of treatment from a bag of yard waste.

This number of doses may sound like many until a number of factors are considered. First, young leaves are not as abundant as the leaves that fall to the ground in fall. Continual harvesting of young leaves will also kill the plant. Second, treatments often last up to four cycles for a given patient. Third, having to extract those precious few milligrams of camptothecin from leaves is a major source of expense. Lastly, *Camptotheca* does not grow well outside of its native range, and overharvesting has led to its endangerment (9). Consequently, developing an alternative source to camptothecin besides *Camptotheca* is necessary.

Because camptothecin is a biomolecule synthesized within *Camptotheca*, knowing which enzymes constitute the biosynthetic pathway that makes camptothecin would enable producing this important TIA in an organism besides *Camptotheca*. My work has contributed to this larger goal by first identifying and then characterizing one set of P450s that catalyze a C-C bond breaking reaction occurring *en route* to camptothecin (Fig. 8.2). As it turns out, this CYP-catalyzed reaction is required for the biosynthesis of many other bioactive compounds and pharmaceuticals that, like camptothecin, are expensive to source from the producing plant species. By better studying the molecular machines performing this crucial biosynthetic step, my work promises to enable others to produce these important compounds in easily-cultured organisms like *E. coli*, brewer's yeast (*Saccharomyces cerevisiae*), or a crop like some relative of tobacco (*Nicotiana benthamiana*).

The applications from my other project are more immediate and straightforward. The human P450 CYP17A1 catalyzes several transformations important to steroid processing. One of these is a C-C lyase reaction at the 17^{th} carbon of this scaffold (hence, 17A1) required to produce the sex-specific hormones (Fig. 8.2). Because several kinds of pancreatic cancer depend on testosterone and other steroids for growth, depriving these cancers of such steroids by inhibiting

CYP17A1 has emerged as a major treatment paradigm (10–12). CYP17A1 contributes to the production of other steroids, too, so finding a means of targeting only the activity that leads to the sex-specific hormones (*i.e.*, the C-C lyase chemistry) can improve these therapies. My work has sought to learn about the C-C bond breaking reaction facilitated by CYP17A1 by reverse engineering other P450s to do the same reaction. By better understanding how this molecular machine does C-C bond breaking work instead of C-H hydroxylation work, this work aims to enable others to develop better chemotherapeutics targeting CYP17A1.

8.3. YEARS OF WORK CONDENSED TO A FEW PAGES

Investigating the C-C bond breaking CYPs from *Camptotheca* first required identifying which of the 244 P450s (13) are capable of facilitating this reaction. After identifying three good candidates, closer assessment revealed two as capable of breaking a C-C bond in loganic acid to produce secologanic acid (Fig. 8.2) and perform the same reaction on another substrate, loganin. Intrigued by this versatility, further work revealed a pair of places within these enzymes that could select for one reactivity or another. Lastly, I probed whether a string of amino acids missing in the third, non-active CYP was responsible for its inability to turnover loganic acid.

Modeling the C-C lyase chemistry of CYP17A1 (Fig. 8.2) with bacterial P450s was filled with difficulties. Neither CYP101A1 nor CYP102A1 performed the reaction on two commercially available compounds. After I synthesized several molecules more likely to be substrates for CYP101A1, the enzyme did not appear to perform the C-C bond breaking reaction on these either. Some molecular modeling comparable to observing a machine work in slow motion suggested that the probe substrates were too mobile in the active site. Selection of another bacterial P450, CYP199A4, more synthetic work, and the help of some collaborators in

Australia have demonstrated promising early results. This collaboration is ongoing, but reverse engineering the C-C lyase chemistry of CYP17A1 is now in sight.

8.3.1. Identifying the CYPs responsible for breaking a C-C bond in loganic acid, a part of camptothecin biosynthesis (Chapter 2)

The search for which protein(s) conduct(s) a particular reaction in a biosynthetic pathway is a puzzle likely to have fascinated the fictional Sherlock Holmes were they known in his era. The search begins by looking where the product of the reaction of interest is found in the plant—young leaves, old leaves, immature fruit, mature fruit, immature bark, mature bark, roots, *etc*.

Knowing where the product compound is found is like the location of a dead body in a murder mystery; it enables the elimination of suspects not present at the scene of the crime. The alibis of each protein are established in one of two ways: Proteomics directly confirms the identity and concentration of each protein, like a time-stamped video of a suspect in the area of a crime. Transcriptomics indirectly infers the identity and amount of each protein by recording the identity and amount of messenger RNAs that encode its production, like matching shoeprints or fingerprints at a crime scene.

The remaining suspects were sorted by their ability to use the murder weapon. The kind of C-C bond breaking reaction that occurs in *Camptotheca* resembles that done by known enzymes in other medicinal plants. Seeing the wound inflicted upon loganic acid to produce secologanic acid suggested a particular kind of P450 (a CYP72A family member) as the size and shape of a wound indicates the caliber of round used. The result of this string of reasoning was three candidate P450s (CYP72A564, CYP72A565, and CYP72A730) for the C-C bond breaking reaction yielding secologanic acid.

These three proteins remained only suspects at this stage. Further evidence was required to confirm their role in the murdering of loganic acid. First, I expressed these proteins in *E. coli* (see the key terms in Section 8.1). Spectroscopic studies related to the blue-shift usually observed upon substrate binding to P450s (Fig. 8.3) suggested only two of the suspects were capable of binding loganic acid. The *coup de grâce* came from restaging the reaction.

Mixing each of the CYPs with the potential substrates (loganic acid and loganin), adding a partner required for delivering electron equivalents for catalysis, and a source of electrons in a NADPH cofactor supplies all the ingredients necessary for the C-C breaking reaction. The compounds in these mixtures were separated and their identity established by using high-resolution mass spectrometry. This technique very precisely establishes the mass of each analyte and confirmed the presence of secologanic acid. In the end, CYP72A564 and CYP72A565 were found guilty of producing secologanic acid from loganic acid and secologanin from loganin (I will now call them secologanic acid synthases (SLASs); CYP72A730 was cleared of all such charges (Fig. 8.4).

8.3.2. Two sites in these P450s toggle substrate selectivity (Chapter 3)

Moving away from my murder mystery analogy and returning to enzymes as molecular machines, I was intrigued that CYP72A564 and CYP72A565 were capable of using multiple substrates. This broadened substrate scope contrasted with other P450s that perform this kind of C-C bond breaking reaction which can utilize only loganin (14–18). The multifunctionality of these *Camptotheca* SLASs presented the possibility of their use for the production of pharmaceuticals besides camptothecin. This possibility comes with an important caveat—being an enzymatic jack-of-all trades and master of none is problematic for the production of a particular compound.

The goal of my next series of studies was thus to determine what enabled the broadened substrate scope of the *Camptotheca* SLASs and what modifications to these enzymes increase their selectivity for particular substrates. Lacking detailed blueprints for these machines in the form of X-ray crystal structures, I created some three-dimensional models to estimate their structure. I then employed several methods for comparing enzymes to deduce which changes between these *Camptotheca* proteins and other proteins were important for function and which were more likely to be superfluous. These efforts yielded four potentially important sites that I subsequently changed in the sequences of CYP72A564 and CYP72A565 before I expressed these modified proteins ("mutants" in the language of biochemistry) using *E. coli*.

I conducted the same assays as with the original ("wild type" in the language of biochemistry) CYPs with some interesting findings (Fig. 8.5). Almost always, making a change to these molecular machines decreased or eliminated their ability to produce secologanic acid or secologanin. However, the amino acid composition at two adjacent sites toggled the selectivity of the SLASs between both substrates, loganic acid only, and loganin only (Table 8.3). Because loganin specificity is the most useful for producing a plethora of medicinal compounds, the His131Phe mutant is particularly helpful for other researchers seeking to make pharmaceuticals like camptothecin using microorganisms (like yeast) or a crop species (like a tobacco plant).

8.3.3. A broken CYP left unfixed (Chapter 4)

CYP72A564 and CYP72A565 (the SLASs) have a misfit cousin, CYP72A730. Despite being over 80% identical in sequence composition to these other two *Camptotheca* P450s and lurking around where loganic acid is available, CYP72A730 has no interest in the family trade of breaking the C7-C8 bond in loganic acid to make secologanic acid. Observing a distinct three amino acid deletion when comparing CYP72A730 to the *Camptotheca* SLASs, I predicted that

this was the origin of the former's inability to perform the C-C bond scission reaction. As before, I tested this hypothesis by mutating the sites in question: adding three amino acids to CYP72A730 in the appropriate location and removing these same amino acids from CYP72A564 (the SLAS most similar to CYP72A730).

My hypothesis that the three amino acid deletion was the cause of CYP72A730's inactivity was wrong. Neither adding the three amino acids to CYP72A730 nor removing them from CYP72A564 had detectable effects on secologanic acid or secologanin production (Fig. 8.6). My reworked hypothesis is that some combination of the 30+-other differences between CYP72A730 and the *Camptotheca* SLASs produces the former's inability to convert loganic acid to secologanic acid. More specifically, 11 of these changes represent drastic changes in shape, charge, and function increasing the likelihood that they are responsible for this inactivity. Although unable to address it, a question remains as to what CYP72A730 does within *Camptotheca* since it cannot produce secologanic acid, and yet it is highly expressed in camptothecin-producing tissues. This cousin may be on its way to developing new functions without abandoning its inherited expression patterns. Only time (and much hard work) will tell.

8.3.4. Failures mixed with occasional success in reverse engineering CYP17A1 C-C lyase chemistry using bacterial P450s (Chapters 6 and 7)

A definitive demonstration of understanding how a machine functions is the ability to completely disassemble and reassemble it in working order. I took this approach to studying the C-C lyase reaction mediated by human CYP17A1 in steroid processing (Fig. 8.2). My earliest attempts using CYP101A1 and CYP102A1 revealed my ignorance about much of how CYP17A1 functions (the subject of Chapter 6). Even after refining my understanding using some molecular modeling to watch these machines function in slow motion, I could not coax

CYP101A1 to break a C-C bond like CYP17A1 does by using specially designed molecules that I had synthesized.

These failures have further refined my hypotheses for how CYP17A1 achieves C-C bond scission. Steady and precise positioning of the C-C bond to be broken over the heme appears essential to turnover. A search of the scientific literature led to the identification of a bacterial P450 that I hypothesized could meet these requirements—CYP199A4 (Chapter 7). I subsequently synthesized an appropriate model substrate that contained the features required for CYP199A4 to bind it and for it to resemble the substrates of CYP17A1. After shipping this sample to collaborators in Australia who have worked with CYP199A4 extensively, the earliest of results were not promising.

First, the compound that I had synthesized was 95% pure, but a major contaminant was a better substrate for CYP199A4. Second, my model compound did not appear to be turned over by wild type CYP199A4 even after the contaminant was consumed. While I synthesized a new batch of the model compound to 99+% purity, my collaborators screened several mutants of CYP199A4 to see which were better able to consume this compound. That is the current state of this project. With a cleaner substrate sample and a version of CYP199A4 better able to consume it, a series of experiments will be used to probe how CYP199A4 conducts the C-C lyase reaction. The results of these experiments can then be applied to CYP17A1 to inform the development of better drugs that target its C-C bond breaking activity. Moreover, the insights into how P450s in general function will inform those seeking to modify all sorts of CYPs to produce chemicals in ways less harmful for the environment.

8.4. WHERE DOES ONE GO FROM HERE?

The advances in knowledge and understanding about the chemical world are at once sizable and miniscule. My work is summarized in this dissertation over eight chapters and several hundred pages. In relation to the created world considered as a whole, however, its advances appear more minute. Because a stated motivation of my work is to the benefit of society by alleviating burdens associated with disease and stewarding natural resources. I will take this closing section to delineate my most salient (if small) contributions to these efforts.

My work with the *Camptotheca* SLASs shows the greatest promise towards these ends. As I outlined in section 8.2.2., obtaining the cancer chemotherapeutic camptothecin is currently limited to isolating this compound from producing species like *Camptotheca*. This process is lengthy and expensive; also, it has led to overharvesting and the endangerment of this species. By first identifying two P450s that perform a critical step in camptothecin biosynthesis (the breaking of a C-C bond in loganic acid to form secologanic acid) and subsequently determining a means of improving the selectivity of these CYPs, development of a system that incorporates these and other enzymes to produce camptothecin is now possible.

Creating such a system for camptothecin production features two main challenges. First, many enzymes that take compounds derived from secologanic acid towards camptothecin remain unknown. The approach I employed to identify the SLASs will be important for identifying candidates for each of these reactions before testing for the hypothesized activity. Second, optimizing the slower steps of camptothecin biosynthesis (like the one performed by the SLASs) requires knowing what about particular enzymes enables their functions. My work with the SLASs is a major advancement for understanding key positions for modification to improve this particular reaction, but its importance for the production of medicinal compounds remains

theoretical. Others will take up these *Camptotheca* P450s, especially the mutants that I have identified, and optimize their activities for production of TIAs like camptothecin in microbes and/or crop species. Producing camptothecin and other medicines from plants in this way promises to reduce the prices of these drugs, allow for the development of new drugs from these medicines, improve the environmental footprint of obtaining these compounds by reducing the amount of purification required, and reduce the threat of extinction faced by wild populations of many medicinal plants.

The preliminary nature of results from reverse engineering the C-C lyase activity of CYP17A1 renders prognosticating more difficult. Having a model compound and enzyme able to turn it over are an essential part of probing this important reaction that occurs in steroid processing. First, a series of experiments already conducted with CYP17A1 will confirm that the reaction in CYP199A4 occurs in the same way as it does in CYP17A1. From here, another set of experiments too difficult to conduct with CYP17A1 will delineate which of the proposed mechanisms is responsible for the C-C bond breaking reaction. These results, though distant, can then inform the development of better drugs targeting CYP17A1 as a prostate cancer therapy. The insights into how P450s function gained along the way can also inform other researchers seeking to use these versatile enzymes as catalysts for more environmentally friendly production of chemicals.

8.5. TABLES AND FIGURES

Table 8.1. Cytochrome P450 sequences in each domain of life.

Domain	*Number CYPs*	*Number Species*	*Avg. per Species*
Eukaryotes	319,877	4,828	66
Bacteria	123,494	12,992	9.5
Archaea	1,105	413	2.7
Viruses	102	56	1.8
	444,578	18,289	

All values are taken from the InterPro database (2) for superfamily "Cytochrome P450" (IPR001128), accessed 14 March 2022.

Table 8.2. Cytochrome P450 sequences in kingdoms among some eukaryotes.

Kingdom	*Number CYPs*	*Number Species*	*Avg. per Species*
Animals	102,600	1,812	57
Plants	92,971	1,267	73
Fungi	119,750	1,498	80
	315,321	4,577	

All values are taken from the InterPro database (2) for superfamily "Cytochrome P450" (IPR001128), accessed 14 March 2022.

Table 8.3. Adjacent amino acid positions in *Camptotheca* CYP72A564 and CYP72A565 toggle selectivity for loganin and loganic acid.

Position 131	**Position 132**	*Specificity*	*Speed*
His	His	None	-
Phe	His	Loganin	
His	Asp	Loganic acid	
Phe	Asp	Loganin	

Figure 8.1. Cytochrome P450 sequences from across the biosphere.
Sunburst plot of over 446,000 cytochrome P450-enoding gene sequences from eukaryotes (green), archaea (yellow), bacteria (violet), and viruses (pink). The size of each wedge corresponds to the number of cytochrome P450 sequences encoded.

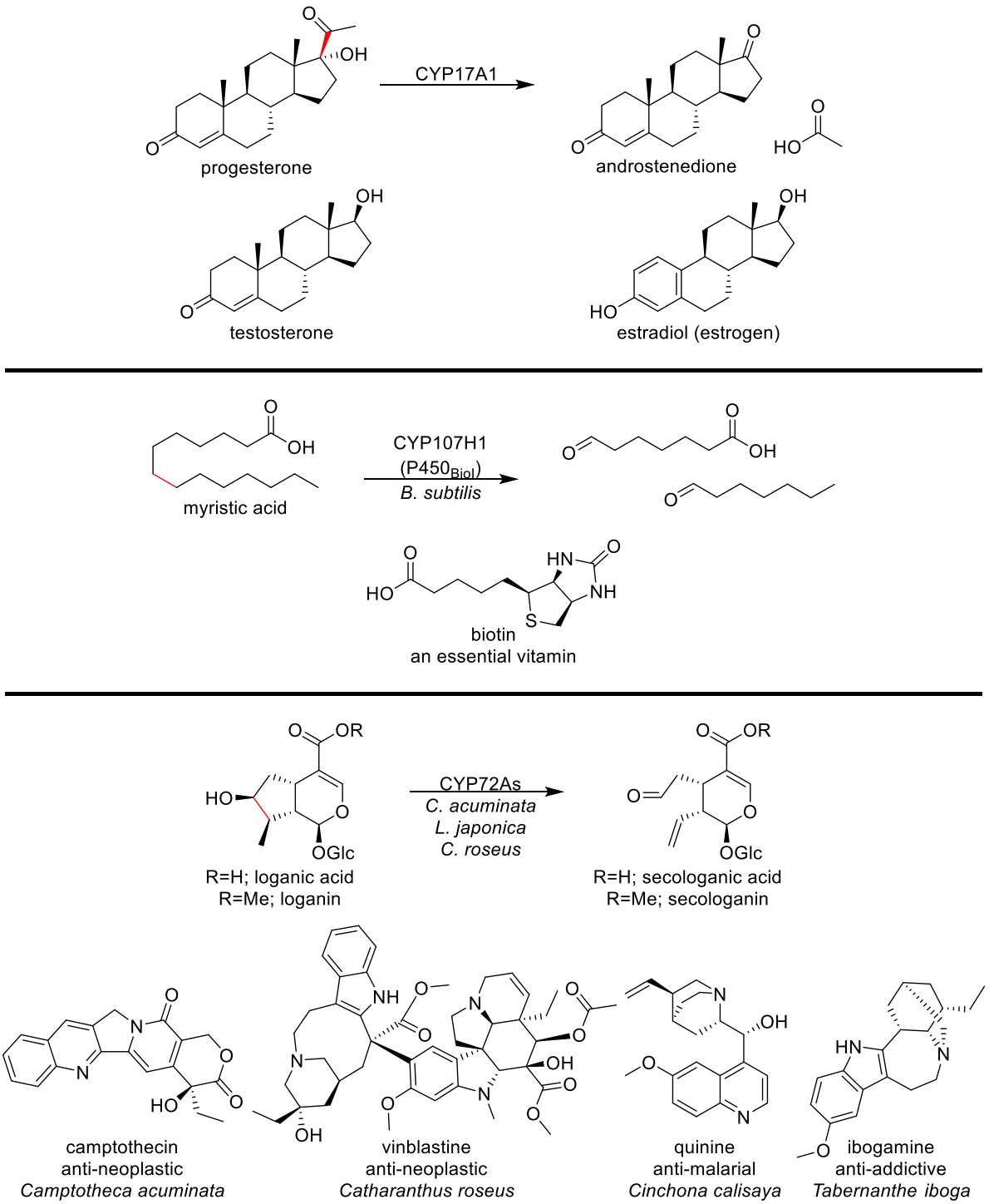


Figure 8.2. The diversity of carbon scaffolds found in biomolecules transformed by C-C bond breaking reactions.

Three cytochrome P450-mediated C-C bond breaking reactions are shown. The C-C bond broken in the reaction is shown in red. Below each reaction are some important biomolecules derived from the product of the shown reaction.

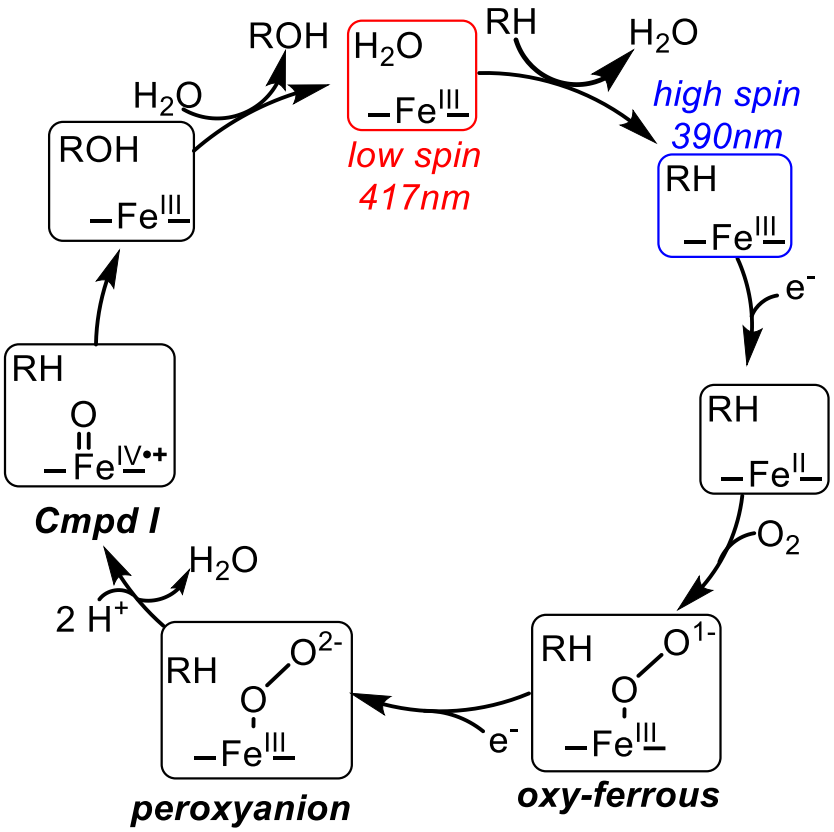


Figure 8.3. Cytochrome P450 catalytic cycle.
Starting with oxidized iron (iron (III)) bound by a water at its sixth position, a generic substrate (RH) binds in the active site. Often, the substrate binding event ejects water from the sixth position and creates a five-coordinate iron(III). The consequence of this loss of water is a spectral shift: The iron shifts from absorbing light around a 417 nm wavelength towards absorbing light closer to the blue range around 390 nm. An electron (e^-) from a partner protein reduce the iron (now iron(II)) enabling oxygen (O_2). Another electron and some protons (H^+) cause the oxygen to break apart into water and a highly reactive species called Compound I. Compound I reacts with the substrate (here, hydroxylating it). Once water replaces the newly formed alcohol, the cycle can repeat.

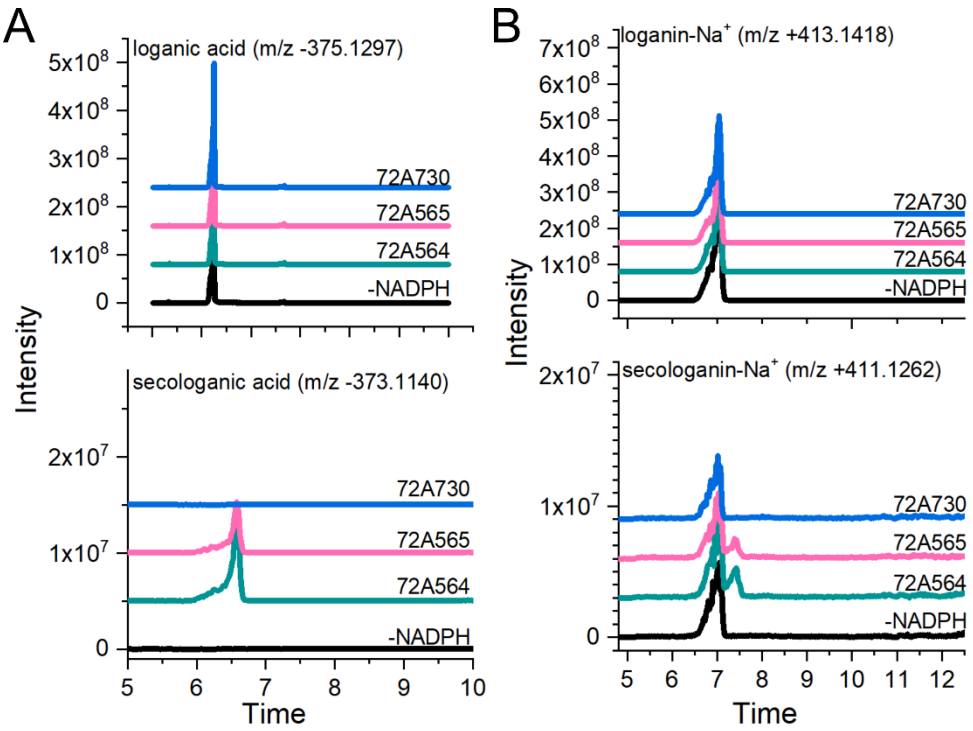


Figure 8.4. Two *Camptotheca* enzymes can perform C-C bond breaking reactions. After mixing all the components required for a reaction and incubating for some time, the compounds in the reaction were separated (corresponding to Time on the x-axis) and their identity inferred by looking for the appropriate mass. Reactions without the electron source NADPH were negative controls where no product can be formed. Reactions containing loganic acid (A) yielded secologanic acid when CYP72A564 and CYP72A565 were present but not when CYP72A730 was present (lower panel). Reactions containing loganin (B) likewise yielded secologanin when CYP72A564 and CYP72A565 were present but not when CYP72A730 was present (lower panel). Taller peaks (Intensity on the y-axis) indicate that more of that compound is present.

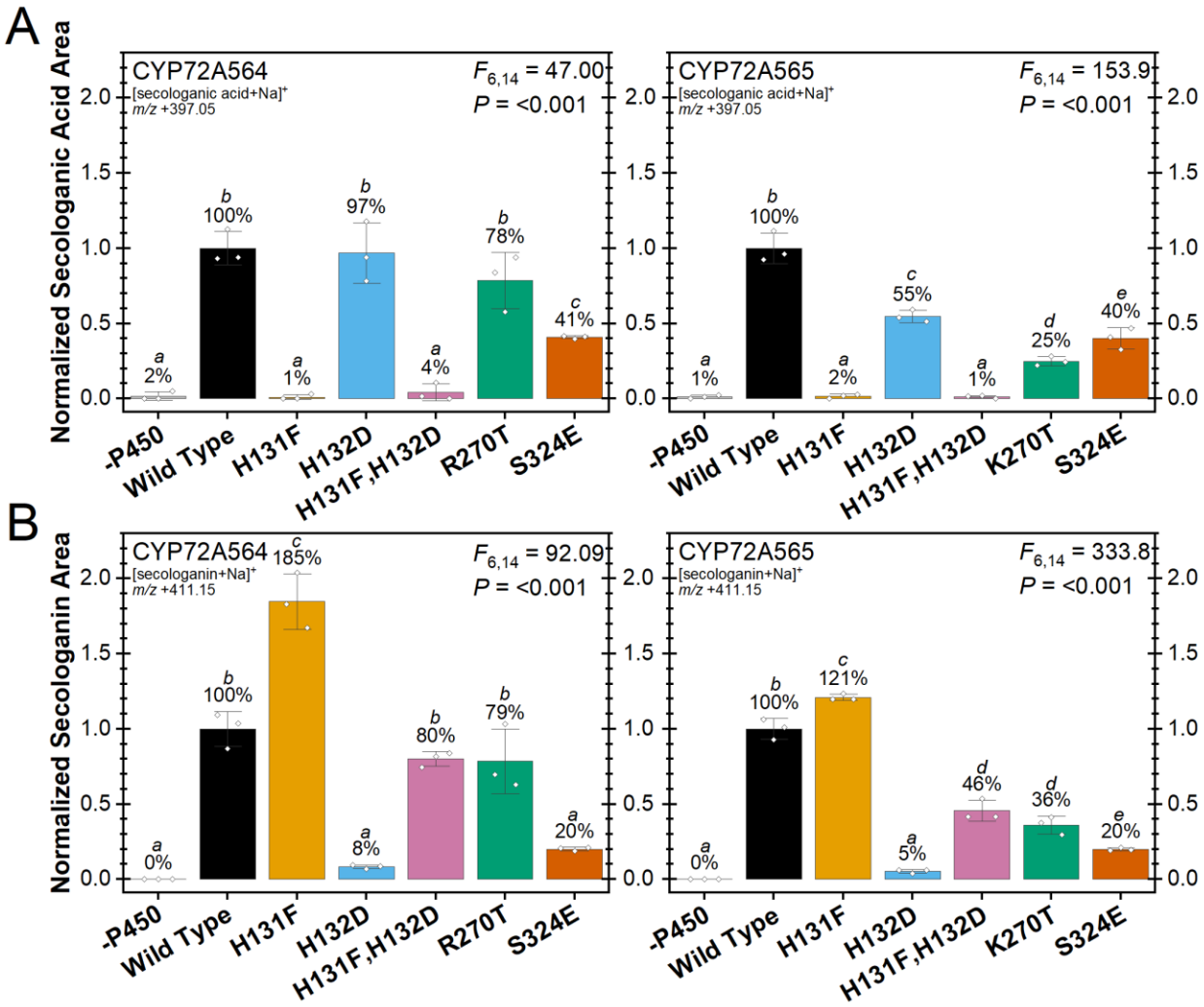


Figure 8.5. Mutating *Camptotheca* SLASs generally decreases activity, but occasionally increases specificity.

These bar graphs display the amount of secologanic acid (A) and secologanin (B) produced by each mutant enzyme relative to the wild type. Any CYP containing H131F did not produce secologanic acid from loganic acid but did produce secologanin from loganin. When only the H132D mutation was made to a CYP, that P450 still produced secologanic acid from loganic acid, but not secologanin from loganin.

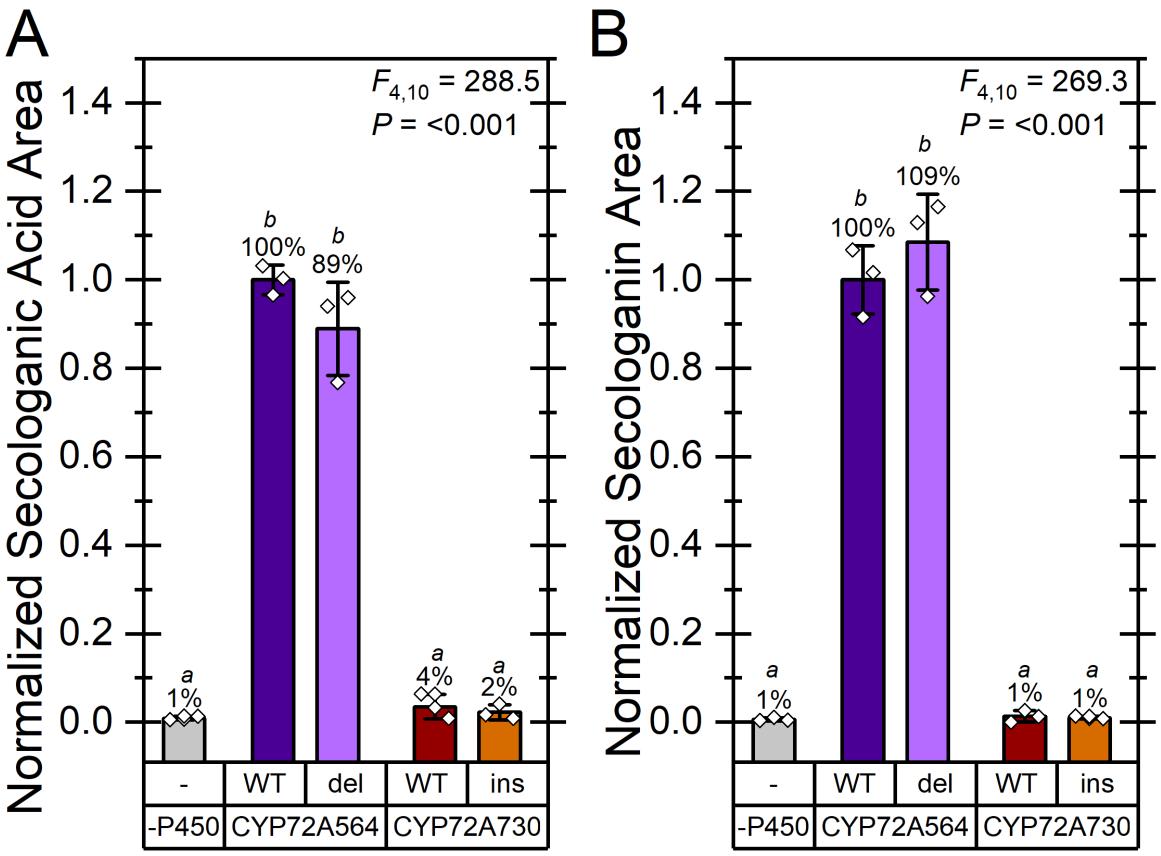


Figure 8.6. The three amino acid deletion in CYP72A730 is not the cause of its inactivity. These bar graphs display the amount of secologanic acid (A) and secologanin (B) produced by each mutant enzyme relative to the CYP72A564 wild type. Whether the three amino acids in question were removed from CYP72A564 or added to CYP72A730, they did not significantly affect the ability or inability of these enzymes to produce secologanic acid from loganic acid (A) or secologanin from loganin (B).

8.6. REFERENCES

1. Classification of protein families - InterPro *InterPro*. [online] https://www.ebi.ac.uk/interpro/entry/InterPro/#table (Accessed March 16, 2022)
2. Cytochrome P450 (IPR001128) - InterPro entry - InterPro *InterPro*. [online] https://www.ebi.ac.uk/interpro/entry/InterPro/IPR001128/ (Accessed March 14, 2022)
3. Guengerich, F. P., and Yoshimoto, F. K. (2018) Formation and cleavage of C–C bonds by enzymatic oxidation–reduction reactions. *Chem. Rev.* **118**, 6573–6655
4. Lorence, A., and Nessler, C. L. (2004) Camptothecin, over four decades of surprising findings. *Phytochemistry*. **65**, 2735–2749
5. Bailly, C. (2019) Irinotecan: 25 years of cancer treatment. *Pharmacol. Res.* **148**, 104398
6. Garst, J. (2007) Topotecan: An evolving option in the treatment of relapsed small cell lung cancer. *Ther. Clin. Risk Manag.* **3**, 1087–1095
7. Jaaks, P., Coker, E. A., Vis, D. J., Edwards, O., Carpenter, E. F., Leto, S. M., Dwane, L., Sassi, F., Lightfoot, H., Barthorpe, S., van der Meer, D., Yang, W., Beck, A., Mironenko, T., Hall, C., Hall, J., Mali, I., Richardson, L., Tolley, C., Morris, J., Thomas, F., Lleshi, E., Aben, N., Benes, C. H., Bertotti, A., Trusolino, L., Wessels, L., and Garnett, M. J. (2022) Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature*. **603**, 166–173
8. Shipley, B., and Vu, T.-T. (2002) Dry matter content as a measure of dry matter concentration in plants and their parts. *New Phytol.* **153**, 359–364
9. Li, S., Zhang, W., Northrup, K., and Zhang, D. (2017) Distribution of Camptotheca Decaisne: Endangered Status. *Pharm. Crops*. **5**, 135–139
10. Gomez, L., Kovac, J. R., and Lamb, D. J. (2015) CYP17A1 inhibitors in castration-resistant prostate cancer. *Steroids*. **95**, 80–87
11. Njar, V. C. O., and Brodie, A. M. H. (2015) Discovery and development of galeterone (TOK-001 or VN/124-1) for the treatment of all stages of prostate cancer. *J. Med. Chem.* **58**, 2077–2087
12. Vogt, C. D., Bart, A. G., Yadav, R., Scott, E. E., and Aubé, J. (2021) Effects of fluorine substitution on substrate conversion by cytochromes P450 17A1 and 21A2. *Org. Biomol. Chem.* **19**, 7664–7669
13. Kang, M., Fu, R., Zhang, P., Lou, S., Yang, X., Chen, Y., Ma, T., Zhang, Y., Xi, Z., and Liu, J. (2021) A chromosome-level *Camptotheca acuminata* genome assembly provides insights into the evolutionary origin of camptothecin biosynthesis. *Nat. Commun.* **12**, 3531
14. Yamamoto, H., Katano, N., Ooi, A., and Inoue, K. (2000) Secologanin synthase which catalyzes the oxidative cleavage of loganin into secologanin is a cytochrome P450. *Phytochemistry*. **53**, 7–12
15. Irmler, S., Schröder, G., St-Pierre, B., Crouch, N. P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J. (2000) Indole alkaloid biosynthesis in *Catharanthus roseus*: New enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. *Plant J.* **24**, 797–804
16. Dugé de Bernonville, T., Foureau, E., Parage, C., Lanoue, A., Clastre, M., Londono, M. A., Oudin, A., Houillé, B., Papon, N., Besseau, S., Glévarec, G., Atehortùa, L., Giglioli-Guivarc'h, N., St-Pierre, B., De Luca, V., O'Connor, S. E., and Courdavault, V. (2015) Characterization of a second secologanin synthase isoform producing both secologanin and

secoxyloganin allows enhanced *de novo* assembly of a *Catharanthus roseus* transcriptome. *BMC Genomics*. **16**, 619
17. Rather, G. A., Sharma, A., Misra, P., Kumar, A., Kaul, V., and Lattoo, S. K. (2020) Molecular characterization and overexpression analyses of secologanin synthase to understand the regulation of camptothecin biosynthesis in *Nothapodytes nimmoniana* (Graham.) Mabb. *Protoplasma*. **257**, 391–405
18. Yang, M., Wang, Q., Liu, Y., Hao, X., Wang, C., Liang, Y., Chen, J., Xiao, Y., and Kai, G. (2021) Divergent camptothecin biosynthetic pathway in *Ophiorrhiza pumila*. *BMC Biol.* **19**, 122

APPENDIX A. CYP72A1_LOGANIN.PDB

The supplementary file contains a protein structure file (*.pdb) developed as a homology model of *Catharanthus roseus* CYP72A1 with loganin docked in the active site. Chapter 2 supplies information about the construction of this homology model.

APPENDIX B: CYP72A564_LOGANIC-ACID.PDB

The supplementary file contains a protein structure file (*.pdb) developed as a homology model of *Camptotheca acuminata* CYP72A564 with loganic acid docked in the active site. Chapter 2 supplies information about the construction of this homology model.

APPENDIX C: CYP72A565_LOGANIC-ACID.PDB

The supplementary file contains a protein structure file (*.pdb) developed as a homology model of *Camptotheca acuminata* CYP72A565 with loganic acid docked in the active site. Chapter 2 supplies information about the construction of this homology model.

APPENDIX D: CYP72A730_LOGANIC-ACID.PDB

The supplementary file contains a protein structure file (*.pdb) developed as a homology model of *Camptotheca acuminata* CYP72A730 with loganic acid docked in the active site. Chapter 2 supplies information about the construction of this homology model.

APPENDIX E. MULTIPLE SEQUENCE ALIGNMENT WITH ALL SLS, SLAS, 7DLH, ANCESTRAL SEQUENCES.XLSX

The supplementary file contains a multiple sequence alignment of all genes labeled secologanin synthase (SLS), SLS-like, secologanic acid synthase (SLS), 7-deoxyloganic acid hydroxylase (7DLH), or 7DLH-like from secoiridoid producing species in GenBank as of July 2020 along with the sequences from an ancestral sequence reconstruction. Chapter 3 provides more details including how sequences were selected and aligned, the node to which each of the sequences corresponds with the produced phylogeny, and how the ancestral sequences were inferred.

APPENDIX F: ANCESTRAL SEQUENCES SITE-BY-SITE PROBABILITIES.TXT

The supplementary file contains the site-by-site amino acid probabilities for the ancestral sequences inferred in Chapter 3. Chapter 3 provides more details about the process of ancestral sequence reconstruction that resulted in these data.

APPENDIX G: SLS-SLAS_ANCESTOR-LOGANIC-ACID.PDB

The supplementary file contains a protein structure file (*.pdb) developed as a homology model of the secologanin synthase, secologanic acid synthase (SLS/SLAS) common ancestor with loganic acid docked in the active site. Chapter 3 supplies information about inferring this ancestral sequence and construction of this homology model.