Summer 8-15-2018

Plant Metabolic Pathways and Regulatory Networks for Aromatic Amino Acids and Hormones

Cynthia Holland
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds
Part of the Agriculture Commons, Biochemistry Commons, and the Plant Sciences Commons

Recommended Citation
https://openscholarship.wustl.edu/art_sci_etds/1627

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Plant Metabolic Pathways and Regulatory Networks for Aromatic Amino Acids and Hormones
by
Cynthia K. Holland

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2018
St. Louis, Missouri
# Table of Contents

List of Figures.................................................................................................................. vi
List of Tables ...................................................................................................................... ix
Acknowledgements .......................................................................................................... x
Abstract ........................................................................................................................... xii

Chapter 1: Introduction: Plant Metabolic Pathways and Regulatory Networks for Aromatic
Amino Acids and Hormones ................................................................................................. 1
  Overview of aromatic amino acids in plants ................................................................. 3
  Chorismate, the branch point metabolite ....................................................................... 5
  Chorismate mutase in plants ....................................................................................... 6
  Tyrosine and phenylalanine biosynthesis in plants ................................................... 8
  Prephenate and arogenate dehydrogenases .............................................................. 9
  Prephenate aminotransferase .................................................................................. 12
  Introduction to plant hormones .............................................................................. 14
  Auxins: Indole-3-acetic acid and phenylacetic acid ............................................... 14
  Benzoates ................................................................................................................ 17
  Jasmonates .............................................................................................................. 18
  GH3 proteins regulate the pre-receptor concentration of plant hormones ........... 20
  Structure/function studies of GH3 proteins ........................................................... 22
  Conclusions ........................................................................................................... 25
  References ............................................................................................................. 28

Chapter 2: Molecular Basis of the Evolution of Alternative Tyrosine Biosynthetic Pathways in
Plants ................................................................................................................................ 40
  Abstract ................................................................................................................... 41
  Introduction ............................................................................................................ 42
  Methods ................................................................................................................... 45
    Identification of ncADH enzymes from plants .................................................... 45
    Recombinant protein expression and purification and site directed mutagenesis .... 46
    GmPDH1 crystallization ....................................................................................... 48
    ADH and PDH assay .......................................................................................... 50
    Computational Substrate Docking ...................................................................... 51
    Data availability ................................................................................................. 51
  Results ..................................................................................................................... 52
Identification and biochemical analysis of noncanonical ADH in legumes.............52
X-ray crystal structure of soybean PDH............................................................53
Identification of a residue that confers TyrA substrate specificity .....................58
Altered substrate specificity simultaneously affects Tyr-sensitivity.........................61
Mutating Asp218 introduces PDH activity in divergent plant ADH enzymes...........63
Discussion...........................................................................................................65
Acknowledgements............................................................................................72
References..........................................................................................................74

Chapter 3: Reaction mechanism of prephenate dehydrogenase from the alternate tyrosine biosynthesis pathway in plants ........................................................80
Abstract............................................................................................................81
Introduction.........................................................................................................82
Results ................................................................................................................83
Discussion..........................................................................................................90
Methods................................................................................................................93
  Computational substrate docking ......................................................................93
  Generation of site-directed mutants .................................................................93
  Protein expression and purification ..................................................................93
  Kinetic analysis of wild-type and mutant proteins.............................................94
Acknowledgements............................................................................................94
References..........................................................................................................95

Chapter 4: Evolution of Allosteric Regulation in Chorismate Mutases from Early Plants......99
Abstract............................................................................................................100
Introduction.........................................................................................................101
Methods...............................................................................................................104
  Generation of expression constructs ..............................................................104
  Protein expression and purification .................................................................104
  Preparation of chorismate ...............................................................................106
  Chorismate mutase assays ..............................................................................107
  X-ray crystallography of Physcomitrella patens chorismate mutase 1 ...............107
Results ................................................................................................................108
  Functional comparisons of chorismate mutases from early plants ................108
  Differential regulation of basal plant chorismate mutases by amino acid effectors....109
  Overall structure, active site, and effector site of P. patens chorismate mutase 1......112
Discussion..........................................................................................................116
Acknowledgements............................................................................................123
References..........................................................................................................124

Chapter 5: Structural Basis for Substrate Recognition and Inhibition of Prephenate
Aminotransferase from Arabidopsis.......................................................................127
Abstract............................................................................................................128
Chapter 7: Characterization of a Chorismate-Conjugating Class of Brassicaceae GH3 Proteins

Abstract .......................................................................................................................... 186
Introduction .................................................................................................................... 187
Methods .......................................................................................................................... 188
Protein expression and purification.................................................................190
Enzyme assays ..............................................................................................191
Pyrophosphate release assays.........................................................................192
Mass spectrometry..........................................................................................192
X-ray crystallography.......................................................................................192
Rapid chemical-quench kinetics of AtGH3.12 ................................................193
Plant and pathogen materials and growth conditions........................................193
P. syringae inoculation and quantification of bacterial growth..............................194
Results ............................................................................................................194
AtGH3.7 is a chorismate-cysteine synthetase......................................................194
Kinetic reevaluation of AtGH3.12 reveals it is a chorismate-glutamate synthetase...195
Salicylic acid inhibits formation of both chorismate conjugates.........................201
Structural analysis of AtGH3.12 co-crystalized with chorismate in the active site...201
Site-directed Mutant Analysis..........................................................................204
Arabidopsis gh3.7 mutants show wild-type pathogen response when infected with
PstDC3000........................................................................................................206
Arabidopsis gh3.7/gh3.12 double mutants are more susceptible to infection than wild-
type................................................................................................................207
Discussion........................................................................................................208
Acknowledgements.........................................................................................213
References.......................................................................................................214
Chapter 8: Conclusions and Future Directions..................................................218
Abstract..........................................................................................................219
Chapter 2: Summary of the regulation of soybean prephenate and arogenate dehydrogenases
........................................................................................................................220
Chapter 2: Future Directions..........................................................................221
Chapter 3: Summary of the enzymatic mechanism of soybean prephenate dehydrogenase223
Chapter 3: Future Directions..........................................................................225
Chapter 4: Summary of the regulation and evolution of chorismate mutases from early
land plants.......................................................................................................226
Chapter 4: Future Directions..........................................................................228
Chapter 5: Summary of the regulation of prephenate aminotransferase................229
Chapter 5: Future Directions..........................................................................231
Chapter 6: Summary of the substrate fidelity of auxin-using GH3 proteins.........234
Chapter 6: Future Directions..........................................................................236
Chapter 7: Summary of the chorismate-conjugating GH3 proteins AtGH3.7 and AtGH3.12
..........................................................................................................................238
Chapter 7: Future Directions..........................................................................240
Conclusions of the thesis ...............................................................................242
References.......................................................................................................245
Appendix I: Identification of a noroxomaritidine reductase with Amaryllidaceae alkaloid biosynthesis related activities ................................................................. 249
Appendix II: Dissonance strikes a chord in stilbene synthesizers ................................. 263
Appendix III: Arabidopsis: The original plant chassis organism .................................... 266
Appendix IV: Structural biology of jasmonic acid metabolism and responses in plants .... 275
Curriculum Vitae .................................................................................................. 300
List of Figures

Chapter 1
Figure 1. Aromatic amino acid pathways contribute to the biosynthesis of many plant secondary metabolites .......................................................... 3
Figure 2. Overview of tyrosine and phenylalanine biosynthesis in plants ........................................ 5
Figure 3. Chorismate mutase generates prephenate, the precursor to tyrosine and phenylalanine ........................................................................ 6
Figure 4. Diversity of chorismate mutases across plants ................................................................. 8
Figure 5. Prephenate dehydrogenase reaction ...................................................................... 10
Figure 6. Prephenate aminotransferase (PAT) reaction ............................................................ 13
Figure 7. Overview of plant hormones and hormone homeostasis .......................................... 15
Figure 8. Comparison of the IAA and PAA biosynthetic from tryptophan and phenylalanine, respectively ......................................................................................... 16
Figure 9. Pre-receptor modulation of plant hormones occurs by GH3 acyl acid amidohydrolases .............................................................................................. 21
Figure 10. GH3 protein reaction mechanism ........................................................................ 22
Figure 11. Three-dimensional structure of a GH3 protein ..................................................... 23
Figure 12. Amino acid sequence alignment of the hormone-binding pocket sequences of GH3 proteins from Arabidopsis thaliana (AtGH3), rice (OsGH3), moss (PpGH3), lycophyte (SmGH3), soybean (GmGH3), and maize (ZmGH3) ...................................................................... 24

Chapter 2
Figure 1. Tyr biosynthesis pathways in plants and identification and characterization of noncanonical ADHs ................................................................. 44
Figure 2. X-ray crystal structure of GmPDH1 .............................................................. 55
Figure 3. Electron density maps of citrate and Tyr ligands .................................................. 57
Figure 4. Identification of Asn222 as a determinant of PDH activity and Tyr sensitivity .... 60
Figure 5. Crystal structures of GmPDH1 N222D and M219T/N222D reveal Tyr binding interactions ......................................................................................... 61
Figure 6. Asn222 confers PDH activity to divergent plant ADHs while simultaneously relaxing Tyr sensitivity ........................................................................ 64
Figure 7. Asp222 is conserved in plant ADHs and proteobacteria orthologs .................... 69
Figure 8. Structural comparison of plant PDH, cyanobacterial ADH, and bacterial PDH .... 71

Chapter 3
Figure 1. Molecular docking of prephenate (grey) into the x-ray crystal structure of GmPDH1 in complex with NADP+ (green) ...................................................... 84
Figure 2. PDH sequence alignment of *Medicago truncatula* (XP_003601003.1; MtPDH), *Glycine max* (NP_001304525.1; GmPDH11), and *Phaseolus vulgaris* (XP_007163590.1; PvPDH).

Figure 3. PDH sequence alignment of *Glycine max* (NP_001304525.1; GmPDH11), *Ochrobactrum intermedium* LMG 3301 (EEQ93947.1; OiTyrA), *Sediminispirochaeta smaragdinae* DSM 11293 (ADK80640.1; SsTyrA), and *Methanoseta harundinacea* (KUK94425.1; MhTyrA).

Figure 4. Proposed reaction mechanism for GmPDH1 and the non-canonical PDH of the alternative tyrosine biosynthesis pathway found in legumes.

Chapter 4

Figure 1. Aromatic amino acid biosynthesis from chorismate and the evolutionary relationships of plant chorismate mutases.

Figure 2. Sequence comparisons among the three chorismate mutases from *A. thaliana* (AtCM1-3), *S. cerevisiae* (ScCM), and the five basal plant chorismate mutases from *P. patens* (PpCM1 and PpCM2), *Am. trichopoda* (AmtCM1 and AmtCM2), and *S. moellendorfii* (SmCM).

Figure 3. X-ray crystal structure of the PpCM1•tryptophan complex.

Figure 4. Active site and allosteric effector sites of PpCM1.

Figure 5. Allosteric structural changes in plant chorismate mutases.

Chapter 5

Figure 1. Prephenate aminotransferase catalytic reaction.

Figure 2. Three-dimensional structures of AtPAT wild-type and K306A mutant.

Figure 3. Electron densities of ligands in AtPAT structures.

Figure 4. Sequence comparison of AtPAT and Arabidopsis AATs.

Figure 5. Amino acid titration of AtPAT PLP signal.

Figure 6. Effect of amino acids on AtPAT activity.

Figure 7. Inhibition of plant and microbial PAT by cysteine.

Figure 8. Three-dimensional structures of the AtPAT T84V and AtPAT T84V/K169V mutants.

Figure 9. Summary of AtPAT reaction sequence and roles of active site residues.

Chapter 6

Figure 1. The GH3 enzymatic reaction.

Figure 2. Visualization of the first half and full reactions of GH3.2, GH3.5, and GH3.11 with cognate substrates.

Figure 3. Single turnover kinetics of GH3.2 with IAA, GH3.5 with IAA, and GH3.11 with JA.

Figure 4. Characterization of chimeric GH3.2 and GH3.5 proteins.

Figure 5. GH3 reaction mechanism, including proofreading of adenylated intermediates.
Chapter 7
Figure 1. AtGH3.7 is a chorismate-cysteine synthetase ......................................................196
Figure 2. AtGH3.12 is a chorismate-glutamate synthetase ....................................................198
Figure 3. Adenylation activity of AtGH3.12 ........................................................................199
Figure 4. Mass spectra of chorismate-glutamate ..................................................................199
Figure 5. Single-turnover kinetics of the first half-reaction of GH3.12 .................................200
Figure 6. Active site of AtGH3.12 with chorismate and AMP bound ..................................202
Figure 7. Electron density for chorismate (left) and AMP (right) in the AtGH3.12 active site using 2Fo-Fc maps contoured to 1.0 σ .................................................................203
Figure 8. Arabidopsis gh3.7-1 does not have a pathogen infection phenotype ..................206
Figure 9. Arabidopsis gh3.7-1/gh3.12 double mutant lines are more susceptible to pathogen infection than wild-type or sid2-2 .................................................................208
Figure 10. Chorismate is a precursor to salicylate and the three aromatic amino acids in plants and is synthesized in both the plastid and the cytosol, where GH3 proteins are found ..........211

Chapter 8
Figure 1. Overview of the regulatory network of aromatic amino acid metabolic pathways ....243
List of Tables

Chapter 2
Table 1. Data collection and refinement statistics .............................................................56
Table 2. Kinetic analysis of representative plant PDHs and ADHs ......................................59
Table 3. Steady-state kinetic parameters and effect of tyrosine on mutant GmPDH1, MtPDH1, GmncADH, MtncADH, and SolyncADH ........................................................................65

Chapter 3
Table 1. Steady-state kinetics parameters for GmPDH1 and its mutants ............................85

Chapter 4
Table 1. Steady-state kinetic parameters of chorismate mutases from basal plants ..........109
Table 2. Summary of amino acid effector regulation in chorismate mutases from basal plants ..............................................................................................................................110
Table 3. Summary of crystallographic data collection and refinement statistics ...............114

Chapter 5
Table 1. Summary of AtPAT crystallographic statistics ......................................................136
Table 2. Summary of steady-state kinetics of wild-type and mutant AtPAT .......................140
Table 3. Summary of $K_D$ values for selected amino acids binding to wild-type and mutant AtPAT ..........................................................................................................................143
Table 4. Summary of $K_D$ values for cysteine binding to wild-type and mutant PAT ..........145

Chapter 6
Table 1. Single turnover kinetics data for AtGH3.2, AtGH3.5, and AtGH3 N2/C5 for IAA, BA, and PAA first half-reactions (n=3) .........................................................................................172
Table 2. Steady state kinetics for GH3 N2/C5 ........................................................................175
Table 3. Single turnover kinetics data for AtGH3 N2/C5 first half-reactions (n=3) ..........175

Chapter 7
Table 1. Summary of steady-state kinetics of AtGH3.7 ........................................................195
Table 2. Summary of steady-state kinetics of AtGH3.12 .....................................................197
Table 3. Summary of AtGH3.7 and AtGH3.12 inhibition kinetics with salicylic acid .......201
Table 4. Summary of AtGH3.12·AMP·Chorismate crystallographic statistics ................204
Table 5. Summary of steady-state AtGH3.12 mutant kinetics with chorismate ...............205
The work described in this dissertation was supported by funding from the National Science Foundation Graduate Research Fellowship Program (DGE-1143954 to C.K.H.), the William H. Danforth Plant Sciences Fellowship, and an NSF grant (MCB-1157771 to J.M.J). Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center of the Advanced Photon Source, a national use facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research under Grant DE-AC02-06CH11357.

Cynthia Holland

Washington University in St. Louis

August 2018
Dedicated to the strong women in my life, both past and present: my mom, Alice Pedigo Turner, my grandmother, Martha Pedigo, and Marti Pedigo. None of this would have been possible without you.
ABSTRACT OF THE DISSERTATION

Plant Metabolic Pathways and Regulatory Networks for Aromatic Amino Acids and Hormones

by

Cynthia K. Holland

Doctor of Philosophy in Biology and Biomedical Sciences

Plant and Microbial Biosciences

Washington University in St. Louis, 2018

Professor Joseph Jez, Chair

Unlike humans and other metazoans, organisms such as fungi, bacteria, and plants have retained the enzymatic machinery necessary to synthesize their aromatic amino acids de novo. Chorismate, the final product of the shikimate pathway, is the precursor to the three aromatic amino acids—tryptophan, tyrosine, and phenylalanine—and is upstream of a number of plant growth hormones, including auxins and benzoates. Phenylalanine and tyrosine both stem from the precursor prephenate, which is formed from chorismate by chorismate mutase, and use dehydrogenases, aminotransferases, and dehydratases in their biosynthetic pathways. Although aromatic amino acid biosynthesis is important for protein synthesis, secondary metabolism, and human health, much of what is known about plant aromatic amino acid biosynthesis and its regulation has been inferred from microbial investigations of the pathway.

L-Tyrosine is essential for protein synthesis and is a precursor of numerous specialized metabolites crucial for plant and human health. Tyrosine can be synthesized via two alternative routes by a key regulatory TyrA family enzyme, prephenate or arogenate dehydrogenase (PDH/TyrA_p or ADH/TyrA_a), representing a unique divergence of primary metabolic pathways.
The molecular foundation underlying the evolution of the alternative Tyr pathway is currently unknown. Here we characterized recently-diverged plant PDH and ADHs, obtained the x-ray crystal structure of soybean PDH, and identified a single amino acid residue that defines TyrA substrate specificity and regulation. Structures of mutated PDHs co-crystallized with Tyr indicate that substitutions of Asn222 confers ADH activity and Tyr-sensitivity. Subsequent mutagenesis of the corresponding residue in divergent plant ADHs introduced PDH activity and relaxed Tyr sensitivity, highlighting the critical role of this residue in TyrA substrate specificity that underlies the evolution of alternative Tyr biosynthetic pathways in plants.

The three-dimensional structure of soybean PDH1 allowed for the identification of both the cofactor- and ligand-binding sites. Here, we present steady-state kinetic analysis of twenty site-directed active site mutants of the soybean (Glycine max) PDH compared to wild-type. Molecular docking of the substrate, prephenate, into the active site of the enzyme reveals its potential interactions with the active site residues and makes a case for the importance of each residues in substrate recognition and/or catalysis, most likely through transition state stabilization. Overall, these results suggest that the active site of the enzyme is highly sensitive to any changes, as even subtle alterations substantially reduced the catalytic efficiency of the enzyme.

Chorismate mutase catalyzes the branch point reaction of phenylalanine and tyrosine biosynthesis to generate prephenate. In Arabidopsis thaliana, there are two plastid-localized chorismate mutases that are allosterically regulated (AtCM1 and AtCM3) and one cytosolic isoform (AtCM2) that is unregulated. Previous analysis of plant chorismate mutases suggested that the enzymes from early plants (i.e., bryophytes/moss, lycophytes, and basal angiosperms) formed a clade distinct from the isoforms found in flowering plants; however, no biochemical
information on these enzymes is available. To understand the evolution of allosteric regulation in plant chorismate mutases, we analyzed a basal lineage of plant enzymes homologous to AtCM1 based on sequence similarity. The chorismate mutases from the moss/bryophyte Physcomitrella patens (PpCM1 and PpCM2), the lycophyte Selaginella moellendorffii (SmCM), and the basal angiosperm Amborella trichopoda (AmtCM1 and AmtCM2) were characterized biochemically. Tryptophan was a positive effector for each of the five enzymes examined. Histidine was a weak positive effector for PpCM1 and AmtCM1. Neither tyrosine nor phenylalanine altered the activity of SmCM; however, tyrosine was a negative regulator of the other four enzymes. Phenylalanine down-regulates both moss enzymes and AmtCM2. The 2.0 Å x-ray crystal structure of PpCM1 in complex with the tryptophan identified the allosteric effector site and reveals structural differences between the R- (more active) and T-state (less active) forms of plant chorismate mutases. Molecular insight into the basal plant chorismate mutases guides our understanding of the evolution of allosteric regulation in these enzymes.

Plants make tyrosine and phenylalanine by a different pathway from many microbes, which requires prephenate aminotransferase (PAT) as the key enzyme. PAT produces arogenate, the unique and immediate precursor for both tyrosine and phenylalanine in plants, and also has aspartate aminotransferase (AAT) activity. The molecular mechanisms governing the substrate specificity and activation or inhibition of PAT are currently unknown. Here we present the x-ray crystal structures of wild-type and various mutants of PAT from Arabidopsis thaliana (AtPAT). Steady-state kinetic and ligand binding analyses identified key residues, such as Glu108, that are involved in both keto acid and amino acid substrate specificities and likely contributed to the evolution of PAT activity among class Ib AAT enzymes. Structures of AtPAT mutants co-crystallized with either α-ketoglutarate or pyridoxamine 5’-phosphate (PMP) and glutamate
further define the molecular mechanisms underlying keto acid and amino acid substrate recognition. Furthermore, cysteine was identified as an inhibitor of PAT from *A. thaliana* and *Antirrhinum majus* plants as well as *Chlorobium tepidum* bacterium, uncovering a potential new effector of PAT.

As a way of regulating the concentration of a given plant hormone within the cell, plants contain a class of enzymes known as GH3 proteins that conjugate amino acids to acyl acid hormones to activate or inactivate the hormone molecule. Grouping of the GH3 proteins by substrate preference shows that the largest class of GH3 proteins use the growth-promoting hormone auxin (indole-3-acetic acid, or IAA), and the second largest class uses the defense hormone jasmonic acid; however, some of the auxin-specific GH3 proteins, notably *Arabidopsis thaliana* GH3.5, have been found to use multiple acyl acid substrates, including the auxins IAA and phenylacetic acid and benzoic acid, with roughly the same catalytic efficiencies. Despite the promiscuity in acyl acid substrate preference, the active site residues that are in contact with a ligand in the crystal structure of AtGH3.5 do not vary from those found in the AtGH3.2 active site. Using single-turnover kinetics, the rate data suggests that both proteins prefer IAA as the substrate. In conducting these experiments, we found that the adenylated reaction intermediate was being hydrolyzed into acyl acid and AMP in the absence of the amino acid, a typical feature of pre-transfer editing. Here we show that a non-cognate acyl acid-adenylate intermediate is more easily hydrolysable than the cognate acyl acid-adenylate, likely due to a slowed structural switch that provides a checkpoint for fidelity before the full reaction proceeds.

There are 19 GH3 proteins in the model plant Arabidopsis, but only 9 of those proteins have confirmed substrates. One clade of GH3 proteins is predicted to use benzoate as substrate and includes AtGH3.7 and AtGH3.12. Previously, AtGH3.12 was identified as a 4-hydroxybenzoic
acid-glutamate synthetase and was found to influence pathogen defense responses through the plant defense hormone salicylic acid. Here, we screened 15 plant hormones with 20 amino acids to find that AtGH3.7 is a chorismate-cysteine synthetase. Rescreening of AtGH3.12 found that it too uses chorismate and is a chorismate-glutamate synthetase. These results were confirmed using a combination of x-ray crystallography, site-directed mutagenesis, single-turnover kinetics, and mass spectrometry. Because chorismate is a precursor to the three aromatic amino acids in plants and a precursor to salicylic acid, these proteins are modulating the concentration of a pathway intermediate. Interestingly, gh3.7 mutants did not have the same pathogen susceptibility phenotype as gh3.12 mutants, but the gh3.7/gh3.12 double mutant was more susceptible to pathogen infection than gh3.12 alone. This study reveals how GH3 proteins, which traditionally regulate plant hormones, can also use hormone intermediates to regulate multiple hormones.

The overarching goal of my thesis research was to understand how plants regulate the biosynthesis of aromatic amino acids, which are essential to both protein synthesis and the production of downstream secondary metabolites, including phytohormones. This was accomplished by structural and biochemical studies of prephenate dehydrogenase, arogenate dehydrogenase, chorismate mutase, and prephenate aminotransferase. Also, biochemical assays were used to unveil the basis of substrate fidelity in the GH3 proteins. Finally, the role of aromatic amino acid biosynthesis pathway intermediates was further explored when AtGH3.7 and AtGH3.12 were found to conjugate amino acids to the aromatic branchpoint metabolite chorismate. This research as a whole has advanced our understanding of the regulation of plant aromatic amino acid and hormone metabolic pathways, unveiling how plants have evolved specific regulatory mechanisms and networks that are more complex than previously thought.
Chapter 1

Introduction: Plant Metabolic Pathways and Regulatory Networks for Aromatic Amino Acids and Hormones
Plants have spent the past 425 million years evolving the complex enzymatic machinery needed to generate biosynthetic pathways to synthesize an array of specialized metabolites. These specialized, or secondary, metabolites are used by the plant for development, growth, defense against biotic and abiotic stressors, attracting pollinators and seed dispersers, and responding to the environment. Numerous secondary metabolites are aromatic compounds that use one of the three aromatic amino acids as starting materials (Maeda and Dudareva, 2012). Some of these aromatic amino acid-derived compounds include defense metabolites such as alkaloids, indole glucosinolates, salicylic acid, and cyanogenic glycosides; growth hormones such as indole-3-acetic acid and phenylacetic acid; structural polymers such as lignin; vitamins; electron carriers such as ubiquinone and plastoquinone; pigments; and flavonoids (Dixon, 1999; Winkel-Shirley, 2001; Bonawitz and Chapple, 2010; Nowicka and Kruk, 2010; Vogt, 2010; DellaPenna and Méné-Saffrané, 2011; Mithöfer and Boland, 2012).

To generate this chemical diversity, plants take primary metabolites and modify them to make new compounds with different functions. For example, in the case of amino acid-derived aromatic compounds, the shikimate pathway uses erythrose-4-phosphate and phosphoenoylpyruvate from the pentose phosphate pathway and glycolysis, respectively, to generate chorismate, the precursor to the defense phytohormone salicylic acid and the three aromatic amino acids, \(L\)-phenylalanine, \(L\)-tyrosine, and \(L\)-tryptophan (Maeda and Dudareva, 2012) (Figure 1). The shikimate pathway receives about 30% of the photosynthetically fixed carbon in vascular plants.
Figure 1. Aromatic amino acid pathways contribute to the biosynthesis of many plant secondary metabolites.

The shikimate pathway (green) generates chorismate, the precursor for the tryptophan (Trp) pathway (blue), the phenylalanine/tyrosine (Phe/Tyr) pathways (red), and the pathways leading to folate, phylloquinone, and salicylate. Trp, Phe, and Tyr are precursors to a number of plant metabolites, including some that are essential human nutrients (bold). Other abbreviations: ADCS, aminodeoxychorismate synthase; AS, anthranilate synthase; CM, chorismate mutase; CoA, coenzyme A; ICS, isochorismate synthase. Taken from Maeda and Dudareva, 2012.

Overview of aromatic amino acids in plants

All organisms use the three aromatic amino acids, phenylalanine, tyrosine, and tryptophan, as components of protein synthesis and as precursors of indoles and other aromatics (Figure 2). Tryptophan is the precursor of auxininc hormones (indole-3-acetic acid), alkaloids, niacin
(vitamin B₃), and, in humans, the neurotransmitter serotonin (Maeda and Dudareva, 2012). The structural polymer lignin, flavonoids, isoflavonoids, phenolics, anthocyanins, and benzoic acid are derived from phenylalanine (Dixon, 1999; Winkel-Shirley, 2001; Bonawitz and Chapple, 2010; Nowicka and Kruk, 2010; Vogt, 2010; DellaPenna and Mène-Saffrané, 2011; Mithöfer and Boland, 2012). Similarly, tyrosine-derived plant metabolites include tocopherols (vitamin E), plastoquinone, cyanogenic glycosides, and suberin waxes (Maeda and Dudareva, 2012; Schenck and Maeda, 2018).

Organisms that do not synthesize their own aromatic amino acids must obtain them from the environment. Because aromatic amino acid biosynthesis genes have been lost in metazoans, humans must obtain these amino acids from their diet. In animals, tyrosine can be synthesized from phenylalanine by phenylalanine hydroxylase (Fitzpatrick, 1999). Animals use tryptophan and tyrosine as precursors for the neurotransmitters serotonin and catecholamine, respectively (Fitzpatrick, 1999; Fernstrom and Fernstrom, 2007). Additionally, the absence of this pathway in humans makes the pathway more desirable for developing herbicides and antimicrobials (Baylis, 2000). Glyphosate, a chemical inhibitor of 5-enoylpyruvylshikimate-3-phosphate synthase (EPSPS), leads to a hyper-accumulation of shikimate, thus diverting resources from aromatic amino acid biosynthesis (Steinrücken and Amrhein, 1980). Although aromatic amino acids are important for both agriculture and human health, much of what is known about aromatic amino acid biosynthesis in plants is inferred from microbial studies. As a result, the regulation of aromatic amino acid biosynthesis in plants is not well understood.
**Chorismate, the branch point metabolite**

Chorismate, the product of the shikimate pathway, is the precursor of L-phenylalanine, L-tyrosine, and L-tryptophan (Tzin and Galili, 2010; Maeda and Dudareva, 2012) (**Figure 1**). In this pathway, phosphoenolpyruvate and D-erythrose 4-phosphate bridge central carbon metabolism to aromatic amino acid biosynthesis. This pathway occurs in the plastids of all plants and in microorganisms. Chorismate has three fates in plants as a precursor to the defense phytohormone salicylic acid, the aromatic amino acids tyrosine and phenylalanine, and the indole aromatic amino acid tryptophan.

![Figure 2. Overview of tyrosine and phenylalanine biosynthesis in plants.](image)

Plants synthesize tyrosine and phenylalanine from chorismate, the product of the shikimate pathway. This pathway uses phosphoenolpyruvate (PEP) from glycolysis and erythrose-4-phosphate (E4P) from the pentose phosphate pathway to generate shikimate. Tyr and Phe can also be synthesized outside of the plastid in the cytosol. Arrows spanning membranes indicate transporters. A chorismate mutase transporter has yet to be identified. Abbreviations: chorismate mutase (CM), prephenate aminotransferase (PAT), arogenate dehydrogenase (ADH), arogenate dehydratase (ADT), prephenate dehydrogenase (PDH), prephenate dehydratase (PDT), hydroxyphenylpyruvate aminotransferase (HPP-AT), phenylpyruvate aminotransferase (PPY-AT).
Chorismate exists at the branchpoint of tyrosine and phenylalanine biosynthesis and tryptophan biosynthesis in plants and microorganisms (Westfall et al., 2014). Chorismate mutase catalyzes the rearrangement of chorismate to prephenate, the precursor of phenylalanine and tyrosine biosynthesis (Figure 3). Conversely, anthranilate synthase competes for chorismate in the plastid to generate tryptophan.

Chorismate mutase in plants

The enzyme that converts chorismate into the precursor molecule in tyrosine and phenylalanine biosynthesis, prephenate, is chorismate mutase (Westfall et al., 2014). This enzyme catalyzes a pericyclic Claisen rearrangement of chorismate in the first committed step of tyrosine and phenylalanine biosynthesis (Sträter et al., 1997). Chorismate mutases have been identified in plants, fungi, bacteria, and nematodes with varying sequence and three-dimensional structures. Microbial chorismate mutases are either dimers or monomers depending on the

![Figure 3. Chorismate mutase generates prephenate, the precursor to tyrosine and phenylalanine.](image)

Chorismate mutase catalyses the pericyclic Claisen rearrangement of chorismate to prephenate. Taken from Kroll and Holland et al., 2017.
presence of a regulatory domain (Xue et al., 1994). Amino acid effectors bind in this regulatory region, or effector site, to either activate or inhibit the enzyme.

In the model plant Arabidopsis, there are two plastidial chorismate mutases (AtCM1 and AtCM3) and one cytosolic isoform (AtCM2) (Eberhard et al., 1993; Eberhard et al., 1996; Mobley et al., 1999). Functional analysis of these enzymes has shown that AtCM1 is activated by tryptophan and inhibited by phenylalanine and tyrosine, while AtCM3 uses the same positive and negative effectors but is also activated by cysteine and histidine; however, AtCM2 is a nonallosteric isoform (Westfall et al., 2014).

The AtCM1 x-ray crystal structure has been solved in complex with the inhibitors tyrosine and phenylalanine (Westfall et al., 2014) (PDB: 4PPU; 4PPV). The enzyme is a dimer with the effector site at the dimer interface with each chain contributing residues to the sides of the effector site. Functional analysis in combination with sequence comparisons of the three chorismate mutase isoforms identified key residues involved in allosteric regulation by amino acids.

Structure-guided phylogenetic analysis of the diversity of chorismate mutases across the green lineage revealed that the clade containing AtCM3 was specific to Brassicaceae, suggesting a specialized role for these proteins in this plant family (Westfall et al., 2014) (Figure 4). Homologs of AtCM2 were found in all plants and are likely to not be allosterically regulated and to localize to the plastid. Some early plants, including the bryophyte Physcomitrella patens, the lycophyte Selaginella moellendorfii, and the early flowering plant Amborella trichopodo, do not have homologs of AtCM2 (Westfall et al., 2014). The clade containing AtCM1 contains a putative N-terminal chloroplast signal peptide and likely all would have the same effectors based on sequence comparisons.
**Figure 4. Diversity of chorismate mutases across plants.**

Colors depict both phylogeny and the presence of key glycine residues. Red sequences include chlorophytes and yeast from the clade. Purple sequences are part of the AtCM2-like clade, and dark green sequences include CMs from mosses, ferns, monocots, and Amborella. The AtCM1-like clade is light green, while the AtCM3-like clade is blue. Taken from Westfall et al., 2014.

**Tyrosine and phenylalanine biosynthesis in plants**

$L$-Phenylalanine and $L$-tyrosine are both downstream of prephenate, the product of the chorismate mutase reaction (Westfall et al., 2014). In the prephenate pathway, prephenate is decarboxylated to form either 4-hydroxyphenylpyruvate (HPP) or phenylpyruvate (PPY) by either prephenate dehydratase (PDT) or prephenate dehydrogenase (PDH), respectively.
Subsequently, an aminotransferase converts the keto acid to an amino acid, forming tyrosine from HPP with HPP-aminotransferase (HPP-AT) and phenylalanine from PPY using PPY-aminotransferase (PPY-AT).

Alternatively, plants can synthesize phenylalanine and tyrosine using the arogenate pathway (Eberhard et al., 1996; Mobley et al., 1999; Colquhoun et al., 2010). Prephenate is aminated to form arogenate by the enzyme prephenate aminotransferase (PAT). Arogenate is the direct precursor of both phenylalanine and tyrosine and can be decarboxylated by either an arogenate dehydrogenase (ADH) or an arogenate dehydratase (ADT) to form tyrosine or phenylalanine, respectively (Rippert and Matringe, 2002; Cho et al., 2007; Yamada et al., 2008; Rippert et al., 2009; Maeda et al., 2010). Recently, a plastidial cationic amino acid transporter in plants was identified that exports phenylalanine and tyrosine to the cytosol (Widhalm et al., 2015a).

**Prephenate and Arogenate Dehydrogenases**

The two dehydrogenases involved in tyrosine biosynthesis, ADH and PDH, catalyze the oxidative decarboxylation of arogenate and prephenate to tyrosine and 4-HPP, respectively (Figure 1) (Maeda and Dudareva, 2012). Most fungi and bacteria contain PDHs, while legumes (Leguminosae) are the only land plants found to contain functional PDHs (Stenmark et al., 1974; Fazel and Jensen, 1979; Schenck et al., 2015). PDHs exist as either monofunctional or bifunctional enzymes, in which PDH is fused to chorismate mutase, EPSPS, or an aminotransferase (Song et al., 2005).
Figure 5. Prephenate dehydrogenase reaction.
Prephenate dehydrogenase is a cytosolic, tyrosine-insensitive enzyme in legume that catalyzes the oxidative decarboxylation of prephenate using NADP\(^+\) to generate 4-hydroxyphenylpyruvate, CO\(_2\), and NADPH on route to tyrosine.

Aromatic amino acid biosynthesis has long been accepted to occur exclusively in plastids where it is subject to feedback regulation; however, it was recently demonstrated that 4-HPP is formed in the cytosol through the decarboxylation of prephenate by PDH (Figure 5). Traditionally, PDHs are allosterically inhibited by tyrosine (Sun et al., 2009). Tyrosine binds an allosteric site in these microbial enzymes and is antagonistic, and the binding of two tyrosines inactivates the enzyme (Turnbull et al., 1991). Surprisingly, plant PDHs have been found to be insensitive to regulation by Tyr, unlike its bacterial homologs, and have only been discovered in leguminous plants (Gamborg and Keeley, 1966; Rubin and Jensen, 1979; Schenck et al., 2015).

The biochemical function of PDH may also be useful as an herbicide tolerance trait. Increased expression of PDH could potentially allow plants to resist 4-HPP dioxygenase (HPPD) herbicides, which include the triketone class of herbicides such as sulcotrione, mesotrione, fluorochloridone, and isoxaflutole (Schenck and Maeda, 2014). Due to the insensitivity of PDH to product feedback inhibition, unlike their microbial homologs, this enzyme may be used to increase production of downstream tyrosine- and 4-HPP-derived natural products when introduced to non-legume plants. Additionally, plants expressing PDH may exhibit resistance to
herbicides by increased 4-HPP and plastoquinone synthesis. Plastoquinone, a 4-HPP derivative, is an electron carrier in photosynthesis and a cofactor in carotenoid biosynthesis. Inhibiting plastoquinone biosynthesis causes photobleaching and lethal phenotypes, which has led to the development of herbicides that target HPPD. By overexpressing PDH, additional feedback-insensitive 4-HPP is produced, which could lead to resistance to HPPD-targeted herbicides.

In plants found to contain the PDH pathway, which includes the legumes *Glycine max* (soybean) and *Medicago truncatula* (barrelclover), HPP can be directly synthesized using one enzyme instead of three that would be required using the arogenate pathway - PAT, ADH, and HPP-AT (Schenck and Maeda, 2014). Because PDH is cytosolic, the enzyme evades competition with PAT and PDT for prephenate in the plastid.

Unlike PDH, ADH activity has been detected in corn, sorghum, mung bean, soybean, alfalfa, and Arabidopsis and is thought to be ubiquitous across Viridiplantae (Rubin and Jensen, 1979; Byng et al., 1981; Connelly and Conn, 1986; Rippert and Matringe, 2002; Holding et al., 2010; Schenck et al., 2015). Arabidopsis ADHs are plastid-localized (Rippert et al., 2009). Characterization of Arabidopsis ADHs revealed that ADH1 is specific for arogenate, while ADH2 can accept prephenate as well, although at three orders of magnitude lower catalytic efficiency than arogenate (Rippert and Matringe, 2002).

Similar to PDH, ADH catalyzes an oxidative decarboxylation reaction using NAD$^+$ or NADP$^+$ as a cofactor (Legrand et al., 2006). This enzyme uses arogenate as a substrate to generate tyrosine as the last step in the prephenate pathway for tyrosine biosynthesis in the chloroplast. Prior to my thesis research, no structural information about arogenate or prephenate dehydrogenases from plants was available. The closest available structure was that of an ADH from the cyanobacterium *Synechocystis*, which varies from the plant sequences (Legrand et al.,
2006) (PDB: 2F1K). Additionally, the evolutionary basis of the evolution of PDH function in legumes, as well as the evolution of tyrosine-insensitivity, was unknown.

**Prephenate aminotransferase**

There are three PLP-dependent aminotransferases involved in tyrosine and phenylalanine biosynthesis (Figure 2). Aminotransferases in aromatic amino acid biosynthesis pathways use either L-glutamate or L-aspartate as amino donors and a keto acid, such as prephenate, 4-HPP, or phenylpyruvate (Figure 6) (Bonner and Jensen, 1985; Siehl et al., 1986; Maeda et al., 2011). Prephenate aminotransferase (PAT) is a key enzyme in aromatic amino acid biosynthesis because its product, arogenate, is the immediate precursor for both tyrosine and phenylalanine. Interestingly, the enzyme has a dual role in metabolism and can also function as a classical aspartate aminotransferase (Bonner and Jensen, 1985; Graindorge et al., 2010; Maeda et al., 2011). This enzyme was only recently discovered as the last missing enzyme in aromatic amino acid biosynthesis in plants (Graindorge et al., 2010). PAT activity has been detected in bacteria (Stenmark et al., 1974; Fazel and Jensen, 1979) and in various plants, including tobacco, Arabidopsis, petunia, tomato, mung bean, and sorghum (Rubin and Jensen, 1979; Bonner and Jensen, 1985; Siehl et al., 1986; Graindorge et al., 2010; Dal Cin et al., 2011; Maeda et al., 2011), and is considered to be ubiquitous in plants. PAT belongs to the class Ib aspartate aminotransferases that use a catalytic lysine for the amine group transfer from an amino acid, aspartate, to a keto acid, prephenate or α-ketoglutarate (Nobe et al., 1998; de la Torre et al., 2009).

PAT was identified in an embryo-lethal mutant screen, and as there is only one copy in the model plant Arabidopsis, it is expected that this enzyme is essential for plant growth and
development (Pagnussat et al., 2005). Therefore, the only way to study PAT activity in planta is through knock-down experiments. RNAi suppression of PAT in petunia petals did not affect aspartate aminotransferase activity, suggesting that PAT does not significantly contribute to total aspartate aminotransferase activity in plants but is contributing to aromatic amino acids (Maeda et al., 2011). A similar experiment used virus-induced gene silencing (VIGS) of PAT in tobacco and found that the plants that were significantly reduced in stature and had decreased lignin content, reduced chlorophyll content, and less free asparagine (de la Torre et al., 2014). In agreement with the petunia RNAi experiment, aspartate and glutamate levels remained unchanged in the tobacco VIGS lines compared to wild-type (de la Torre et al., 2014).

![Figure 6. Prephenate aminotransferase (PAT) reaction.](image)

PAT catalyzes the PLP-dependent transfer of an amine from aspartate (Asp) to the keto acid prephenate to generate a new amino acid, arogenate, and a new keto acid, oxaloacetate (OAA). This reaction occurs in the plastid of all plants upstream of phenylalanine and tyrosine.

A phylobiochemical analysis of prephenate aminotransferases in plants and microorganisms revealed that the plant enzyme likely evolved as a lateral gene transfer from Chlorobi/Bacteroidetes (Dornfeld et al., 2014). This study also identified two residues that were conserved in functional PATs but variable in aspartate aminotransferases—Thr84 and Val169. Mutations in single residues yielded reduced activity as a prephenate aminotransferase enzyme; however, the double mutant was a gain-of-function mutant in that it preferred 4-hydroxyphenylypyruvate (4-HPP) as a keto acid substrate and tryptophan as the amino donor and
thus was a functional 4-HPP-aminotransferase. Despite this interesting finding, no structural data was available for a plant PAT or aspartate aminotransferase to determine where these residues are in relation to the active site. Additionally, this is the only biochemical investigation of plant PAT, and no effectors were screened or identified for this enzyme.

Introduction to plant hormones

To grow, develop, and respond to the environment, plants use specialized metabolites known as phytohormones (Santer et al, 2009). These hormones vary in chemical structure and in biological function and are generated from primary metabolites. While some hormones are active in their unmodified form, such as the growth hormone indole-3-acetic acid (IAA), hormones can be modified by glycosylation, methylation, hydroxylation, sulfonation, carboxylation, or amino acid conjugation to generate either the active form of the hormone or a storage or degradation form (Silverman et al., 1995; Bajguz and Piotrowska, 2009; Piotrowska and Bajguz, 2011; Koo and Howe, 2012; Korasick et al., 2013). Our understanding of the function of the various possible forms of plant hormones has only recently begun to be understood. To date, there are nine classes of recognized plant hormones— auxins, jasmonates, benzoates, ethylene, cytokinins, brassinosteroids, abscisic acid, gibberellins, and strigalactones—that regulate every aspect of plant development and responses to biotic and abiotic stressors from germination to senescence (Figure 7).

Auxins: Indole-3-acetic acid and phenylacetic acid

Auxins are plant phytohormones involved in almost every aspects of plant growth and development, including cell division and proliferation, gravitropism, phototropism, and
Figure 7. Overview of plant hormones and hormone homeostasis.
A. Representative chemical structures of the nine classes of plant hormones. B. Plant hormones, including the auxin indole-3-acetic acid (IAA), can be modified to form new structures that alter the physiological response to the hormone. Active (green), inactive (orange), and anti-auxin (red) forms of IAA are shown. Taken from Westfall et al., 2013.

development and maintenance of root and shoot morphology (Woodward and Bartel, 2005). The primary active form of auxin, indole-3-acetic acid (IAA), is synthesized from the aromatic amino acid L-tryptophan or from aromatic amino acid biosynthetic precursors like anthranilate. Active forms of auxins also include phenylacetic acid (PAA) and 4-chloroindole acetic acid, and IAA can be found in methylated, glycosylated, and amino acid conjugated forms in plants (Korasick et al., 2013). In Arabidopsis thaliana, about 90% of IAA is found in an amide linkage, with another 10% in an ester-linked conjugate; only about 1% of IAA is found in a free, unconjugated form (Normanly et al., 1993; Tam et al., 2000).

To balance the levels of IAA within a cell requires regulation of synthesis, transport, storage, and catabolism (Korasick et al., 2013). Modifications to the chemical structure of IAA regulate the concentration of the active hormone within the plant cell at a given time. Methylation, glycosylation, or conjugation with alanine or leucine convert free IAA into a storage form for
later use; however, conjugation with aspartate and glutamate lead to the degradation (Staswick et al., 2005; Korasick et al., 2013). Interestingly, IAA-tryptophan conjugates act as an anti-auxin that inhibits physiological responses typical of the plant growth hormone (Staswick, 2009).

While other IAA conjugates exist in Arabidopsis, their role in regulating auxin levels remains unknown.

Figure 8. Comparison of the IAA and PAA biosynthetic from tryptophan and phenylalanine, respectively.
Catalytic enzymes are AtTAA1 (Trp aminotransferase) and PsTAR1 (Trp aminotransferase related), YUC (flavin mono-oxygenase), and PhPPY-AT (phenylpyruvate aminotransferase) and PsArAT (aromatic aminotransferase). The dashed line represents an as yet unknown catalytic mechanism. Adapted from Cook et al., 2016.

A second form of auxin, phenylacetic acid (PAA), has received considerably less attention in the auxin field (Cook et al., 2016). PAA is an active auxin in plants and is present at similar physiological levels as IAA in tomatoes, tobacco, sunflowers, peas, barley, and maize (Wightman, 1977; Wightman and Lighty, 1982; Schneider et al., 1985; Ludwig-Muller and Cohen, 2002; Schneider and Wightman, 1986; Sugawara et al., 2015). Recently, PAA-tryptophan conjugates were found at 17-fold higher levels than IAA-tryptophan in Arabidopsis seedlings suggesting it could play a role in plant physiology (Staswick et al., 2017).

PAA is derived from phenylpyruvate, an intermediate in phenylalanine biosynthesis, using enzymes that are distinct from those used in IAA biosynthesis (Cook et al., 2016) (Figure 8).
The biosynthesis, inactivation, transport, and role in signal transduction of PAA remains unknown (Korasick et al., 2013). Unlike IAA, PAA is not actively and directionally transported, but it does, however, form concentration gradients (Sugawara et al., 2015).

**Benzoates**

Benzoates are plant phytohormones that are derived from the aromatic amino acid precursor chorismate (Widhalm and Dudareva, 2015b). Acetylsalicylic acid, or aspirin, reduces pain, inflammation, and fever in humans and can also lower the risk of heart attack, stroke, and cancers if taken regularly (Vlot et al., 2009); however, plants produce the benzoate salicylic acid (SA) as a phenolic defense hormone to regulate a number of physiological responses including disease resistance, flowering, thermogenesis, cell growth, stomatal aperture, seed germination, senescence-related gene expression, and responses to biotrophic pathogens (Yalpani et al., 1991; An and Mou, 2011; Boatwright and Pajerowska-Mukhtar, 2013).

SA can be synthesized in plants by two routes, both of which use chorismate as a precursor (Vlot et al., 2009). Phenylalanine can be converted into SA by benzoate intermediates or coumaric acid, or chorismate can be converted into SA via isochorismate using isochorismate synthase and a predicted, yet unidentified, isochorismate pyruvate lyase (Strawn et al., 2007; Verberne et al., 2000; Wildermuth et al., 2001). There are two *iscochorismate synthase* genes in Arabidopsis, but the *ics1/ics2* double mutant line still contains SA, suggesting that both SA biosynthetic pathways are functional in plants (Garcion et al., 2008).

SA can be modified by SA glucosyltransferase (SAGT) to form either SA O-β-glucoside (SAG) or salicyloyl glucose ester (Dean and Mills, 2004; Dean et al., 2005; Dean et al., 2003; Lee and Raskin, 1998; Lee and Raskin, 1999; and Song 2006). While SA biosynthesis is
predicted to occur in the chloroplast (Wildermuth et al., 2001; Garcion et al., 2005; Strawn et al., 2007), SAG can be actively transported to the vacuole where it presumably functions as a storage form of SA (Dean and Mills, 2004; Dean et al., 2005; Hennig et al., 1993). Methylated SA (MeSA), hydroxylated SA, and aspartate-conjugated SA forms also exist in plants (Vlot et al., 2009; Westfall and Sherp et al., 2016). The physiological significance of SA-aspartate and other presumed SA-amino acid conjugates remains unknown.

Another benzoate that is found in amino acid conjugates in plants is benzoic acid (BA) (Westfall and Sherp et al., 2016). While BA itself is not a known plant hormone, it is a precursor to phenylalanine-derived SA in plants. A BA-2-hydroxylase soluble P450 monooxygenase converts BA to SA (Widhalm and Dudareva, 2015b; Leon et al., 1995). GH3.5, a hormone regulating protein that conjugates amino acids to auxins, is responsible for forming BA-aspartate conjugates that can be detected in planta, and the catalytic efficiencies of the enzyme using IAA and BA are about the same. However, the physiological role of these BA-aspartate conjugates remains to be fully understood. It is speculated that these conjugates are a SA precursor pool so the plant can generate SA quickly in response to pathogen attack (Westfall and Sherp et al., 2016).

**Jasmonates**

To coordinate growth and development during times of stress and predation, plants evolved the capability of synthesizing and responding to an array of complex modified oxylipins, or oxygenated fatty acid derivates, known as jasmonates (Kombrink, 2012). Methyl jasmonate (MeJA) was first isolated from the oil of jasmine (*Jasminum grandiflorum*), although its effect on plant physiology remained elusive for nearly two decades (Demole et al., 1962; Ueda and
Kato, 1980; Dathe, 1981). Since this discovery, jasmonates have been recognized as critical phytohormones involved in plant defense and fertility.

Chemically, the jasmonates encompass jasmonic acid (JA) and its derivatives, including MeJA; 3R, 7S-jasmonyl-L-isoleucine (JA-Ile); cis-jasmon; and jasmonyl-1-aminocyclopropane-1-carboxylic acid (Browse, 2009). Jasmonate biosynthesis proceeds through the lipoxygenase-catalyzed peroxidation of linoleic acid to form 13-hydroperoxide, which is modified and cyclized to form 12-oxo-phytodienoic acid (OPDA). OPDA is reduced and then undergoes three rounds of β-oxidation to form JA. Further biochemical modifications of JA, such as glycosylation, decarboxylation, reduction of either the C6 carbonyl or C9,10 double bond, or hydroxylation of either C11 or C12, can convert the core JA scaffold into over 30 different jasmonate compounds with diverse physiological roles.

Investigations of the role of jasmonates in plant physiology reveal their role in stamen and trichome development, vegetative growth, cell cycle regulation, senescence, regulation of anthocyanin pigment biosynthesis, and responses to various biotic and abiotic stresses, most notably response to herbivory (Creelman and Mullet, 1997; Wasternack, 2007; Howe and Jander, 2008; Browse, 2009; Avanci et al., 2010; Pauwels and Goossens, 2011). Jasmonates are also key players in induced resistance of plants (Kunkel and Brooks, 2002), including Rhizobacteria-mediated induced systemic resistance (Pieterse et al., 1998).

As one of the major plant defense compounds, jasmonates induce defense mechanisms to ward off insects, herbivores, and a broad spectrum of fungal and bacterial pathogens (Howe and Jander, 2008; Browse, 2009). Wounding of the plant tissue leads to elevated levels of jasmonates at the site of attack, which in turn induces a systemic increase of JAs through the regulation of JA-responsive genes that allow for production of diverse compounds as the attack
continues (Reymond et al., 2004; Glauser et al., 2008). These compounds include glucosinolates, camalexins, alkaloids, artemisinin, and volatile organic compounds, which provide a chemical arsenal to combat predators (Bolter, 1993; Paré and Tumlinson, 1999; Engelberth et al., 2004; De Geyter et al., 2012).

When stressed or otherwise stimulated, various jasmonates accumulate and are converted to JA-Ile, the only known bioactive jasmonate hormone (Figure 9) (Wasternack and Hause, 2007; Fonseca et al., 2009a; Fonseca et al., 2009b). JA-Ile forms a co-receptor complex with the F-box protein CORONATINE INSENSITIVE 1(COI1) and a transcriptional repressor JASMONATE-ZIM DOMAIN PROTEIN (JAZ). This complex allows for ubiquitylation of the JAZ protein, which results in its degradation by the 26S proteasome (Fonseca et al., 2009a; Thines et al., 2007; Chini et al., 2007). Degradation of JAZ proteins thereby alters transcriptional patterns in response to JA-mediated signals.

Although JA-Ile is the active form of the hormone, conjugated forms with other amino acids, including alanine, leucine, valine, phenylalanine, tyrosine, and tryptophan, also occur in plants and function in defense, eliciting phytoalexin biosynthesis, and agravitropic root growth in seedlings (Piotrowska and Bajguz, 2011; Staswick et al., 2017). Recently it was shown that JA-tryptophan likely inhibits IAA transport by negatively regulating the AUX1 cellular auxin importer (Staswick et al., 2017). Glycosylated JA is predicted to be a storage or transport form of the hormone.

**GH3 proteins regulate the pre-receptor concentration of plant hormones**

In plants, the GH3 (Gretchen Hagen 3) acyl acid-amido synthetases contribute to maintaining hormone levels in plants by conjugating an acyl acid hormone, including auxins, jasmonates, and
benzoates, to an amino acid as either a storage or activation mechanism (Hagen and Guilfoyle, 1985; Westfall et al., 2013). The biological function of this class of enzymes was first established when the jasmonate resistant 1 (jar1-1) mutant was identified in Arabidopsis thaliana and shown to exhibit decreased sensitivity to exogenous MeJA and reduced male fertility (Staswick et al., 1992, 2002). The A. thaliana JAR1 protein (also known as AtGH3.11) functions as a JA-amido synthetase and conjugates isoleucine to JA (Staswick and Tiryaki, 2004).

Figure 9. Pre-receptor modulation of plant hormones occurs by GH3 acyl acid amido synthetases.
JA in its free form is an inactive molecule, while the isoleucine conjugated form is active. This active molecule goes on to form the active hormone receptor complex with the COI receptor and JAZ proteins to lead to downstream physiological responses. In the case of the auxin IAA, the conjugates are inactive and no longer bind to the TIR receptor. These conjugates are generally storage or degradation forms of the hormone.

GH3 proteins functions as a monomer to catalyze the ATP-dependent conjugation of an acyl acid and an amino acid (Chen et al., 2010). While some conjugates are “active” and can regulate responsive genes through binding the receptor protein, such as JA-Ile (Feys et al., 1994; Xie et al., 1998), other conjugates are “inactive” storage or degradation forms of the hormone, such as
IAA-Asp (LeClere et al., 2002). GH3 proteins are part of the ANL superfamily (Acyl-CoA synthetase, Non-ribosomal peptide synthetase adenylating domain, and Luciferase) and carry out an ATP-dependent adenylation reaction (Gulick, 2009).

This family of proteins generates conjugates through a two-step ping-pong mechanism that involves both an adenylation and transfer reaction (Staswick et al., 2002; Staswick et al., 2005; Terol et al., 2006; Chen et al., 2010; Okrent et al., 2011; Westfall et al., 2010) (Figure 10). In the first half-reaction, the acyl acid carboxylate is activated by adenylation using ATP, and pyrophosphate is released (Chen et al., 2010; Westfall et al., 2010); this is followed by a transfer reaction where the nucleophilic amine group of an amino acid displaces AMP to form an acyl acid-amino acid conjugate.

![Figure 10. GH3 protein reaction mechanism.](image)

**Figure 10. GH3 protein reaction mechanism.**
GH3 proteins form acyl acid-amino acid (Acyl-AA) conjugates through a two-step mechanism; in the first half-reaction ($k_1$), an adenylated acyl acid intermediate (Acyl-AMP) forms and pyrophosphate ($PP_i$) is released. The amino acid (AA) comes in during the second half-reaction to form the final conjugate, and adenosyl-monophosphate (AMP) is released.

**Structure/function studies of GH3 proteins**

Several crystal structures of Arabidopsis and grapevine (*Vitus vinifera*) have provided insight on the sequence-structure-function relationship among the GH3 family of proteins (Westfall and Sherp et al., 2016; Westfall and Zubieta et al., 2012; Chen et al., 2017; Round et al., 2013; Peat et al., 2012). GH3 proteins share a common tertiary structure containing two domains—a large N-terminal domain and a smaller C-terminal domain characteristic of the adenylating firefly luciferase (ANL) enzyme superfamily—with the active site at the interface of the two domains.
A flexible hinge loop connects the two domains and pivots the C-terminal domain during the reaction. In the open conformation, ATP, Mg\(^{2+}\), and the acyl acid bind for the adenylation reaction to occur, and a solvent-accessible channel allows for pyrophosphate release. Upon adenylation and pyrophosphate release, the C-terminal domain rotates about the hinge loop and closes the active site, repositioning residues for the transferase reaction to occur (Figure 11).

**Figure 11. Three-dimensional structure of a GH3 protein.**
GH3 proteins contain a larger N-terminal domain connected to a smaller C-terminal domain through a flexible hinge loop. The active site forms at the interface of the two domains. During the reaction, ATP binds when the protein is in the open conformation. When pyrophosphate is released and the adenylated intermediate is formed, the protein ratchets to form a closed conformation, where the hinge loop region covers the active site for the full reaction to proceed. When the conjugate is released, the protein returns to the open conformation. Taken from Westfall and Zubieta et al., 2012.

From Arabidopsis, structures of GH3.5, GH3.11, and GH3.12 have been determined, and the substrates of uncharacterized GH3 proteins can be predicted based on analysis of the active site.
residues (Westfall and Zubieta et al., 2012). In addition to these structures, small-angle x-ray scattering analysis show that the C-terminal domain of GH3 proteins toggles between the open (ATP-bound) and closed (AMP-bound) conformations during the two half-reactions, followed by biochemical studies that investigated the roles of the active site residues in each of the two half-reactions (Westfall and Zubieta et al., 2012; Round et al., 2013).

Figure 12. Amino acid sequence alignment of the hormone-binding pocket sequences of GH3 proteins from Arabidopsis thaliana (AtGH3), rice (OsGH3), moss (PpGH3), lycophyte (SmGH3), soybean (GmGH3), and maize (ZmGH3). These alignments cluster the GH3 protein based on their presumed function: box 1 represents predicted jasmonate-using GH3 proteins; box 2 represents auxin-using GH3 proteins; box 4 represents benzoate-using GH3 proteins; box 6 represents indole-butyric acid using GH3 proteins; and boxes 3, 5, 7, and 8 represent GH3 proteins with unidentified substrates. Taken from Westfall and Zubieta et al., 2012.
Genome and transcriptome data has identified the presence of 19 putative GH3 genes in the model plant Arabidopsis thaliana (thale cress) (Hagen and Guilfoyle, 2002). GH3 proteins have been identified in all plants studied so far, including the lycophyte Selaginella moellendorffii and the bryophyte Physcomitrella patens (Figure 12). Once a three-dimensional structure of a GH3 protein was available, the active sites could be identified and compared based on amino acid sequence similarity. This led to the reclassification of the GH3 proteins based on the sequences found on α-helices 5 and 6, residues 217 and 239 and β-sheets 8 and 9 (Westfall and Zubieta et al., 2012). The largest class of GH3 proteins are auxin using (Figure 12, box 2), followed by the jasmonate-using GH3 proteins.

Interestingly, some classes of GH3 proteins seem to be specific to Brassica species, including classes 4-7 (Figure 12). AtGH3.15 was recently found to be specific for an IAA biosynthetic intermediate, indole-3-butyric acid, as its acyl acid substrate (Sherp et al., 2018). Although class 3 is not exclusive to Brassicaceae, the only member found in this cluster are from either a Brassica species (Arabidopsis thaliana) or from soybean (Glycine max), suggesting that this class of GH3s is potentially not widespread in plants. The acyl acid substrate for class 3 remains to be identified.

Conclusions

As sessile organisms, plant have been constantly evolving to ward off stressors, both biotic and abiotic. As the environments around plants have evolved, they have developed new and ever improving mechanisms to grow to compete with other plants for physical space, to attract pollinators and seed dispersers, to defend themselves against pathogens and predators, and to
protect themselves against Mother Nature. Part of the secret to plants success on Earth stems from the chemical arsenal they have built and refined over the past 425 million years.

Plants have developed the molecular machinery in the form of enzymes necessary to decorate and modify primary metabolites, such as amino acids, to give them new functions. In the case of the three aromatic amino acids—tyrosine, phenylalanine, and tryptophan—these molecules have been repeatedly introduced into the plant chemical library in vastly different forms. For example, tryptophan is the precursor to both the growth promoting hormone indole-3-acetic acid (IAA) as well as the indole glucosinolate defense metabolites (Bolter, 1993). Also, phenylalanine is a precursor to the defense hormone salicylic acid as well as the structural polymer lignin. Over time, plants have woven an intricate web that delicately balances growth and everyday plant physiology with the ability to defend itself using a similar scaffold.

Despite the importance of aromatic amino acids as precursors to important plant chemicals and polymers and as components of proteins, little is known about the evolution of their biosynthetic pathways and their regulation in plants. While the microbial pathways were studied extensively in the 1980s and 1990s, much of what we know about aromatic amino acid biosynthesis in plants is based on the notion that the pathways are likely the same as those of microorganisms. The only three-dimensional structure from an aromatic amino acid biosynthetic enzyme that existed before my thesis research began was that of chorismate mutase from Arabidopsis thaliana (Westfall et al., 2013). Structure-guided functional analysis with an eye towards phylogeny would greatly inform our understanding of the evolution of these essential plant enzymes.

Of similar importance is our understanding of how plants regulate the growth-defense tradeoff by modulating the concentrations of growth and defense hormones. The expansive GH3
family of proteins affords plants the ability to add an amino acid to a hormone to change its function. Interestingly, some of these proteins, like AtGH3.5, accept acyl acid substrates that are both defense hormones, like SA, and growth hormones, like IAA. In comparing the active sites residues of IAA-specific GH3s to those of AtGH3.5, there are no real differences that guide our understanding of how essentially the same active can recognize and adenylate two very different acyl acids.

In general, our understanding of the substrate-specificities of the GH3 proteins remains to be fully explored. Studying role of Brassicaceae-specific classes of GH3 proteins could lead to potential new insights on how this family of plants has evolved separate regulatory mechanism for controlling hormones or plant metabolites.
REFERENCES


Cook SD, Nichols DS, Smith J, Chourey PS, McAdams EL, Quittenden L, Ross JJ (2016) Auxin biosynthesis: are the indole-3-acetic acid and phenylacetic acid biosynthesis pathways mirror images? *Plant Physiol.* 171(2), 1230-1241


Feys BJF, Benedetti CE, Penfold CN, Turner JG (1994) Arabidopsis mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. *Plant Cell* 6, 751-759


Schenck CA, Maeda HA (2014) Prephenate dehydrogenases and arogenate dehydrogenases that are insensitive to tyrosine inhibition and methods of using the same. US. Patent P130167US02, USSN: 14/548216


Staswick PE, Yuen GY, Lehman CC (1998) Jasmonate signaling mutants of Arabidopsis are susceptible to the soil fungus *Pythium irregular*. *Plant J.* 15, 747-54

Staswick PE, Tiryaki I, Rowe ML (2002) Jasmonate response locus JAR1 and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. *Plant Cell* 14, 1405-15


Chapter 2

Molecular Basis of the Evolution of Alternative Tyrosine Biosynthetic Pathways in Plants


ABSTRACT

L-Tyrosine (Tyr) is essential for protein synthesis and a precursor of numerous specialized metabolites crucial for plant and human health. Tyr can be synthesized via two alternative routes by a key regulatory TyrA family enzyme, prephenate or arogenate dehydrogenase (PDH/TyrA) or ADH/TyrA), representing a unique divergence of primary metabolic pathways. The molecular foundation underlying the evolution of the alternative Tyr pathways is currently unknown. Here we characterized recently-diverged plant PDH and ADHs, obtained the x-ray crystal structure of soybean PDH, and identified a single amino acid residue that defines TyrA substrate specificity and regulation. Structures of mutated PDHs co-crystallized with Tyr indicate that substitutions of Asn222 confers ADH activity and Tyr-sensitivity. Subsequent mutagenesis of the corresponding residue in divergent plant ADHs introduced PDH activity and relaxed Tyr sensitivity, highlighting the critical role of this residue in TyrA substrate specificity that underlies the evolution of alternative Tyr biosynthetic pathways in plants.
INTRODUCTION

Unlike recently-evolved and lineage-specific diverse specialized (secondary) metabolic pathways (Moghe and Last, 2015; Weng et al., 2012), primary metabolism such as amino acid biosynthesis are ubiquitous and usually conserved among organisms. There are some exceptions to this notion (Gowik and Westhoff, 2011; Torruella et al., 2009), and L-tyrosine (Tyr) biosynthetic pathway is one example in which variations have long been described in microbes and plants (Jensen and Pierson, 1975; Schenck et al., 2015). Elucidation of evolutionary diversification of primary metabolism not only addresses the extent of metabolic plasticity but also provides useful engineering tools to modify core metabolic pathways.

Tyr is an essential aromatic amino acid required for protein synthesis in all organisms but, synthesized de novo only in plants and microorganisms (Maeda and Dudareva, 2012; Tzin and Galili, 2010). The neurotransmitter catecholamines in metazoans are derived from Tyr, which must be obtained from their diet, as they cannot synthesize Tyr de novo (Fernstrom and Fernstrom, 2007). In plants, Tyr and a Tyr pathway product, 4-hydroxyphenylpyruvate (HPP), serve as precursors to numerous specialized metabolites crucial for both plant and human health, such as antioxidants vitamin E, the photosynthetic electron carrier plastoquinone, betalain pigments, and defense compounds, including dhurrin, rosmarinic acid, and isoquinoline alkaloids (e.g. morphine) (Millner and Barber, 1984; Strack et al., 2003; Hunter and Cahoon, 2007; Hagel and Facchini, 2013; Gleadow and Møller, 2014). The major plant cell wall component lignin can also be synthesized from Tyr in grasses (Barros et al., 2016).

Tyr is synthesized from prephenate downstream of the shikimate pathway, by two reactions, an oxidative decarboxylation and a transamination (Fig. 1a). The TyrA enzymes catalyze the oxidative decarboxylation step and are the key regulatory enzymes of Tyr biosynthesis, as they
are usually inhibited by Tyr in a competitive manner and compete for substrates that are also used in L-phenylalanine biosynthesis (Fig. 1a) (Hudson et al., 1984; Fischer and Jensen, 1987; Bonner et al., 2008). In many microbes an NAD(H)-dependent prephenate dehydrogenase/TyrA (PDH/TyrAp; EC 1.3.1.13) converts prephenate into HPP followed by transamination to Tyr by Tyr aminotransferase (TAT/TyrB, Fig. 1a) (Kuramitsu et al., 1985; Fischer and Jensen, 1987). In plants, these two reactions occur in the reverse order, with prephenate first being transaminated to arogenate by prephenate aminotransferase (PPA-AT) (Connelly and Conn, 1986; Graindorge et al., 2010; Dal Cin et al., 2011; Maeda et al., 2011; Dornfeld et al., 2014), followed by oxidative decarboxylation to Tyr by an NADP(H)-dependent arogenate dehydrogenase/TyrA (ADH/TyrAa; EC 1.3.1.78, Fig. 1a) (Gaines et al., 1982; Rippert and Matringe, 2002). Some exceptions to these “textbook” models are found in nature including microbes that use ADH to synthesize Tyr (Keller et al., 1985; Mayer et al., 1985) and plants such as legumes having PDH activity (Gamborg et al., 1966; Rubin and Jensen, 1979; Schenck et al., 2015). Also, some microbial TyrAs prefer NADP(H) cofactor (Fischer and Jensen, 1987; Legrand et al., 2006). Thus, variations exist in the TyrA enzymes in diverse organisms, yet the molecular basis underlying TyrA substrate specificity and the alternative Tyr pathways is currently unknown.

Comparison of microbial TyrA sequences identified an aspartate residue downstream of the NAD(P)(H) binding motif that correlated with NADP(H) cofactor specificity of TyrA (Song et al., 2005; Bonner et al., 2008). Site-directed mutagenesis of Escherichia coli PDH and structural analysis of Aquifex aeolicus PDH identified an active site histidine, which interacts with substrate C4-hydroxyl and is critical for catalysis in each PDH (Christendat et al., 1998; Sun et al., 2009). Structural analysis and mutagenesis also identified an active site arginine that is necessary for substrate binding, but not for substrate specificity (Christendat et al., 1999; Sun et
Figure 1. Tyr biosynthesis pathways in plants and identification and characterization of noncanonical ADHs.

(a) Two Tyr biosynthetic routes from prephenate are shown. The PDH (blue) pathway is present in most microbes and legumes, whereas the ADH (red) pathway is ubiquitous in plants. Dashed line represents feedback inhibition by Tyr. (b) Simplified phylogenetic analysis of TyrA homologs from various eudicot lineages identified a clade of ADH/PDH homologs (noncanonical, gray) distinct from previously characterized plant ADH (canonical). Plant PDHs form a subgroup in the noncanonical clade. Stars mark enzymes used in this study. (c) PDH (blue) and ADH (red) activity of PDH, and noncanonical ADHs with NADP+ cofactor. Catalytic efficiency ($k_{cat}/K_m$) is expressed as mM⁻¹ s⁻¹ ± SEM of n ≥ 3. Data without bars had detectable activity unless N.D. (below detection limit) is indicated (see Table 2 for values). (d) Effect of Tyr on plant ADH and PDHs. Data are shown as IC₅₀ plots with enzymatic activity determined at increasing amounts of L-Tyr (at the following concentrations: 0, 0.01, 0.1, 0.5, 1, 5, and 8 mM). Activity was normalized to an assay with no L-Tyr and expressed as percent activity of n = 3 ± SEM.
al., 2006; Sun et al., 2009). Besides their varied substrate and cofactor specificities, TyrA enzymes also exhibit different regulatory properties. Mutation of another active site histidine, which is present in the E. coli and A. aeolicus PDHs but absent in Tyr-insensitive Synechocystis ADH, relieved Tyr inhibition but simultaneously reduced PDH activity (Sun et al., 2009). Random mutagenesis of the E. coli enzyme identified additional residues that relaxed Tyr inhibition; however, PDH activity was also reduced in these mutants (Lütke-eversloh and Stephanopoulos, 2005). Sequence and structural comparisons of divergent TyrA homologs have been unable to identify specific determinants of Tyr-sensitivity and substrate specificity.

Recent work described legume PDHs that were insensitive to Tyr regulation (Schenck et al., 2015). Here, we used phylogeny-guided structure-function analyses of ADHs from legumes and eudicots that are phylogenetically related to legume PDHs and identified an active site residue that determines prephenate versus arogenate specificity and simultaneously alter Tyr inhibition in these enzymes. The structures of mutant PDH enzymes co-crystallized with Tyr reveal the molecular basis of TyrA substrate specificity and feedback-regulation that underlies the evolution of two alternative Tyr pathways in plants.

**MATERIALS AND METHODS**

**Identification of ncADH enzymes from plants**

BlastP searches were performed using the amino acid sequences of GmPDH1/G. max_18g02650 (KM507071) and MtPDH/M. truncatula_3g071980 (KM507076) as queries against various plant lineages found within the Phytozome (www.phytozome.net) (Goodstein et al., 2012), 1KP (www.onekp.com)(Matasci et al., 2014), legume information system (https://legumeinfo.org/) (Dash et al., 2016), and SustainPine DB
A phylogenetic analysis (Fig. 1b) was performed using selected homologs identified through BlastP searches. Evolutionary distances were estimated based on maximum likelihood (Whelan and Goldman, 2001). Phylogenetic analysis was performed in MEGA6 (Tamura et al., 2013) from an amino acid alignment using MUSCLE (Edgar, 2004). All positions with <75% site coverage were removed, leaving 263 positions in the final analysis from 32 sequences, the tree was estimated with 1,000 bootstrap replicates (Fig. 1b). Numbers at the branches represent the percent of replicate trees in which sequences grouped accordingly, branches with less than 50% bootstrap support are not shown. Additional phylogenetic analyses were conducted using both maximum likelihood and neighbor-joining (Saitou and Nei, 1987) methods, with all 131 TyrA homologs from the green plant lineage, including eudicots, monocots, basal angiosperms, gymnosperms, non-seed plants, and green algae, which were manually rooted onto green algae orthologs. Phylogenetic analysis was performed in MEGA6 (Tamura et al., 2013) from an amino acid alignment using MUSCLE (Edgar, 2004). All positions with <75% site coverage were removed.

**Recombinant protein expression and purification and site directed mutagenesis**

Full-length coding sequences of GmPDH1, GmncADH, MtPDH, MtncADH, SlncADH, and AiPDH/ADH, which do not contain a chloroplast transit peptide sequence, were amplified using gene-specific primers with Phusion DNA polymerase (Thermo). The PCR products were purified using QIAquick gel extraction kit (Qiagen) and ligated into pET28a vector (Novagen) at EcoRI and NdeI sites, in frame with an N-terminal 6x-His tag using In-Fusion HD cloning kit and protocol (Clontech). For site directed mutagenesis, plasmid template was diluted 100-fold,
mixed with 1 U Phusion DNA polymerase (Thermo), 0.2 mM dNTP’s, 0.5 µM forward and reverse mutagenesis primers, and 1x Phusion reaction buffer (Thermo), and then placed in a thermocycler for 98ºC for 30 s followed by 20 cycles of 10 s at 98ºC, 20 s at 70ºC , 4.5 min at 72ºC with a final extension at 72ºC for 10 min. The PCR products were purified using a QIAquick Gel Exraction Kit (Qiagen), treated with DpnI (Thermo) to digest methylated template DNA for 30 min at 37ºC, and then introduced into E. coli XL1-Blue cells. Plasmids with wild-type or site-directed mutant genes were sequenced to confirm that no errors were introduced during PCR and cloning.

For recombinant protein expression, E. coli Rosetta2 (DE3) cells (Novagen) transformed with the confirmed plasmids were culutred in 10 mL LB media supplemented with 100 µg/mL kanamycin were grown at 37ºC and 200 r.p.m. One milliliter of the overnight culture was added into 50 mL of fresh LB without antibiotics, allowed to grow at 37ºC with 200 r.p.m., 10 mL of which was further transferred to 500 mL of fresh LB with kanamycin (100 µg/mL) and grown until the OD$_{600}$ reached 0.3, when the temperature of the incubator was changed to 18ºC. After 1 hour isopropyl β-D-1-thiogalactopyranoside (IPTG, 0.4 mM final concentration) was added to induce recombinant protein expression and grown for an additional 20 hours. The cultures were spun at 10,000 x g for 10 minutes, and the pellet was washed in 100 mL of 0.9 M NaCl once and redissolved in 25 mL lysis buffer (25 mM HEPES pH 7.6, 50 mM NaCl, 10% (v/v) ethylene glycol) and 0.5 mM phenylmethylsulfonyl flouride. After two cycles of freeze-thaw, 25 mg of lysozyme (Dot Scientific) was added and cells were sonicated for 3 min. Cell debris was pelleted by centrifugation for 30 min at 50,000 x g, and supernatant was applied to a 1 mL HisTrap FF column for purification of the His-tagged recombinant protein using an ÄKTA FPLC system (GE Healthcare Bio-Sciences). After loading the supernatant the column was
washed with 20 column volumes of 90 % buffer A (0.5 M NaCl, 0.2 M sodium phosphate and 20 mM imidazole) and 10 % buffer B (0.5 M NaCl, 0.2 M sodium phosphate and 0.5 M imidazole) followed by elution with 100 % buffer B. Fractions containing purified recombinant proteins were pooled and desalted by Sephadex G50 column (GE Healthcare) size-exclusion chromatography into lysis buffer. The purity of purified proteins were analyzed by SDS-PAGE. All protein purification steps were performed at 4°C unless stated otherwise.

**GmPDH1 crystallization**

Purified protein (see above) was loaded onto a Superdex-75 26/60 HiLoad FPLC size-exclusion column (GE Healthcare) equilibrated with 25 mM Hepes, pH 7.5, and 100 mM NaCl. Protein concentration was determined by the Bradford method (Protein Assay, Bio-Rad) with bovine serum albumin as a standard. For selenomethionine (SeMet) GmPDH1 expression, *E. coli* Rosetta-2 (DE3) cells containing the PDH construct were grown to an OD$_{600}$ ~0.6 in M9 minimal media, at which point the media was supplemented with 60 mg SeMet, Val, Leu, and Ile and 100 mg of Lys, Phe, and Thr and induced with 1 mM IPTG for 16-18 hours at 16°C. SeMet GmPDH1 was purified as described for native GmPDH1.

Purified enzyme was concentrated to 10 mg ml$^{-1}$ and crystallized using the hanging-drop vapor-diffusion method with a 2-µl drop (1:1 protein and crystallization buffer). Tyr (3 mM final) was added to both GmPDH1 M219T/N222D and GmPDH1 N222D. Diffraction quality crystals of the native GmPDH1 were obtained at 4°C with a crystallization buffer of 20% PEG-4000, 30% (w/v) D-sorbitol, and 100 mM sodium citrate, pH 5.5. Crystals of SeMet PDH1 formed at 4°C with a crystallization buffer of 20% (w/v) PEG-3350, 100 mM sodium citrate, pH 4.0, and 200 mM sodium citrate tribasic. Crystals of GmPDH1 N222D formed in 2 mM of an
oxometalates solution containing 0.005 M sodium chromate tetrahydrate, 0.005 M sodium molybdate dihydrate, 0.005 M sodium tungstate dihydrate, and 0.005 M sodium orthovanadate, 0.1 M of MOPS and bis-Tris, pH 6.5, and 50% (v/v) of a precipitant mixture of 20% (w/v) PEG-8000 and 40% (v/v) 1,5-pentanediol (Gorrec, 2013). Crystals of GmPDH1 M219T/N222D formed in 16% (w/v) PEG 8000, 40 mM potassium phosphate dibasic, and 20% (v/v) glycerol. All crystals were flash-frozen in liquid nitrogen with mother liquor supplemented with 25% glycerol as a cryoprotectant.

The GmPDH1 structure was solved by single-wavelength anomalous dispersion (SAD) phasing. Diffraction data collected at beamline 19ID of the Argonne National Laboratory Advanced Photon Source were indexed, integrated, and scaled using HKL3000 (Minor et al., 2006). SHELX (Sheldrick, 2008) was used to determine initial SeMet positions and to estimate initial phases from the peak wavelength data set. SeMet positions and parameters were refined in MLPHARE (Collaborative Computational Project, 1994). Solvent flattening was performed with DM (Terwillinger, 2000), and ARP/wARP (Morris et al., 2003) was used to build an initial model. Iterative rounds of manual model building and refinement were performed with COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010), respectively. The resulting model was used for molecular replacement into the higher resolution native data set using PHASER (McCoy et al., 2007). Iterative rounds of manual model building and refinement, which included translation-libration-screen (TLS) models, used COOT and PHENIX, respectively. The native GmPDH1 structure was used for molecular replacement to solve the GmPDH1 N222D and GmPDH1 M219T/N222D structures. Each mutant structure was built and refined using the same method as the wild-type enzyme. Data collection and refinement data are summarized in Table 1. The final model of SeMet-substituted GmPDH1 included residues Ser9
to Gln258 and NADP⁺ for both molecules in the asymmetric unit and 228 waters with no Ramachandran outliers (97.8% favored and 2.2% allowed). The final model of the GmPDH1•NADP⁺•citrate complex included residues Gln8 to Ile257 for chain A and residues Gln8 to Thr260 for chain B, NADP⁺ and citrate in both chains, and 605 waters with no Ramachandran outliers (97.7% favored and 2.3% allowed). The structure was intended to be an apoenzyme, but NADP⁺ and citrate were bound in the active site. The final model of the GmPDH1 N222D•NADP⁺•Tyr complex included residues Ser9 to Met258 for chain A and residues Gln8 to Thr260 for chain B, NADP⁺ and Tyr in both chains, and 435 waters with no Ramachandran outliers (97.8% favored and 2.2% allowed). The final model of the GmPDH1 M219T/N222D•NADP⁺•Tyr complex included residues Ser9 to Ile257 for chain A and residues Gln8 to Ile257 for chain B, NADP⁺ and Tyr in both chains, and 616 waters with no Ramachandran outliers (97.5% favored and 2.5% allowed).

**ADH and PDH assay**

Kinetic parameters of purified recombinant enzymes were determined from assays containing varying concentrations of arogenate (19.5 µM - 5 mM) or prephenate (23.4 µM - 6 mM) substrate together with 25 mM HEPES pH 7.6, 50 mM KCl, 10% (v/v) ethylene glycol, and 0.5 mM NADP⁺ cofactor. The reactions were initiated by addition of enzyme and incubated at 37°C and monitored every 10-15 seconds at A₃₄₀nm using a microplate reader (Tecan Genios). Kinetic parameters were determined by fitting initial velocity data to the Michaelis-Menten equation using Origin software (OriginLab) from technical replicate assays (n ≥ 3). Arogenate was prepared by enzymatic conversion of prephenate (Sigma-Aldrich) as previously reported (Maeda et al., 2010). For Tyr inhibition assays, varying amounts of Tyr (10 µM - 8 mM) were mixed
with 0.5 mM NADP$^+$ and either 1 mM arogenate or 0.8 mM prephenate, followed by addition of enzyme and monitoring $A_{340\text{nm}}$ as above from technical replicate assays ($n = 3$). Tyr was dissolved in a basic solution (0.025 N NaOH) for sufficient solubility, and the HEPES buffer concentration was increased to 500 mM to maintain a constant pH 8.5. For HPP inhibition assays, varying amounts of HPP (10 µM - 2 mM) were mixed with 0.5 mM NADP$^+$ and 0.8 mM prephenate, followed by addition of enzyme and monitoring $A_{340\text{nm}}$ as above from technical replicate assays ($n = 3$). HPP was dissolved in a 20% solution of dimethyl sulfoxide for sufficient solubility and further diluted to 2% final concentration in the assays. All enzyme assays were conducted at a reaction time and protein concentration that were in the linear range and proportional to reaction velocity.

**Computational Substrate Docking**

Molecular docking of arogenate into the GmPDH1 M219T/N222D•NADP$^+$•Tyr three-dimensional model with Tyr removed was performed using AutoDock Vina (ver. 1.1.2) (Trott and Olson, 2010). The positions of NADP$^+$ and Tyr in the structure was used to guide docking with a grid box of 30 x 30 x 30 Å and the level of exhaustiveness set to 8.

**Data Availability**

Genebank accession codes for the sequences mentioned in this article are as follows: *G.max*_18g02650 (GmPDH1), KM507071; *G.max*_14g05990 (GmncADH), KM507073; *M.truncatula*_3g071980 (MtPDH), KM507076; *M.trunctula*_5g083530 (MtncADH), KX957934; *A.ipaensis*_VYE8T (AiPDH/ADH), KX957935; *S.lycopersicum*_6g050630 (SlncADH), KX957936. Coordinates and structure factors for SeMet GmPDH1•NADP$^+$ (PDB:
5T9E), GmPDH1•NADP⁺•citrate (PDB: 5T8X), GmPDH1 N222D•NADP⁺•Tyr (PDB: 5T9F), and GmPDH1 M219T/N222D•NADP⁺•Tyr (PDB: 5T95) were deposited in the RCSB Protein Data Bank.

RESULTS

Identification and biochemical analysis of noncanonical ADH in legumes

Our previous phylogenetic analysis of plant TyrA enzymes identified a “noncanonical” clade (gray box in Fig. 1b) containing legume PDHs that was distinct from the “canonical” ADHs present in all plant lineages (Schenck et al., 2015). The noncanonical clade also contained additional homologs from some angiosperms (through absent in e.g. Brassicales, Cucurbitales, and Rosales; Fig. 1b). For comparison of the biochemical properties of PDHs and their noncanonical TyrA homologs, representative members of each group were expressed as recombinant His-tagged proteins without chloroplast transit peptides (if present), purified for steady-state kinetic analysis, and compared with previously characterized PDHs and a canonical ADH (Fig. 1c, Table 2). PDHs from *Glycine max* (GmPDH1: G.max_18g02650) and *Medicago truncatula* (MtPDH; M.truncatula_3g071980) preferred prephenate versus arogenate as substrates with 139-fold and 21-fold higher $k_{cat}/K_m$ values, respectively, consistent with previous characterization. The noncanonical TyrA homolog from soybean (G.max_14g05990) only displayed activity with arogenate, whereas that from *M. truncatula* (M.truncatula_5g083530) accepted both substrates but was ~6,200-fold more efficient with arogenate, similar to previously characterized ADH from *Arabidopsis thaliana* (AtADH2; A.thaliana_1g15710) (Rippert and Matringe, 2002). Thus, G.max_14g05990 and M.truncatula_5g083530 are noncanonical ADHs (GmncADH and MtncADH, respectively). Each of the legume noncanonical ADH strongly
prefers NADP$^+$ over NAD$^+$ as cofactor consistent with previously reported plant ADH and PDHs (Gaines et al., 1982; Rippert and Matringe, 2002; Schenck et al., 2015). In addition to substrate specificity, these three types of plant ADHs and PDHs differ in feedback inhibition by Tyr (Fig. 1d; Tables 2). The canonical AtADH2, was highly sensitive to Tyr (IC$_{50}$ = 38 µM), whereas GmPDH1 and MtPDH were insensitive to feedback inhibition by Tyr (up to 8 mM in assays, Fig. 1d), consistent with prior studies. The noncanonical ADHs, GmncADH and MtncADH, were sensitive to Tyr but with IC$_{50}$ values in the mM range. Thus, unlike PDHs, legume noncanonical ADHs are partially inhibited by Tyr.

To further define the phylogenetic boundaries of noncanonical ADH and PDHs additional homologs from *Arachis ipaensis* (peanut, *A. ipaensis* _VYE8T*) and *Solanum lycopersicum* (tomato, *S. lycopersicum* _6g050630*), which exist at key phylogenetic boundaries (Fig. 1b), were biochemically characterized. *A. ipaensis* _VYE8T* (AiPDH/ADH) used both arogenate and prephenate to similar degrees ($k_{cat}/K_m = 11.6$ and 14.9 mM$^{-1}$ s$^{-1}$, respectively), whereas *S. lycopersicum* _6g050630* (SlncADH) exhibited ADH but no PDH activity (Fig. 1c). AiPDH/ADH was insensitive to Tyr inhibition, whereas SlncADH showed relaxed sensitivity to Tyr with an IC$_{50}$ = 12.8 mM (Fig. 1d; Tables 2), similar to legume ncADHs. Thus, legume enzymes having considerable PDH activity are Tyr insensitive and form a subclade within the noncanonical clade likely due to a recent gene duplication of an ncADH within legumes (Fig. 1b).

**X-ray crystal structure of soybean PDH**

To understand the structure-sequence relationship of legume PDHs and ADHs, and because TyrA structures from plants are not available (Legrand et al., 2006; Sun et al., 2006; Chiu et al.,...
2010; Ku et al., 2010), the x-ray crystal structure of GmPDH1 was determined by single-wavelength anomalous dispersion phasing using selenomethionine-substituted protein (Table 1). The resulting model was then used for molecular replacement with a 1.69 Å resolution native data set to solve the structure of the GmPDH1•NADP⁺•citrate complex (Fig. 2a; Table 1).

GmPDH1 forms a homodimer with each 257 amino acid monomer adopting a N-terminal Rossman fold domain (residues 8-171) that shapes the NADP(H)-binding domain and an α-helical C-terminal dimerization domain (residues 172-257) (Fig. 2a). The PDH dimer is formed by two tail-to-tail monomers that pack closely resulting in a dumbbell-shaped molecule (Fig. 2a). The N-terminal domain is made up of seven β-strands sandwiched between two sets of three α-helices. The C-terminal dimerization domain consists of an entirely helical architecture of four α-helices. The active site in each monomer is found at the interface of the two domains.

Consistent with the NADP⁺ specificity of GmPDH16, the crystal structure of GmPDH1 shows clear electron density for this ligand in the N-terminal domain of each monomer (Fig. 2b) and extensive protein-ligand binding interactions (Fig. 2c). The β1a-α1 loop (residues 16-21) is the conserved GxGxxG motif characteristic of NAD(P)(H)-dependent oxidoreductases (Wierenga et al., 1985) and contributes interactions with the pyrophosphate moiety and the nicotinamide ring. The main-chain amides of Asn19 and Phe20 hydrogen bond with an oxygen atom in the diphosphate linker. The hydroxyl group of Ser223 interacts with another phosphate oxygen. Additionally, contacts with five water molecules further stabilize the disphosphate linker. The syn-conformation of the nicotinamide ring is stabilized by π-π stacking interactions with Phe20 and by polar contacts between N1 and the side-chain of Ser101. Water molecules also interact with the carboxamide oxygen and nitrogen. These interactions orient the B-face of the nicotinamide ring toward the substrate binding pocket.
Figure 2. X-ray crystal structure of GmPDH1.

(a) Ribbon diagram showing the monomeric units (colored gold and white, respectively) of the homodimer. NADP\(^+\) (green) and citrate (purple) are depicted as space-filling models. The N- and C-terminal domains are also indicated. (b) Electron density for NADP\(^+\). The 2Fo-Fc omit map (1.5 \(\sigma\)) for the ligand is shown. (c) Nicotinamide cofactor binding pocket of GmPDH1. Residues surrounding the bound NADP\(^+\) (green) and water molecules (red spheres) are shown. Ligand interactions are indicated by dotted lines. (d) Active site residues in GmPDH1 in contact with citrate (purple) identify the proposed prephenate binding site.

Other interactions complete the cofactor binding site (Fig. 2c). The adenine ring, which is in the \textit{anti}-conformation, hydrogen bonds to the side-chain of Glu80 and a water molecule through its exocyclic N6 and to the hydroxyl group of Thr73 via N3 and N9. Water molecules form polar interactions with the adenine N3 and N7. Extensive charge-charge interactions are formed between the 2′-phosphate of the adenine ribose and the side-chain of Arg40, the hydroxyl groups of Ser39, Ser41, and Tyr43, the backbone amide nitrogen of Ser41, and three water molecules. These interactions form the phosphate binding site that favors NADP(H) over NAD(H). The 3′-phosphate of the adenine ribose interacts with the main-chain amide of Gly18 and the ring
<table>
<thead>
<tr>
<th></th>
<th>GmPDH1(SeMet) •NADP⁺</th>
<th>GmPDH1•NA DP⁺•citrate</th>
<th>GmPDH1 N222D •NADP⁺•Tyr</th>
<th>GmPDH M219T/N222D •NADP⁺•Tyr</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data collection</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Space group</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
<td>P1</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a, b, c (Å)</td>
<td>46.51, 55.13, 68.59</td>
<td>46.00, 55.28, 67.94</td>
<td>46.46, 55.05, 68.93</td>
</tr>
<tr>
<td></td>
<td>α, β, γ (°)</td>
<td>107.3, 98.9, 103.6</td>
<td>107.4, 98.9, 103.2</td>
<td>107.8, 99.6, 102.6</td>
</tr>
<tr>
<td>Wavelength</td>
<td></td>
<td>0.979</td>
<td>0.979</td>
<td>0.979</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td></td>
<td>34.1 - 2.03</td>
<td>32.4 - 1.69</td>
<td>33.9 - 1.99</td>
</tr>
<tr>
<td>R_sym (%)</td>
<td>6.1 (43.2)</td>
<td>4.4 (43.9)</td>
<td>8.5 (28.5)</td>
<td>4.9 (34.4)</td>
</tr>
<tr>
<td>I / σ</td>
<td>7.6 (1.6)</td>
<td>13.8 (1.1)</td>
<td>10.2 (2.3)</td>
<td>12.3 (1.7)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>96.0 (88.7)</td>
<td>97.0 (94.6)</td>
<td>88.9 (85.3)</td>
<td>88.9 (87.2)</td>
</tr>
<tr>
<td>Redundancy</td>
<td>1.80</td>
<td>1.96</td>
<td>1.69</td>
<td>1.78</td>
</tr>
<tr>
<td><strong>Refinement</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td></td>
<td>34.1 - 2.03</td>
<td>32.4 - 1.69</td>
<td>33.9 - 1.99</td>
</tr>
<tr>
<td>No. unique reflections</td>
<td>37,512</td>
<td>64,687</td>
<td>36,694</td>
<td>59,535</td>
</tr>
<tr>
<td>R_work / R_free (%)</td>
<td>18.8 / 22.9</td>
<td>15.3 / 18.2</td>
<td>15.8 / 20.6</td>
<td>15.4 / 18.4</td>
</tr>
<tr>
<td>No. atoms</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>4017</td>
<td>4094</td>
<td>4054</td>
<td>4084</td>
</tr>
<tr>
<td>Ligand</td>
<td>224</td>
<td>604</td>
<td>435</td>
<td>616</td>
</tr>
<tr>
<td>Water</td>
<td>96</td>
<td>122</td>
<td>122</td>
<td>122</td>
</tr>
<tr>
<td>B-factors</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>29.1</td>
<td>23.7</td>
<td>32.5</td>
<td>21.4</td>
</tr>
<tr>
<td>Ligand</td>
<td>25.3</td>
<td>19.6</td>
<td>32.0</td>
<td>15.3</td>
</tr>
<tr>
<td>Water</td>
<td>37.4</td>
<td>37.5</td>
<td>41.3</td>
<td>37.3</td>
</tr>
<tr>
<td>R.m.s. deviations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.010</td>
<td>0.007</td>
<td>0.008</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>1.17</td>
<td>1.17</td>
<td>0.91</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 1. **Data collection and refinement statistics.** *Data for each structure represents a single crystal. *Values in parentheses are for highest-resolution shell.
oxygen of the ribose hydrogen bonds to the hydroxyl group of Thr73. Both the adenine ribose and the nicotinamide ribose adopt the C2′-endo conformation. The 2′-hydroxyl of the nicotinamide ribose interacts with the side-chain hydroxyl and the main-chain nitrogen of Ser101, whereas the 3′-hydroxyl of the nicotinamide ribose hydrogen bonds to the backbone oxygen of Thr73. A water molecule interacts with the 2′- and 3′-hydroxyls of the nicotinamide ribose.

Figure 3. Electron density maps of citrate and Tyr ligands. (a) The 2Fo-Fc omit map (1.25 σ) for the citrate ligand from GmPDH1•NADP+•citrate complex is shown. (b) The 2Fo-Fc omit map (1.25 σ) for the Tyr ligand from GmPDH M219T/N222D•NADP•Tyr complex is shown.

Although efforts to obtain crystals with different ligands (e.g. prephenate and HPP) were not successful, the structure of PDH complexed with NADP+ and citrate, contributed from the crystallization buffer, suggests how substrates may bind within the active site (Fig. 2d). The citrate is positioned in a pocket proximal to the nicotinamide ring and the putative catalytic histidine (His124). The Ne of His124 and the side-chain amine of Gln184 form polar contacts with the α-carboxyl group of citrate. Similarly, the side-chain nitrogen of Gln184 and Ne of His188 contact the γ-hydroxyl of citrate. The ζ-carboxyl group of citrate interacts with the hydroxyl of Thr206, which is provided by the other subunit at the dimer interface. Additional polar contacts are made between the ε-carboxyl and the hydroxyl of Thr131 and the side-chain
amide of Asn222. The binding of citrate, which mimics the dicarboxylate portion of prephenate, identifies potential residues in the substrate binding site.

**Identification of a residue that confers TyrA substrate specificity**

Next, the predicted substrate binding site (Fig. 2d) and the phylogenetic distribution of PDH and ADHs (Fig. 1b) were used together to identify residues responsible for differences in substrate specificity. Amino acid alignment of the plant TyrA enzymes showed highly conserved residues responsible for NADP⁺ binding, including the GxGxxG motif, and residues proposed to function in catalysis (e.g. Ser101 and His124) (Christendat et al., 1998; Christendat and Turnbull, 1999; Sun et al., 2006; Sun et al., 2009). Within the PDH active site, residues uniquely conserved in either ADHs or PDHs were also identified (Fig. 5a). Asp218 in GmncADH, which corresponds to Asn222 in GmPDH1, was highly conserved among ADHs but not in PDHs (Fig. 5a). Similarly, Thr215 of GmncADH was generally conserved among ADHs but replaced by either Met or Val in PDHs, although peanut AiPDH/ADH retains a Thr at the corresponding position (Fig. 5a). These comparisons suggest that either Met219 or Asn222 (or both) may determine prephenate specificity in PDH.

To experimentally test the roles of the two residues in prephenate versus arogenate substrate specificity, site-directed mutagenesis was performed on GmPDH1 to convert Asn222 and Met219 into the corresponding residues in GmncADH (N222D and M219T). The M219T mutant had very similar kinetic parameters to wild-type enzyme preferring prephenate over arogenate substrate (Fig. 5b; Tables 3). The N222D mutant, however, showed a 115-fold reduction in $k_{cat}/K_m$ with prephenate and gained ADH activity (Fig. 5b; Table 3). The turnover rate ($k_{cat}$) of N222D for arogenate (27.8 s⁻¹) was comparable to wild-type GmPDH1 and
GmncADH for prephenate and arogenate, respectively (30.4 and 27.7 s\(^{-1}\); Table 2). The M219T/N222D double mutant, exhibited very similar \(k_{\text{cat}}/K_m\) values for PDH and ADH activity compared to the N222D single mutant (Fig. 5b; Table 3), suggesting that the M219T substitution had little effect on substrate specificity alone and in combination with the N222D mutation.

To test if the analogous mutation alters substrate specificity outside of soybean PDH, the Asp residue was introduced to the corresponding Cys on *Medicago* PDH, MtPDH. Similar to the GmPDH1 N222D mutant, the C220D mutation on MtPDH reduced PDH activity by >1000-fold and enhanced ADH activity by ~10-fold (Fig. 5b), which is also reflected by a 31-fold higher and 3-fold lower \(K_m\) toward prephenate and arogenate, respectively, compared to wild-type (Tables 2 and 3). To examine if an acidic Asp residue was necessary for converting PDH to ADH activity, an alanine mutation was introduced at Asn222 in GmPDH1 (N222A). The N222A mutant reduced PDH activity, but did not introduce ADH activity, unlike N222D (Fig. 5b; Table 3). These results suggest that the corresponding 222 position in legume PDH

### Table 2. Kinetic analysis of representative plant PDHs and ADHs.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Substrate</th>
<th>(k_{\text{cat}}) (s(^{-1}))</th>
<th>(K_m) (mM)</th>
<th>(k_{\text{cat}}/K_m) (M(^{-1}) s(^{-1}))</th>
<th>(IC_{50,\text{Tyr}}) (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmPDH1</td>
<td>Prephenate</td>
<td>30.4 ± 0.7</td>
<td>0.09 ± 0.01</td>
<td>337,800</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Arogenate</td>
<td>6.3 ± 0.7</td>
<td>2.59 ± 0.09</td>
<td>2,430</td>
<td>31.3 ± 9.3</td>
</tr>
<tr>
<td>MtPDH1</td>
<td>Prephenate</td>
<td>18.5 ± 3.0</td>
<td>0.05 ± 0.01</td>
<td>370,000</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Arogenate</td>
<td>16.9 ± 1.1</td>
<td>0.94 ± 0.04</td>
<td>17,980</td>
<td>--</td>
</tr>
<tr>
<td>Peanut PDH/ADH</td>
<td>Prephenate</td>
<td>2.8 ± 0.2</td>
<td>0.19 ± 0.01</td>
<td>14,740</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Arogenate</td>
<td>3.2 ± 0.1</td>
<td>0.28 ± 0.03</td>
<td>11,430</td>
<td>--</td>
</tr>
<tr>
<td>GmncADH</td>
<td>Prephenate</td>
<td>27.7 ± 1.1</td>
<td>0.41 ± 0.03</td>
<td>67,560</td>
<td>15.5 ± 0.8</td>
</tr>
<tr>
<td></td>
<td>Arogenate</td>
<td>39.0 ± 7.6</td>
<td>0.14 ± 0.02</td>
<td>278,600</td>
<td>2.2 ± 0.5</td>
</tr>
<tr>
<td>MtncADH</td>
<td>Prephenate</td>
<td>0.3 ± 0.1</td>
<td>6.69 ± 0.27</td>
<td>45</td>
<td>0.6 ± 0.1</td>
</tr>
<tr>
<td></td>
<td>Arogenate</td>
<td>15.9 ± 2.1</td>
<td>0.45 ± 0.05</td>
<td>35,330</td>
<td>12.8 ± 2.0</td>
</tr>
</tbody>
</table>

\(^{a}\)No inhibition was detected. 
\(^{b}\)Not detectable.
enzymes is the key determinant for their substrate specificity, where an acidic Asp residue is crucial for ADH activity.

Figure 4. Identification of Asn222 as a determinant of PDH activity and Tyr sensitivity. (a) Trimmed amino acid alignment corresponding to the phylogeny in Fig. 1b highlighting residues Met219 and Asn222 (number based on GmPDH1). Identical amino acids present in >50% of the sequences are shaded black and biochemically similar residues present in >50% of the sequences are shaded gray. (b) PDH (red) and ADH (blue) activity of GmPDH1, MtPDH and corresponding site-directed mutants. Bars represent average catalytic efficiency ($k_{cat}/K_m$) in mM$^{-1}$ s$^{-1}$ ± SEM of $n = 3$ replications. Data without bars had detectable activity unless N.D. (below detection limit) is indicated (see Tables 2 and 3 for values). (c) Effect of Tyr on PDH activity of wild-type and mutant GmPDH1 and MtPDH. Data are shown as IC$_{50}$ plots with enzymatic activity determined at increasing concentrations of L-Tyr (at the following concentrations: 0, 0.01, 0.1, 0.5, 1, 5, and 8 mM). Activity was normalized to an assay with no L-Tyr and expressed as percent activity of $n = 3$ ± SEM. Open symbols correspond to wild-type enzymes, with dashed lines. Mutant enzymes have filled symbols with solid lines.
Altered substrate specificity simultaneously affects Tyr-sensitivity

The mutations on legume PDHs were also tested for their effect on Tyr sensitivity. Similar to wild-type GmPDH1, the M219T and N222A single mutants, which did not alter substrate specificity, were not inhibited by Tyr (Fig. 5c; Table 3). In contrast, the GmPDH1 N222D and M219T/N222D mutants, as well as the MtPDH C220D mutant, exhibited Tyr inhibition with IC<sub>50</sub> values of 5 to 11 mM (Fig. 5c; Table 3). Thus, mutating Asn222 and Cys220 of GmPDH1 and MtPDH, respectively, into an Asp not only introduced ADH activity, but also Tyr sensitivity.

Figure 5. Crystal structures of GmPDH1 N222D and M219T/N222D reveal Tyr binding interactions.
(a) Ribbon diagrams of overlay of GmPDH1 (blue), GmPDH1 N222D (rose), and GmPDH1 M219T/N222D (white) with NADP<sup>+</sup> (green) shown as a space-filling model. (b) Active site overlay of wild-type and mutant GmPDH1 shows a conserved architecture. Coloring of side-chains is the same as panel a. (c) Active site residues in GmPDH1 M219T/N222D in contact with Tyr (purple). (d) Molecular docking of arogenate (rose) into the active site of GmPDH1 M219T/N222D. The surface of the active site pocket is shown with the surface corresponding to Asp222 colored red.
The GmPDH1 mutants that bind to Tyr can now be used to test the role of the active site Asp222 in ADH activity and Tyr sensitivity. The GmPDH1 N222D and M219T/N222D mutants were successfully co-crystallized with Tyr and NADP⁺ bound in their active site at 1.99 and 1.69 Å resolution, respectively (Table 1, Fig. 5). An overlay of these two mutants with the wild-type structure revealed no global conformational changes (Fig. 5a). Likewise, the substitutions did not drastically alter the active site structure of either mutant (Fig. 5b).

In the GmPDH1 M219T/N222D structure, the ring hydroxyl of the Tyr ligand contacts Ne of His124, the hydroxyl of Ser101, and the amine group of Gln184 (Fig. 5c). The side chain carboxylate of Tyr interacts with the hydroxyl group and backbone amide of Thr131, as well as the carbonyl and backbone amide of Gln130. The position of the bound Tyr is also stabilized by π-π stacking interaction with the nicotinamide ring of NADP⁺. The amine nitrogen of Tyr forms polar contacts with a water molecule, the carbonyl of Gln130, and the carboxylate of the mutated Asp222 residue. Identical contacts were observed in the GmPDH1 N222D structure. Neither Met219 nor the mutated Thr219 makes a direct contact with the ligand.

In the GmPDH1 mutant structures, the active site pocket near the site of hydride transfer from the substrate to the nicotinamide via His124 is composed of a wall of nitrogen atoms (e.g. of Gln184 and His188), and Asp222 adds a negatively charged region to the side of the pocket to recognize the amine of Tyr (Fig. 5c). Computational docking of arogenate into the crystallographic structure of GmPDH1 M219T/N222D shows that the hydroxyl of arogenate can anchor itself between His124 and the nicotinamide ring, similar to Tyr (Fig. 5d). Also, the carboxylate of Asp222 could form a polar interaction with the amine of arogenate (Fig. 5d). By mutating the 222 residue from a neutral Asn to a negatively charged Asp, the specificity in
substrate recognition changes to preferentially recognize the amine of arogenate over the carbonyl of prephenate and also introduce sensitivity to Tyr.

**Mutating Asp218 introduces PDH activity in divergent plant ADH enyzmes**

To test if PDH activity can be introduced to legume ncADHs, the reciprocal mutation was made on GmncADH at position Asp218 (corresponding to Asn222 of GmPDH1) to generate the D218N mutant. The D218N substitution reduced $k_{cat}/K_m$ for ADH by ~6-fold while introducing PDH activity (**Fig. 6a; Table 3**) into an enzyme which was originally unable to use prephenate (**Fig. 1c; Table 2**). The corresponding Asp to Cys mutation on MtncADH (D220C) showed similar results, e.g. reduced ADH activity and enhanced PDH activity (**Fig. 6a**). While wild-type MtncADH had a 6,190-fold preference for arogenate, MtncADH D220C was switched to prefer prephenate over arogenate by 1.7-fold (**Table 3**). These results suggest that multiple residues at the corresponding 222 position are able to confer PDH activity, whereas only an Asp residue is present in all ADHs (**Figure 4a**).

The corresponding Asp residue was also mutated to Asn in divergent ADH from the basal noncanonical clade, tomato (SlncADH D224N), and canonical ADH clade, Arabidopsis (AtADH2 D241N, **Fig. 1b**). Similar to the results observed with the legume ncADHs, the tomato and Arabidopsis mutant enzymes gained PDH activity at the expense of reduced ADH activity (**Fig. 6a; Table 3**). Additionally, each ADH mutant (GmncADH D218N, MtncADH D220C, SlncADH D224N, and AtADH2 D241N) were less sensitive to Tyr inhibition than the respective wild-type enzymes (**Fig. 6b; Tables 2 and 2**). Instead, GmncADHD218N and MtncADH D220C were slightly inhibited by their new product HPP. Thus, the alteration of the
Figure 6. Asn222 confers PDH activity to divergent plant ADHs while simultaneously relaxing Tyr sensitivity.  

(a) ADH activity from wild-type ADH enzymes and their mutants that remove Asp at the corresponding 222 position. Bars represent average catalytic efficiency ($k_{cat}/K_m$) in mM$^{-1}$ s$^{-1}$ ± SEM for n = 3. Activity from AtADH2 is shown as specific activity (nkat/mg ± SEM for n = 3) as kinetics were unable to be determined. Data without bars had detectable activity unless N.D. (below detection limit) is indicated (see Tables 2 and 3 for values).  

(b) IC$_{50}$ plots show Tyr sensitivity of ADH activity from wild-type and mutated ADHs. Enzymes were tested for ADH activity at increasing concentrations of Tyr (at the following concentrations: 0, 0.01, 0.1, 0.5, 1, 5, and 8 mM) and were normalized to the 0 mM assay. Bars are average activity ± SEM for n = 3. Open symbols correspond to wild-type enzymes, with dashed curves. Mutant enzymes have filled symbols with solid curves.
key active site Asp residue is the evolutionary switch needed to introduce PDH activity in diverse plant ADH enzymes while simultaneously relieving feedback inhibition by Tyr.

<table>
<thead>
<tr>
<th>protein</th>
<th>substrate</th>
<th>$k_{cat}$ (s$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat} / K_m$ (M$^{-1}$s$^{-1}$)</th>
<th>IC$_{50}^{Tyr}$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmPDH1 M219T</td>
<td>prephenate</td>
<td>30.3 ± 0.8</td>
<td>0.10 ± 0.02</td>
<td>303.000</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>arogenate</td>
<td>4.6 ± 0.3</td>
<td>1.55 ± 0.14</td>
<td>2,968</td>
<td>--</td>
</tr>
<tr>
<td>GmPDH1 N222D</td>
<td>prephenate</td>
<td>19.1 ± 1.4</td>
<td>7.58 ± 1.13</td>
<td>2,520</td>
<td>5.3 ± 0.4</td>
</tr>
<tr>
<td></td>
<td>arogenate</td>
<td>27.8 ± 3.5</td>
<td>0.53 ± 0.18</td>
<td>52,450</td>
<td>4.7 ± 0.4</td>
</tr>
<tr>
<td>GmPDH1 N222A</td>
<td>prephenate</td>
<td>6.6 ± 0.3</td>
<td>0.19 ± 0.04</td>
<td>34,740</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>arogenate</td>
<td>4.6 ± 0.2</td>
<td>0.19 ± 0.04</td>
<td>34,740</td>
<td>--</td>
</tr>
<tr>
<td>GmPDH1 M219T/N222D</td>
<td>prephenate</td>
<td>2.5 ± 0.3</td>
<td>1.18 ± 0.12</td>
<td>2,119</td>
<td>11.1 ± 1.2</td>
</tr>
<tr>
<td>MtPDH C220D</td>
<td>prephenate</td>
<td>29.0 ± 4.4</td>
<td>0.63 ± 0.18</td>
<td>46,030</td>
<td>5.9 ± 0.5</td>
</tr>
<tr>
<td></td>
<td>arogenate</td>
<td>46.8 ± 4.2</td>
<td>0.27 ± 0.01</td>
<td>173,300</td>
<td>--</td>
</tr>
<tr>
<td>GmncADH D218N</td>
<td>prephenate</td>
<td>11.5 ± 0.5</td>
<td>1.98 ± 0.05</td>
<td>5,810</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>arogenate</td>
<td>8.6 ± 0.2</td>
<td>0.74 ± 0.14</td>
<td>11,620</td>
<td>--</td>
</tr>
<tr>
<td>MtncaADH D220C</td>
<td>prephenate</td>
<td>10.1 ± 0.5</td>
<td>0.74 ± 0.03</td>
<td>13,650</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>arogenate</td>
<td>7.0 ± 0.8</td>
<td>0.87 ± 0.03</td>
<td>8,046</td>
<td>7.7 ± 1.5</td>
</tr>
<tr>
<td>SolyncADH D224N</td>
<td>prephenate</td>
<td>2.4 ± 1.0</td>
<td>2.31 ± 0.10</td>
<td>1,040</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>arogenate</td>
<td>11.7 ± 0.2</td>
<td>1.34 ± 0.48</td>
<td>8,730</td>
<td>--</td>
</tr>
</tbody>
</table>

*No inhibition was detected.

Table 3. Steady-state kinetic parameters and effect of tyrosine on mutant GmPDH1, MtPDH1, GmncADH, MtncaADH, and SolyncADH.

DISCUSSION

In plants, aromatic amino acid biosynthesis provides essential building blocks for proteins and diverse primary and specialized metabolites (Tzin and Galili, 2010; Maeda and Dudareva et al., 2012); however, the biochemical pathways for production of these compounds can vary, as exemplified in Tyr biosynthesis. While all plants have canonical ADH for Tyr synthesis (Gaines et al., 1982; Connelly and Conn, 1986; Rippert and Matringe, 2002), our studies found that some angiosperms have noncanonical ADH (ncADH) and some legumes additionally have PDH (Fig. 1b-c) (Schenck et al., 2015). The three types of TyrA dehydrogenases share similar catalytic properties, but with distinct arogenate versus prephenate specificities (Fig. 1c; Table 2) (Rubin and Jensen, 1979; Connelly and Conn, 1986; Rippert and Matringe, 2002; Schenck et al., 2015).
Moreover, the final pathway product, Tyr, strongly feedback inhibits the canonical ADHs and partially inhibits ncADHs (Fig. 1d), whereas legume PDHs are completely insensitive to Tyr (Fig. 1d) (Schenck et al., 2015). Also, unlike plastid-localized canonical ADHs (Rippert et al., 2009), ncADH from eudicots and PDH lack an N-terminal chloroplast transit peptide and localize in the cytosol (Schenck et al., 2015), as also shown for cytosolic CM and TAT isoforms that function before and after PDH, respectively (Westfall et al., 2014; Wang et al., 2016). Our data suggest that alternative Tyr pathways with distinct regulation and cellular localization have evolved in different plants, though their physiological functions remain unknown.

Phylogenetic analyses of plant TyrAs showed that the legume PDH genes evolved through duplication of an ancestral plant ADH that likely occurred before divergence of basal angiosperms (Fig. 1), followed by subfunctionalization, rather than horizontal gene transfer of a bacterial PDH (Schenck et al., 2015). PDH enzymes are restricted to legumes, particularly in the more recently-diverged species, such as peanut and soybean (Fig. 1b) (Wojciechowski et al., 2004; Cardoso et al., 2012). Therefore, the PDH genes evolved through an ancient duplication event giving rise to the angiosperm noncanonical clade, which was followed by a second duplication within the legume family.

The current study demonstrates that alteration of Asp218 (into either Asn or Cys) played a key role during the subfunctionalization of the duplicated gene from ADH to PDH (Figs. 5 and 6). Comparison of the x-ray crystal structures of the wild type and N222D mutants of GmPDH1 (Figs. 2 and 6) showed that the Asp substitution is readily accommodated in the active site without significant conformational changes (Fig. 5a-b). Prephenate and arogenate are nearly identical with the exception of a carbonyl versus an amine, respectively (Fig. 1a). Positioning of the carboxylate side-chain of the Asp residue in the GmPDH1 mutants provides an energetically
dominant ionic interaction with the amine of arogenate substrate (Fig. 5d), which would be protonated at physiological pH, compared to a hydrogen bond with the prephenate carbonyl group in the wild-type enzyme. The same charge-charge interaction is also critical for feedback inhibition in the GmPDH1 mutants (Fig. 5c; Table 3) and binding with Tyr, which also has the side-chain amine (Fig. 5c).

Although introduction of Asp218 into GmPDH1 restored ADH activity to near wild-type levels of GmncADH ($k_{cat}/K_m$ of 52.5 vs 67.5, respectively, Fig. 5b; Tables 2 and 3), that of Asn222 into GmncADH was insufficient to obtain PDH activity comparable to wild-type GmPDH1 level (Figs. 5b and 6a). An additional mutation of Met219, which covaries with Asn222, on GmPDH1 wild-type and N222D mutant did not enhance ADH activity (Fig. 5b). Comparisons among GmPDH1, GmncADH, and AtADH2 reveal variety in the amino acid sequence of the $\beta_1e-\beta_1f$ loop (Phe127 to Trp136 in GmPDH1, which is at the opposing side of the active site from Asn222 and consists of residues that interact with the ligand side chain carboxylate (Fig. 5c). Thus, residues on the $\beta_1e-\beta_1f$ loop could be contributing to the correct positioning of the substrate for catalysis, and various combinations of active site mutations on both sides may be needed to convert an ADH to a fully functional PDH.

The residue corresponding to Asp218 that confers ADH activity can now be used to trace the evolutionary origin of the plant ADHs. Asp218 is present in TyrA homologs of all plants and algae (Fig. 7), suggesting that Asp218-containing ADHs are universal to the plant kingdom. Plant TyrA sequences were most similar to those of proteobacteria among other microbial TyrAs (Fig. 7) (Bonner et al., 2008; Reyes-Prieto et al., 2012; Dornfeld et al., 2014). Interestingly, an Asp was present at the corresponding 218 position in the TyrA orthologs of proteobacteria, which was previously shown to have ADH activity (e.g. Phenyllobacterium immobile) (Mayer et
al., 1985), but absent in those of other microbes (Figs. 7b-c, 8) (Legrand et al., 2006). Together these data suggest that ADH containing Asp218 evolved in a bacteria ancestor, which was horizontally transferred to the common ancestor of plants and algae. Together with PPA-ATs acquired from a Chlorobi/Bacteroidetes ancestor (Dornfeld et al., 2014), the Asp218-containing ADHs are maintained in the plant kingdom for synthesis of Tyr via the arogenate pathway (Fig. 1a).

Is the corresponding Asp residue also responsible for substrate specificity and regulation of divergent microbial TyrA dehydrogenases? In conclusion, microbial TyrA dehydrogenases likely have different mechanisms of substrate specificity from plant TyrAs. Here we compared the three-dimensional structure of GmPDH1 (Fig. 2), the first of a plant TyrA structure, to previously reported microbial TyrAs from the cyanobacteria Synechocystis sp. PCC 6803 (SynADH; PDB: 2F1K) (Legrand et al., 2006) and A. aeolicus PDH (AaPDH; PDB: 3GGP) (Sun et al., 2009). SynADH is specific to arogenate and Tyr insensitive, whereas AaPDH prefers prephenate and is inhibited by Tyr (Legrand et al., 2006; Bonvin et al., 2006; Sun et al., 2006; Sun et al., 2009). The overall fold of GmPDH1 is conserved (root mean square deviations of 2.5-3.0 Å for ~235 Cα atoms) with SynADH and AaPDH (Fig. 8a). While the N-terminal Rossmann-fold was highly conserved, some differences in topology were found in the C-terminal dimerization domain: the 3_10 helix (α9) and the long C-terminal helix (α13) of SynADH and AaPDH are missing in the soybean enzyme, and the α7 helix of GmPDH1 is split into two helices in SynADH and AaPDH (α7 and α8, Fig. 8a).

Comparison of cofactor binding sites of these enzymes also reveals structural variation that defines the NADP(H) cofactor specificity of GmPDH1 (Fig. 8a). The β1b-α2 loop of GmPDH1 (Ser39-Tyr43) forms a pocket that allows for specific charge-charge (Arg40) and hydrogen bond
Asp222 is conserved in plant ADHs and proteobacteria orthologs. A sequence similarity network was created using GmPDH1 to identify 318 homologs with a BLAST e-value ≤ 10−5 and visualized in cytoscape. Each circle (node) represents a single TyrA homolog with each line (edge) connecting the nodes representing two proteins that have sequence similarity greater than a given threshold. (a) 100% networks are shown with increasing sequence similarity scores from left to right of ≥ 20, 25, and 30, respectively. (b) 100% network shows that plants (bright green) and brown algae (dark green) are separate, but more close in sequence similarity to bacteria (red) than archaea (blue). The residue at position 222 (corresponding to GMPDH1) is shown for representative TyrA homologs on top of the node that it represents (D, aspartic acid, N, asparagine, and Q, glutamine). P. zucineum (α-proteobacteria ortholog) is from the same genus that contains ADH activity. Algae orthologs fall into the plant group including C. merolae (red algae), A. anophagefferens (brown algae) and C. variabilis (green algae), which is from the same genus that contains ADH activity.
interactions (Ser39, Ser41, and Thr43) with the phosphate group. A similar set of interactions are formed with Arg31, Gln32, and Thr35 in the NADP(H)-dependent SynADH. In contrast, the same loop in the NAD(H)-dependent AaPDH (Asp62, Ile63, and Asn64) is repositioned to fill the corresponding space. This allows for direct interaction with the hydroxyl groups of the adenine ribose of NAD(H) (Fig. 8b). Interestingly, the diphosphate group of NADP(H) adopts a trans-conformation in GmPDH1, where the same cofactor moiety in SynADH and AaPDH are in cis-conformations (Fig. 8b). In SynADH and AaPDH, a 4.5 and 7.7 Å shift in α1 compared to GmPDH, respectively, containing part of the GxGxxG motif, accommodates the cis conformer of cofactor. Thus, the trans-conformation of NADP(H) appear to be a unique feature of GmPDH1 and likely plant TyrAs.

Despite the cofactor binding site variations, each structure maintains the positioning of the ribose and nicotinamide ring relative to a key catalytic histidine (Fig. 5b-c; Fig. 8b-c). The residues that contribute hydrogen bonds to the nicotinamide ribose (Thr73 and Ser101 in GmPDH1; Thr65 and Ser92 in SynADH; Ser99 and Ser126 in AaPDH) are conserved, as is an apolar residue stacking with the nicotinamide ring (Phe28 in GmPDH1; Ile11 in SynADH; Met41 in AaPDH, Fig. 5b-c; Fig. 8b-c). Overall, these interactions position the C4 of the nicotinamide ring in proximity to the conserved catalytic histidine (His124 in GmPDH1; His112 in SynADH; His147 in AaPDH) for the ensuing oxidative decarboxylation reaction (Legrand et al., 2006; Sun et al., 2006; Sun et al., 2009).

Notable differences were found in the architecture of the residues and regions that recognize the side chain of substrates and the Tyr effector (Fig. 8c); part of which reflects the structural variations in the C-terminal dimerization domain (Fig. 8a). SynADH contains an Asn in the 222 position similar to GmPDH1, while AaPDH has Asp255 at the corresponding position. However,
Figure 8. Structural comparison of plant PDH, cyanobacterial ADH, and bacterial PDH. 
(a) Altered dimerization domains. Ribbon diagrams with α-helices shown as cylinders of GmPDH1 (white, PDB: 5T8X), SynADH (purple, PDB: 2F1K), and AaPDH (gold, PDB: 3GGP) with NAD(P)+ (green) in each structure shown as a stick model. (b) Nicotinamide cofactor binding sites of GmPDH1 show trans--conformation, whereas those of SynADH and AaPDH are in cis-conformation. The SynADH structure depicts the diphosphate moiety in two alternate cis-conformations. Coloring of the ribbons and side-chains is the same as in (a). (c) Substrate binding site variations. Residues forming the active site in the GmPDH1-NADP•citrate complex, SynADH•NADP•citrate complex, and the AaPDH•NAD•4-hydroxyphenylpyruvate complex are shown as stick models.

the placement of α-helix adjacent to Asn222 or Asp255 (α11 in SynADH and AaPDH compared to α9 in GmPDH1) varies. This is partly due to a proline residue uniquely present in SynADH and AaPDH but absent in GmPDH1, which kinks the α11 helix to orient the ligand towards the catalytic His. Moreover, the β1e-β1f loop, which is opposite from Asn222 or Asp255, is
condensed in GmPDH1 (Phe127-Trp136) compared to SynADH (Ala115-Leu129) and AaPDH (Ala150-Leu164). These key differences in the active site configuration likely prevent the Asp/Asn residue from being involved in arogenate/prephenate specificity and Tyr inhibition in the microbial structures (Fig. 8c). Thus, microbial TyrA dehydrogenases, which are distantly-related from plant TyrAs (Fig. 7c), have taken different and yet unknown evolutionary paths towards refining substrate specificity as compared to plant TyrAs.

In summary, using a combined phylogenetic and structural approach, we identified the critical residue that controls the substrate specificity and Tyr sensitivity of TyrAs and underlies the functional evolution of alternative Tyr pathways in plants. The high conservation of the Asp residue among all plantae and some microbial TyrA orthologs suggests an ancient evolutionary origin of the ADH Tyr pathway universally present in the plant kingdom today. Structural variations of the ligand side-chain recognition between plant and distant microbial TyrAs further indicate that the Asp-mediated substrate switch is a unique mechanism among plant TyrA orthologs. The identified key residue can now be used to alter Tyr biosynthetic pathways and regulation, as demonstrated in diverse plant TyrAs (Fig. 6), to optimize Tyr availability for the production of its derived natural products, including vitamin E and morphine alkaloid (Millner and Barber, 1984; Strack et al., 2003; Hunter and Cahoon, 2007; Hagel and Facchini, 2013; Gleadow and Møller, 2014).

ACKNOWLEDGEMENTS

We thank Kelly Wallin and Un (Rosia) Schmidt for the cloning of MtnAADH and SlncADH, respectively, and Pierre-Marc Delaux for help with a legume phylogenetic analysis. This work was supported by the National Science Foundation (IOS-1354971 to H.A.M. and MCB-1614539.
to J.M.J.). C.K.H. was supported by the NSF Graduate Research Fellowship Program (DGE-1143954). Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center of the Advanced Photon Source, a national user facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research (DE-AC02-06CH11357).
REFERENCES


Christendat D, Saridakis VC, Turnbull JL (1998) Use of site-directed mutagenesis to identify residues specific for each reaction catalyzed by chorismate mutase-prephenate dehydrogenase from Escherichia coli. Biochem. 37(45), 15703-15712

Christendat D, Turnbull JL (1999) Identifying groups involved in the binding of prephenate to prephenate dehydrogenase from Escherichia coli. Biochem. 38(15), 4782-47893


Chapter 3

Reaction mechanism of prephenate dehydrogenase from the alternate tyrosine biosynthesis pathway in plants

This chapter was previously published as Holland CK and Jez JM (2018) Reaction mechanism of prephenate dehydrogenase from the alternate tyrosine biosynthesis pathway in plants. *ChemBioChem* 19, 1132-6.

Author contributions: C.K.H performed the research; C.K.H. and J.M.J. designed the research, analyzed data, and wrote the paper.
ABSTRACT

Unlike metazoans, plants, bacteria, and fungi retain the enzymatic machinery necessary to synthesize the three aromatic amino acids, L-phenylalanine, L-tyrosine, and L-tryptophan de novo. In legumes, such as soybean, alfalfa, and common bean, prephenate dehydrogenase (PDH) catalyzes the tyrosine-insensitive biosynthesis of 4-hydroxyphenylpyruvate, a precursor to tyrosine. The three-dimensional structure of soybean PDH1 was recently solved in complex with the NADP⁺ cofactor. This structure allowed for the identification of both the cofactor- and ligand-binding sites. Here, we present steady-state kinetic analysis of twenty site-directed active site mutants of the soybean (Glycine max) PDH compared to wild-type. Molecular docking of the substrate, prephenate, into the active site of the enzyme reveals its potential interactions with the active site residues and makes a case for the importance of each residue in substrate recognition and/or catalysis, most likely through transition state stabilization. Overall, these results suggest that the active site of the enzyme is highly sensitive to any changes, as even subtle alterations substantially reduced the catalytic efficiency of the enzyme.
INTRODUCTION

Aromatic amino acids are essential components of proteins and are critical for human health. Despite this importance, the biosynthetic pathways of L-phenylalanine, L-tyrosine, and L-tryptophan in plants have only recently gained attention (Maeda and Dudareva, 2012). Historically, aromatic amino acid biosynthesis in plants was thought to mirror the pathways in microbes, but new discoveries such as the recent discovery of an evolutionarily distinct prephenate dehydrogenase (PDH) in plants highlights how these biosynthesis pathways in plants can differ from those found in microbes (Schenck et al., 2014).

Tyrosine is downstream of the shikimate pathway, which converts phosphoenolpyruvate and D-erythrose 4-phosphate from central carbon metabolism to the aromatic amino acid. The shikimate pathway generates chorismate, the branch point metabolite in aromatic amino acid biosynthesis (Maeda and Dudareva, 2012; Tzin and Galili, 2010; Toghe et al., 2013). Chorismate mutase then converts chorismate to prephenate, the precursor to enzymes in both tyrosine and phenylalanine (Romero et al., 1995; Mobley et al., 1999; Westfall et al., 2014; Kroll and Holland et al., 2017). En route to tyrosine, plants primarily use the arogenate pathway, where prephenate is transaminated to arogenate by prephenate aminotransferase, followed by an oxidative decarboxylation by arogenate dehydrogenase (ADH) (Bommer and Jensen, 1985; Jung et al., 1986; Eberhard et al., 1996; Rippert and Matringe, 2002; Legrand et al., 2006; Cho et al., 2007; Yamada et al., 2008; Rippert et al., 2009; Colquhoun et al., 2010; Graindorge et al., 2010; Maeda et al., 2010; Dal Cin et al., 2011; Maeda et al., 2011; Holland et al., 2018). Some plants contain a PDH that leads to these reactions being performed in the reverse order. PDH oxidatively decarboxylates prephenate to 4-hydroxyphenylpyruvate (4-HPP) and prephenate aminotransferase subsequently converting 4-hydroxyphenylpyruvate (4-HPP) into tyrosine.
(Schenck and Holland et al., 2017). This alternate tyrosine biosynthesis pathway in legumes may provide a means for dedicating cytosolic metabolites for specialized metabolism.

While most fungi and bacteria contain PDH, legumes (*Leguminosae*) are the only land plants found to contain functional dimeric PDH (Stenmark et al., 1974; Fazel and Jensen, 1979). In microbes, PDH exists as either monofunctional or bifunctional homodimeric enzymes, in which PDH is fused to chorismate mutase, 5-enolpyruvylshikimate-3-phosphate (EPSP) synthase, or an aminotransferase (Song et al., 2005). In plants, PDH functions as a homodimer. In microbes, PDHs are allosterically inhibited by tyrosine (Sun et al., 2006; Sun et al., 2009). The plant PDHs are insensitive to regulation by tyrosine, unlike their bacterial homologues, making it an excellent candidate for metabolic engineering in microorganisms to produce tyrosine-derived products (e.g. morphine and vitamin E) (Rubin and Jensen, 1979; Gamborg and Keeley, 1966).

A recent three-dimensional structure of GmPDH1 from soybean (*Glycine max*) has shed light on the mechanisms governing sensitivity to tyrosine and substrate specificity of PDH (Schenck and Holland et al., 2017). A single residue in the active site of PDH (Asn222) was found to mediate binding of both the substrate and the feedback-inhibitor tyrosine, showing that this single residue, which is an aspartate in ADH, is responsible for the evolution of PDH activity and the deregulation of this enzyme in plants. The crystal structure combined with phylogenetic data showed that plants evolved PDH from a plant ADH and did not acquire the same enzyme that microorganisms have.

**RESULTS**

Because this PDH form was only recently discovered, the roles of the active site residues, outside of Asn222, in both binding and catalysis remained to be investigated. Here we use the x-
ray crystal structure of GmPDH1 for computational docking of prephenate in the active site to guide functional analysis of twenty site-directed mutants generated to probe the contributions of active site residues to PDH activity.

Computational docking of prephenate into the three-dimensional structure of the GmPDH1•NADP⁺ complex yielded a solution that places the reactive C4 group of prephenate in proximity to the nicotinamide ring of NADP⁺ and His124 (Figure 1). The ring carboxylate group of C1 is oriented toward Ser101, Gln184, His188, and Thr206 and the pyruvyl moiety positioned to interact with Gln130, Thr131, and Asn222. The prephenate docking solution mimics the observed binding of the dicarboxylic acid citrate in the GmPDH1 active site (Schenck and Holland et al., 2017). In the crystal structure, His124 and His188 contact the α-carboxyl group and γ-hydroxyl group of citrate, respectively, with additional interactions between the ε-carboxyl from Thr131 and Asn222.

Figure 1. Molecular docking of prephenate (grey) into the x-ray crystal structure of GmPDH1 in complex with NADP⁺ (green). Putative protein-ligand interactions are shown as dashed lines. Chains A and B are represented in gold and purple, respectively.
To examine the contributions of the active site residues identified from docking of prephenate into GmPDH1, a series of point mutants (S101A, H124A, H124N, H124Q, Q130A, Q130D, Q130E, Q130S, T131A, T131S, Q184A, Q184D, H188A, H188N, H188Q, T206A, T206S, N222A, and N222Q) were generated, expressed in *E. coli*, and purified for biochemical comparison with wild-type GmPDH1. Wild-type GmPDH1 had a $K_m$ of 270 µM and a turnover rate of 980 min$^{-1}$ with prephenate (Table 1).

### Table 1. Steady-state kinetics parameters for GmPDH1 and its mutants.

<table>
<thead>
<tr>
<th>protein</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat} / K_m$ (M$^{-1}$ s$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GmPDH1</td>
<td>980 ± 76</td>
<td>0.27 ± 0.05</td>
<td>60,500</td>
</tr>
<tr>
<td>S101A</td>
<td>1,100 ± 210</td>
<td>5.2 ± 1.2</td>
<td>3,530</td>
</tr>
<tr>
<td>H124A</td>
<td>5.5 ± 0.2</td>
<td>0.034 ± 0.006</td>
<td>2,700</td>
</tr>
<tr>
<td>H124N</td>
<td>4.5 ± 0.5</td>
<td>0.29 ± 0.07</td>
<td>15</td>
</tr>
<tr>
<td>H124Q</td>
<td>2.4 ± 0.3</td>
<td>0.13 ± 0.06</td>
<td>310</td>
</tr>
<tr>
<td>Q130A</td>
<td>3.3 ± 0.5</td>
<td>0.22 ± 0.09</td>
<td>250</td>
</tr>
<tr>
<td>Q130D</td>
<td>170 ± 8</td>
<td>0.48 ± 0.05</td>
<td>5,900</td>
</tr>
<tr>
<td>Q130E</td>
<td>230 ± 8</td>
<td>0.44 ± 0.03</td>
<td>8,710</td>
</tr>
<tr>
<td>Q130S</td>
<td>28 ± 2</td>
<td>0.13 ± 0.03</td>
<td>3,590</td>
</tr>
<tr>
<td>T131A</td>
<td>7.3 ± 0.4</td>
<td>0.21 ± 0.03</td>
<td>580</td>
</tr>
<tr>
<td>T131S</td>
<td>3.9 ± 0.2</td>
<td>0.11 ± 0.02</td>
<td>590</td>
</tr>
<tr>
<td>Q184A</td>
<td>3.1 ± 0.6</td>
<td>0.37 ± 0.18</td>
<td>140</td>
</tr>
<tr>
<td>H188Q</td>
<td>6.5 ± 0.4</td>
<td>0.07 ± 0.02</td>
<td>1,550</td>
</tr>
<tr>
<td>T206S</td>
<td>1.5 ± 0.1</td>
<td>0.032 ± 0.001</td>
<td>780</td>
</tr>
<tr>
<td>N222Q</td>
<td>30 ± 2</td>
<td>0.49 ± 0.07</td>
<td>1,020</td>
</tr>
</tbody>
</table>

To probe the potential role of His124 in the PDH reaction mechanism, this residue was mutated to an alanine to remove the side-chain and to glutamine and asparagine, which can act as hydrogen-bond donors and acceptors to maintain possible hydrogen bonding patterns. Each of these three mutations led to substantial decreases in catalytic efficiency compared to wild-type (Table 1; Figure 1). These changes primarily result from 22- to 4000-fold reductions in turnover.
rate, which is not surprising considering this residue is well positioned to facilitate hydride transfer from prephenate to NADP⁺.

Substitutions of residues that potentially interact with the prephenate ring carboxylate group (i.e., Ser101, Gln184, His188, and Thr206) also altered the steady-state kinetic parameters of prephenate conversion by GmPDH1 (Table 1). The S101A mutant primarily affected the $K_m$ value for prephenate with a 20-fold increase, which suggests that this residue does not significantly contribute to catalysis. Substitution of Gln184 with acidic residues (Q184D and Q184E) resulted in a loss of detectable PDH activity. These results suggest that the polar, uncharged nature of glutamine is critical for catalysis, as adding a negative charge to the side chain removed all activity. Removal of the side-chain in the Q184A mutant resulted in a 320-fold lower $k_{cat}$ value with a $K_m$ comparable to wild-type. Gln184 is in close proximity to the hydroxyl and carboxylate groups at C1 and C4 on the cyclohexanediene ring (Figure 1). Therefore, this residue likely helps position the ligand for hydride transfer and the ensuing decarboxylation reaction. His188 is positioned in the active site to facilitate the decarboxylation of prephenate following hydride removal (Figure 1). Like the His124 mutant, this residue was mutated to alanine, asparagine, and glutamine. The H188A and H188N mutants were inactive. The H188Q mutant exhibited a 150-fold reduction in turnover rate with modest changes in $K_m$. Thr206 is likely to serve as a hydrogen bond donor to prephenate’s central carboxylate. Mutation of this residue to alanine leads to a loss of catalysis, but the T206S mutant displayed a tremendous loss of activity, with a decrease in turnover from 980 to 1.5 min⁻¹. As serine maintains a polar, uncharged functional group in this position within the active site, our data suggests that the methyl group of the threonine side-chain may be critical for substrate binding and/or orientation.
Gln130, Thr131, and Asn222 are positioned to contact the pyruvyl group of prephenate (Figure 1). Mutation of Gln130 generally had modest effects on reaction kinetics. The Q130D and Q130E mutants had 4- to 6-fold reductions in $k_{\text{cat}}$ and less than 2-fold increases in $K_m$. In contrast, substitution with either alanine or serine resulted in 300- and 35-fold lower turnover rates, respectively. These results suggest that maintaining hydrogen bonding to the substrate from this position of the active site is important for catalysis. In contrast, even modest changes of Thr206 led to either a loss of activity (T206A mutant) or a 650-fold decrease in $k_{\text{cat}}$ (T206S mutant). Asn222 was previously mutated to an aspartate, which was found to introduce sensitivity to tyrosine and to switch the activity from prephenate to arogenate (Schenck and Holland et al., 2017). As such, the complete loss of activity in the N222A mutant was not unexpected. A slight shift in side-chain length with the N222Q mutant resulted in a 30-fold decrease in $k_{\text{cat}}$ and a modest change in $K_m$.

To gain a better understanding of the degree of conservation among PDHs in plants, the GmPDH1 active site residues were compared to the amino acid sequences of PDHs from common bean (*Phaseolus vulgaris*) and alfalfa (*Medicago truncatula*) (Figure 2). Not surprisingly, residues corresponding to Ser101, His124, Gln184, His188, and Thr206 were invariant in all three enzymes. Gln130 and Thr131 in the carboxylate-binding pocket were variable between GmPDH1 and the homologs from bean and alfalfa. Asn222 is conserved in the bean PDH, but is a cysteine in alfalfa, which was previously assayed and found to influence prephenate recognition in the alfalfa enzyme (Schenck and Holland et al., 2017). Gln130 was a valine and Thr131 a serine in the bean and alfalfa enzymes. As noted above, the T131S mutation led to a significant reduction in PDH activity.
Studies of the ADH/TyrA homologs from different microbes (e.g., *Ochrobactrum intermedium* LMG 3301, *Sediminspirochaeta smaragdinae* DSM 11293, and *Methanoseta harundinacea*) indicates that mutation of the aspartate that corresponds to Asn222 of GmPDH1 to an asparagine introduces PDH activity into these proteins (Schenck et al., 2017).

Figure 2. PDH sequence alignment of *Medicago truncatula* (XP_003601003.1; MtPDH), *Glycine max* (NP_001304525.1; GmPDH11), and *Phaseolus vulgaris* (XP_007163590.1; PvPDH). Residues in blue are conserved in all three sequences, while residues in gold are variable.

Sequence alignment of these microbial ADH and GmPDH1 reveals strict conservation of residues corresponding to Ser101, His124, Gln184, and His188 across the homologs (Figure 3). Gln130 is conserved in OiTyrA but is an aspartate in the *Sediminspirochaeta* homolog and a threonine in the *Methanoseta* ADH. Thr131 varies in the microbial enzymes, whereas, Thr206 is conserved except in *Methanoseta*, in which it is a serine. Finally, the residue responsible for recognizing prephenate versus arogenate (Asn222 in GmPDH1), is an aspartate in OiTyrA and SsTyrA and a glutamine in MhTyrA. In MhTyrA, mutation of the glutamine to glutamate introduced ADH activity into the archaeon PDH (Schenck et al., 2017). Thus, as for the plant
enzymes, the identity of residue 222 (GmPDH1 numbering) determines ADH versus PDH activity.

Earlier work on the canonical microbial PDH suggested a concerted mechanism for conversion of prephenate to 4-HPP through the coupling of hydride transfer and oxidative dihydroxylation (Hermes et al., 1984). Subsequent studies identified the role of ionizable groups and implicated a deprotonated histidine ($pK_a = 6.5$) as a key catalytic residue (Turnbull et al., 1991; Christendat et al., 1998; Christendat and Turnbull, 1999). X-ray crystal structures and site-directed mutagenesis of the monofunctional homodimeric chorismate mutase-PDH from the microbe *Aquifex aeolicus* showed that a histidine (corresponding to His124 in GmPDH1) polarizes the substrate C4-hydroxyl group for hydride transfer and that an arginine (not found in GmPDH1) interacts with the C1-carboxylate to facilitate prephenate binding and decarboxylation (Sun et al., 2006; Sun et al., 2009; Christendat et al., 1998). Interestingly, the three-dimensional structure of GmPDH1 revealed that some active site features are retained in the non-canonical PDH (Schenck and Holland et al., 2017).

Mechanistic analyses of the microbial PDH centered on the histidine and arginine proposed to serve as 'catalytic' residues (Sun et al., 2006; Sun et al., 2009; Christendat et al., 1998); however, the stereochemical consequences of the PDH reaction, which leads to conformational changes that flatten the substrate during aromatization and decarboxylation, suggests that other active site residues likely contribute to transition state stabilization during this dynamic process. Site-directed mutagenesis and steady-state kinetic analysis of point mutants targeting each contact residue in the active site suggest a catalytic mechanism for PDH that requires multiple residues for efficient conversion of prephenate to 4-HPP (Figure 4).
Figure 3. PDH sequence alignment of *Glycine max* (NP_001304525.1; GmPDH11), *Ochrobactrum intermedium* LMG 3301 (EEQ93947.1; OiTyrA), *Sediminispirochaeta smaragdinae* DSM 11293 (ADK80640.1; SsTyrA), and *Methanosaeta harundinacea* (KUK94425.1; MhTyrA). Residues in blue are conserved in all sequences, while residues in gold are variable.

**DISCUSSION**

We propose a mechanism for GmPDH1 and the other non-canonical PDH from legumes in which His124 facilitates hydride transfer with His188, Gln184, and Thr206 promoting decarboxylation at C1 (Figure 2). In GmPDH1, critical binding contacts are made between prephenate and active site residues. Gln130 and Thr131 are positioned to stabilize the pyruvyl group carboxylate to orient the carbonyl towards Asn222, which is the selectivity switch between plant PDH and ADH activity, as well as the determinant of sensitivity to the feedback inhibitor
tyrosine (Schenck and Holland et al., 2017). Mutations in these residues had a lesser effect on catalytic efficiency that other positions in the active site (Table 1). As described for the microbial PDH, His124 in GmPDH1 is ideally positioned to polarize the C4-hydroxyl group of prephenate during hydride transfer to NADP⁺ and is the invariant catalytic residue with mutations greatly impacting activity. In addition, Ser101 of GmPDH1 is also conserved and contributes to substrate binding and coordination of the hydrogen-bonding network between His124, cofactor, and prephenate, as proposed for the microbial PDH and suggested by the primary effect on $K_m$ with mutants of this residue (Sun et al., 2006; Sun et al., 2009). In contrast to the microbial enzyme, His188, instead of an arginine, is positioned to interact with the C1-carboxylate group, along with additional contacts provided by Gln184 and Thr206. The effect of mutating His188, Gln184, and Thr206 are consistent with these residues interacting with the C1-carboxylate with His188 as a key residue for this part of the reaction mechanism. We suggest that residues surrounding the bound prephenate stabilize the transition state that leads to formation of the aromatic 4-HPP product and that conformational strain is an important driving force behind the reaction chemistry. This is analogous to the role of active site residues in chorismate mutase and their contributions to stabilization during the reaction that generates prephenate (Westfall et al., 2014; Kroll and Holland et al., 2017).

Understanding the structure-function relationships in the PDH from the alternative tyrosine biosynthesis pathway may prove useful for use of this protein as an herbicide tolerance trait. Increased expression of PDH could potentially allow plants to resist herbicides, such as the triketone herbicides sulcotrione, mesotrione, nitisnone, leptospermone, fluorochloridone, and isoxaflutole, that target 4-HPP dioxygenase (Siehl et al., 2014; Ndikuryayo et al., 2017). Due to the insensitivity of the plant ADH and PDH to product feedback inhibition, unlike their
microbial homologues, these enzymes could be used to increase production of natural products derived from tyrosine and 4-HPP when heterologously expressed in non-legume plants. Additionally, plants expressing ADH and PDH may exhibit resistance to herbicides through increased 4-HPP and plastoquinone synthesis. Plastoquinone, a 4-HPP derivative, is an electron carrier in photosynthesis and a cofactor in carotenoid biosynthesis (Liu et al., 2016). Inhibiting plastoquinone biosynthesis causes photobleaching and lethal phenotypes, which has led to the development of herbicides that target 4-HPP dioxygenase. By overexpressing feedback-insensitive PDH additional 4-HPP could be produced, which could promote resistance to 4-HPP dioxygenase-targeted herbicides. In plants containing the PDH pathway, 4-HPP can be directly synthesized using one enzyme instead of three that would be required by the arogenate pathway.

In summary, our results suggest that the GmPDH1 active site has been highly conserved in the evolution of the function of the enzyme, as small changes in the active site severely disrupt activity. Thus, nature has already engineered the best version of this enzyme, and it is unlikely that the catalytic efficiency of this already proficient enzyme in tyrosine biosynthesis of legumes could be increased further by structural-guided mutagenesis of the active site.
METHODS

Computational Substrate Docking

Prephenate was docked into the active site of the x-ray crystal structure of GmPDH1 complexed with NADP⁺ (PDB: 5WHX; Schenck and Holland et al., 2017) using AutoDockVina (Trott and Olson, 2010) (ver. 1.1.2) in the active site of the GmPDH1 structure (PDB: 5WHX). For docking, a grid box of 30 x 30 x 30 Å and level of exhaustiveness = 8 were used.

Generation of Site-Directed Mutants

Site-directed mutants of GmPDH1 (S101A, H124A, H124N, H124Q, Q130A, Q130D, Q130E, Q130S, T131A, T131S, Q184A, Q184D, Q184E, H188A, H188N, H188Q, T206A, T206S, N222A, N222Q) were generated with the pET-28a-GmPDH1[6] vector as template and oligonucleotides containing mutations using the QuikChange PCR method (Stratagene).

Protein Expression and Purification

Expression constructs were transformed into E. coli Rosetta II (DE3) cells (Novagen). Cells were cultured in Terrific broth until $A_{600nm} \sim 0.6–0.8$ was reached, at which time protein expression was induced by addition of a final concentration of 1 mM isopropyl-β-D-1-thiogalactopyranoside. Cells were grown at 16°C for 16-18 hours and then pelleted by centrifugation and resuspended in CelLytic B (Sigma-Aldrich) (1mL/mg cells). Following sonication, cell debris was removed by centrifugation, and the resulting lysate was passed over a Ni²⁺-nitrilotriacetic acid (Qiagen) column equilibrated in 50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, 10% glycerol (v/v), and 1% Tween-20 (v/v). The column was then washed with 50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, and 10% glycerol (v/v). Bound His-
tagged protein was eluted with 50 mM Tris (pH 8.0), 500 mM NaCl, 250 mM imidazole, and 10% glycerol (v/v). Protein concentration was determined by the Bradford method (Protein Assay, Bio-Rad) with bovine serum albumin as standard. Site-directed mutants of GmPDH1 were expressed and purified using the same methods.

**Kinetic Analysis of Wild-type and Mutant Proteins**

Enzyme assays that monitored the conversion of prephenate to 4-HPP were performed using 1-15 µg of recombinant protein added to a 50-µL reaction mixture of 50 mM Hepes (pH 7.5), 50 mM NaCl, 0.5 mM NADP⁺, and varied concentrations of prephenate (0–1 mM). The reduction of NADP⁺ to NADPH leads to an absorbance increase at A_{340nm} (ε = 6220 M⁻¹ cm⁻¹). Reactions were initiated by addition of the reaction mixture and were conducted at 37ºC using a microplate reader (Tecan) in a 384-well plate. The initial velocity data were fit to the Michaelis-Menten equation using GraphPad Prism (version 7.0c for Mac, GraphPad Software).

**ACKNOWLEDGEMENTS**

This work was supported by the National Science Foundation (MCB-1614539 to J.M.J.). C.K.H. was supported by the NSF Graduate Research Fellowship Program (US) (DGE-1143954).
REFERENCES


Christendat D, Saridakis VC, Turnbull JL (1998) Use of site-directed mutagenesis to identify residues specific for each reaction catalyzed by chorismate mutase-prephenate dehydrogenase from Escherichia coli. Biochem. 37(45), 15703-12

Christendat D, Turnbull JL (1999) Identifying groups involved in the binding of prephenate to prephenate dehydrogenase from Escherichia coli. Biochem. 38(15), 4782-93


Chapter 4

Evolution of Allosteric Regulation in Chorismate Mutases from Early Plants

This chapter was previously published as Kroll K*, Holland CK*, Starks CM, and Jez JM (2017) Evolution of allosteric regulation in chorismate mutases from early plants. *Biochem. J.* 474, 3705-3717

Author contributions: K.K. purified protein, performed enzyme assays, collected x-ray diffraction data, and refined the x-ray crystal structure. C.K.H. collected and analyzed kinetic data and analyzed the structural data. C.M.S. purified chorismate. K.K., C.K.H., and J.M.J contributed to the design of the experiments and analyzed the data. C.K.H. and J.M.J wrote the manuscript with all authors providing feedback.
ABSTRACT

Plants, fungi, and bacteria synthesize the aromatic amino acids: $L$-phenylalanine, $L$-tyrosine, and $L$-tryptophan. Chorismate mutase catalyzes the branch point reaction of phenylalanine and tyrosine biosynthesis to generate prephenate. In *Arabidopsis thaliana*, there are two plastid-localized chorismate mutases that are allosterically regulated (AtCM1 and AtCM3) and one cytosolic isoform (AtCM2) that is unregulated. Previous analysis of plant chorismate mutases suggested that the enzymes from early plants (i.e., bryophytes/moss, lycophytes, and basal angiosperms) formed a clade distinct from the isoforms found in flowering plants; however, no biochemical information on these enzymes is available. To understand the evolution of allosteric regulation in plant chorismate mutases, we analyzed a basal lineage of plant enzymes homologous to AtCM1 based on sequence similarity. The chorismate mutases from the moss/bryophyte *Physcomitrella patens* (PpCM1 and PpCM2), the lycophyte Selaginella moellendorffii (SmCM), and the basal angiosperm *Amborella trichopoda* (AmtCM1 and AmtCM2) were characterized biochemically. Tryptophan was a positive effector for each of the five enzymes examined. Histidine was a weak positive effector for PpCM1 and AmtCM1. Neither tyrosine nor phenylalanine altered the activity of SmCM; however, tyrosine was a negative regulator of the other four enzymes. Phenylalanine down-regulates both moss enzymes and AmtCM2. The 2.0 Å x-ray crystal structure of PpCM1 in complex with the tryptophan identified the allosteric effector site and reveals structural differences between the R- (more active) and T-state (less active) forms of plant chorismate mutases. Molecular insight into the basal plant chorismate mutases guides our understanding of the evolution of allosteric regulation in these enzymes.
INTRODUCTION

All organisms use the three aromatic amino acids, L-phenylalanine, L-tyrosine, and L-tryptophan, for protein synthesis and as precursors of indoles and other aromatic metabolites (Gosset 2009; Maeda and Dudareva, 2012; Tohge et al., 2013). Chorismate, the precursor of aromatic amino acids, vitamins, and the phytohormone salicylic acid, is synthesized in plants, fungi, and bacteria by the shikimic acid pathway, which uses erythrose-4-phosphate from the pentose phosphate pathway and phosphoenolpyruvate from glycolysis as starting molecules (Tzin and Galili et al., 2010). Up to thirty percent of carbon fixed by photosynthesis goes through the shikimate pathway to produce the three aromatic amino acids (Tohge et al., 2013).

Chorismate lies at the metabolic branch point of aromatic amino acid biosynthesis, where chorismate mutase catalyzes the pericyclic Claisen rearrangement of chorismate into prephenate in the first committed step of phenylalanine and tyrosine biosynthesis (Sträter et al., 1997) (Figure 1).

Figure 1. Aromatic amino acid biosynthesis from chorismate and the evolutionary relationships of plant chorismate mutases. (A) Metabolic role of chorismate mutase and the routes to tryptophan, tyrosine, and phenylalanine are shown. (B) Chorismate mutase catalyzes the pericyclic Claisen rearrangement of chorismate to prephenate (adapted from Lee et al., 1995).
Chemically, interactions with charged residues in the active site distort chorismate into a reactive transition state that leads to prephenate (Sträter et al., 1997). Alternatively, anthranilate synthase competes with chorismate mutase for chorismate for the tryptophan biosynthetic pathway (Romero et al., 1995). The two enzymes of this branch point are reciprocally regulated by feedback activation and/or inhibition in higher plants and yeast (Gosset 2009; Maeda and Dudareva, 2012; Tohge et al., 2013; Tzin and Galili, 2010). For example, tryptophan inhibits anthranilate synthase and activates chorismate mutase to avoid buildup of the amino acid (Bernasconi et al., 1994; Westfall et al., 2014).

While chorismate mutase activity has been identified in bacteria, fungi and plants, the sequence similarity and overall structure vary significantly among these enzymes. Recent biochemical and structural characterization of the chorismate mutases from the model plant Arabidopsis thaliana (thale cress) sheds light on the allosteric regulation of these enzymes in plants [8]. Arabidopsis contains three chorismate mutase isoforms (AtCM1-3) (Eberhard et al., 1993; Eberhard et al., 1996; Mobley et al., 1999; Westfall et al., 2014). AtCM1 and AtCM3 are plastid-localized and exhibit allosteric regulation by downstream aromatic amino acids, while AtCM2 is an unregulated cytosolic isoform. Typically, the allosterically regulated chorismate mutases are repressed by tyrosine and phenylalanine and are activated by tryptophan. The aromatic amino acids bind an effector site on the enzyme and regulate the ability of chorismate to bind at the active site for catalysis.

Previous phylogenetic analysis of chorismate mutases across the plant lineage revealed different clades of the enzyme (Westfall et al., 2014). The chorismate mutases from algae grouped with the enzyme from yeast, which is activated by tryptophan and inhibited by tyrosine with phenylalanine having no effect on activity (Schappauf et al., 1998). The plant chorismate
mutases were generally grouped into clades that matched the three isoforms from Arabidopsis. For example, representatives of the cytosolic AtCM2-like clade, which lack the allosteric effector site, and the plastidal, feedback regulated AtCM1-like clade were found across a variety of plants. Unlike the other Arabidopsis isoforms, AtCM3 is unaltered by either phenylalanine or tyrosine but is activated by tryptophan, histidine, and cysteine (Westfall et al., 2014). Phylogenetically, the AtCM3-like clade was found only in the Brassicaceae, which suggests a possible specialized role for this enzyme in those plants. These three groups of chorismate mutases did not include representatives of the basal plant lineages, which formed another distinctive group. This clade contained chorismate mutases from the bryophyte (moss) Physcomitrella patens, the lycophyte Selaginella moellendorfii, and the basal angiosperm Amborella trichopoda, all of which diverged from other flowering plants over 130 million years ago (Amborella Genome Project, 2013). Structural analysis of the AtCM1 effector site revealed two amino acids (Gly149 and Gly213) that are involved in regulating effector binding in the Arabidopsis isoforms (Westfall et al., 2014). Like AtCM1, the chorismate mutases of these basal organisms contain the two glycine residues in these positions that confer allosteric regulation, but also vary in sequence of the site. Given the distinct clade formed by the basal plant chorismate mutases, structural and biochemical analysis of these enzymes could guide our understanding of the evolution of their allosteric regulation.

To understand the origins and molecular basis of allosteric regulation in the plant chorismate mutases, here we examine the steady-state kinetic properties of the enzymes from P. patens (PpCM1 and PpCM2), Am. trichopoda (AmtCM1 and AmtCM2), and S. moellendorfii (SmCM) and determined the x-ray crystal structure of PpCM1 in complex with tryptophan. Biochemical and structural comparisons of the basal plant chorismate mutases to those from Arabidopsis and
yeast provide new information about the molecular evolution of effector regulation control of a key branch point in plant aromatic amino acid biosynthesis.

METHODS

Generation of expression constructs

Codon-optimized coding regions for PpCM1 (XM_001761190.1), PpCM2 (XM_001771774.1), AmtCM1 (XM_006844803.3), AmtCM2 (XM_006857165.3), and SmCM (XM_002962173.1) in the pET-28a bacterial expression vector were synthesized by Genewiz. For each construct, the putative chloroplast transit peptide was identified by sequence comparison with the AtCM1-3 isoforms (Figure 2). The resulting constructs allow for expression of N-terminally hexahistidine-tagged proteins with a thrombin cut site between the His-tag and the first residue of the coding region.

Protein expression and purification

Each pET-28a expression construct was transformed into Escherichia coli Rosetta II (DE3) cells (EMD Millipore), which were cultured in Terrific broth at 37 °C until A600nm of 0.6-0.8 was reached. To induce protein expression, a final concentration of 1 mM isopropyl β-D-1-thiogalactopyranoside was added to the cultures, after which the cells incubated for 16-18 hours at 16 °C. Cells were pelleted by centrifugation (5000 x g) and resuspended in lysis buffer (50mM Tris, pH 8.0, 500 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol, and 1% (v/v) Tween-20). Following sonication, cell debris was removed by centrifugation (23,000 x g) and the resulting supernatant was passed over a Ni2+-nitriloacetic acid (NTA) column (Qiagen)
Figure 2. Sequence comparisons among the three chorismate mutases from *A. thaliana* (AtCM1-3), *S. cerevisiae* (ScCM), and the five basal plant chorismate mutases from *P. patens* (PpCM1 and PpCM2), *Am. trichopoda* (AmtCM1 and AmtCM2), and *S. moellendorffii* (SmCM). Numbering of the sequences is based on the PpCM1 sequence, and helices based on the the PpCM1 structure are labeled in gold above the sequence. Generally conserved residues are colored black, while conserved residues in the active site are red. Variable active site residues are pink. Conserved effector residues are dark blue, and variable effector residues are light blue.
equilibrated in the lysis buffer. The column was then washed with 50 mM Tris, pH 8.0, 500 mM NaCl, 20 mM imidazole, and 10% (v/v) glycerol. Bound His-tagged protein was eluted with 50 mM Tris, pH 8.0, 500 mM NaCl, 250 mM imidazole, and 10% (v/v) glycerol. The His-tag was removed by overnight dialysis at 4 °C using thrombin (1:2000 total protein). Dialyzed protein was loaded onto a mixed benzamidine-sepharose/Ni$^{2+}$-NTA column equilibrated in the above wash buffer to deplete the sample of thrombin and undigested protein. The flow-through of this column was loaded onto a Superdex-75 26/60 HiLoad size-exclusion column (GE Healthcare) equilibrated with 25 mM HEPES, pH 7.5, and 100 mM NaCl using an AktaExplorer FPLC system. Protein concentration was determined by the Bradford method (Protein Assay, Bio-Rad) with bovine serum albumin as a standard.

**Preparation of chorismate**

Chorismate preparation was adapted from a protocol using a chorismate mutase knockout strain of *Aerobacter aerogenes* 62-1 provided by Prof. Tim Wencewicz (Washington U. St. Louis, Chemistry) (Parker and Walsh, 2012). The *A. aerogenes* 62-1 mutant was cultured and extracted as previously described. The pH of the supernatant was adjusted from 6.9 to 2.8 with HCl before loading onto a Phenomenex Strata-C18 E solid phase extraction column that was washed with 60 mL of methanol and equilibrated with 60 mL of water. Once the supernatant was loaded, the column was washed with water and eluted with methanol. The elution was analyzed by HPLC using a 4.6 x 250 mm Fluophase PFP HPLC column at a flow rate of 1.5 mL min$^{-1}$. Solution A consisted of 0.05% (v/v) trifluoroacetic acid and Solution B consisted of 0.05% (v/v) trifluoroacetic acid in acetonitrile. The column was run for 5 min at 90% solution A
and 10% solution B followed by a 15 min gradient to 33% solution B. The elutions were dried overnight by centrifugal evaporation. NMR confirmed the identity of the sample.

**Chorismate mutase assays**

Steady-state kinetic assays monitoring the conversion of chorismate to prephenate were performed as described previously (Westfall et al., 2014). Either 1 µg (AmtCM1) or 5 µg (SmCM1, PpCM1, PpCM2, and AmtCM2) of protein were used in each reaction. Protein was added to the reaction mixture (100 µL) of 50 mM Tris, pH 8.0 and varying concentrations (0 - 8 mM) of chorismate. The disappearance of chorismate leads to an absorbance decrease at $A_{274\text{ nm}}$ ($\varepsilon = 2630 \text{ M}^{-1} \text{ cm}^{-1}$). Initial velocity data were fit to the Michaelis-Menten equation with Prism (GraphPad). To screen for amino acid effectors, each protein was assayed with 50 mM Tris, pH 8, 0.5 mM chorismate, and 2 mM amino acid, as above. Assays showing a 1.5-fold effect (either activation or inhibition) on specific activity were selected for further analysis. EC$_{50}$ values for activation and inhibition by specific amino acids (0-10 mM) were then determined from fitting data to $y = \text{max} / (1 + (E_f/\text{EC}_{50}))$, where max is the maximum observed rate difference and $E_f$ is the effector concentration using Prism.

**X-ray crystallography of *Physcomitrella patens* chorismate mutase 1**

Purified PpCM1 was concentrated to 6 mg mL$^{-1}$ and crystallized using the hanging-drop vapor-diffusion method with a 2-µL drop (1:1 protein:crystallization buffer). Diffraction quality crystals of PpCM1 were obtained at 4 °C with a crystallization buffer of 10% (w/v) polyethylene glycol 4000, 20% (v/v) 2-propanol, and 100 mM HEPES, pH 7.5. Crystals were flash-frozen in liquid nitrogen with mother liquor supplemented with 25% glycerol as a cryoprotectant.
Diffraction data (100 K) was collected at the Argonne National Laboratory Advanced Photon Source Structural Biology Center 19-ID beamline. X-ray data were indexed, scaled, and integrated with HKL3000 (Minor et al., 2006). Molecular replacement implemented in PHASER (Fortelle and de la Bricogne, 1997) used AtCM1 in complex with tyrosine (PDB: 4PPU; Westfall et al., 2014) as a search model to determine the structure of the PpCM1•Trp complex. Iterative rounds of manual model building and refinement, which included translation-libration-screwmotion (TLS) models, used COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010). Data collection and refinement statistics are summarized in Table 3. The final model of the PpCM1•Trp complex included residues Leu73 to Asp316 for chain A, residues Glu69 to Pro152, Arg161 to Glu272, and Arg283 to Asp316 for chain B, two tryptophan effector molecules, one HEPES molecule, and 495 waters. Coordinates and structure factors for the PpCM1•tryptophan complex (PDB: 5W6Y) have been deposited in the RCSB Protein Data Bank.

RESULTS

Functional comparisons of chorismate mutases from early plants

The moss (PpCM1 and PpCM2), lycophyte (SmCM), and basal angiosperm (AmtCM1 and AmtCM2) chorismate mutase isoforms were expressed in E. coli using codon-optimized N-terminally His-tagged constructs. The predicted N-terminal chloroplast signal peptides for PpCM1, PpCM2, AmtCM1, and AmtCM2 were not included in the constructs to aid in protein solubility. SmCM is not predicted to have a chloroplast signal peptide. Each enzyme was purified by Ni^{2+}-affinity and size-exclusion chromatography and was isolated as homodimeric forms (~60 kDa; 28-30 kDa per monomer).
For comparison to the chorismate mutases from Arabidopsis, the steady-state kinetic parameters of purified PpCM1, PpCM2, SmCM, AmtCM1, and AmtCM2 were determined (Table 1). The data indicate that the chorismate mutases from the early plants have turnover rates comparable to AtCM1 ($k_{\text{cat}} = 16$ s$^{-1}$) and AtCM3 ($k_{\text{cat}} = 13$ s$^{-1}$) and within 2-fold of AtCM2 ($k_{\text{cat}} = 39$ s$^{-1}$) (Westfall et al., 2014); however, these rates are 17- to 25-fold lower than the rates reported for the yeast enzyme (Schnappauf et al., 1998). The $K_m$ values for chorismate determined for the five enzymes from early plants are higher than those reported for AtCM1 ($K_m = 0.55$ mM), AtCM2 ($K_m = 0.15$ mM), and AtCM3 ($K_m = 1.10$ mM) (Westfall et al., 2014). Overall, the catalytic efficiencies ($k_{\text{cat}}/K_m$) of the five chorismate mutases from basal plants were comparable to each other and to AtCM3 ($k_{\text{cat}}/K_m = 11,800$ M$^{-1}$ s$^{-1}$), but were 3- to 6-fold lower than AtCM1 ($k_{\text{cat}}/K_m = 29,300$ M$^{-1}$ s$^{-1}$) and 64- to 130-fold less efficient than AtCM2 ($k_{\text{cat}}/K_m = 258,000$ M$^{-1}$ s$^{-1}$).

<table>
<thead>
<tr>
<th>protein</th>
<th>$k_{\text{cat}}$, s$^{-1}$</th>
<th>$K_m$, mM</th>
<th>$k_{\text{cat}}/K_m$, M$^{-1}$ s$^{-1}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>PpCM1</td>
<td>20.7 ± 0.8</td>
<td>2.39 ± 0.24</td>
<td>8,660</td>
</tr>
<tr>
<td>PpCM2</td>
<td>19.5 ± 0.7</td>
<td>2.33 ± 0.45</td>
<td>8,370</td>
</tr>
<tr>
<td>SmCM</td>
<td>18.8 ± 2.9</td>
<td>5.19 ± 1.55</td>
<td>3,620</td>
</tr>
<tr>
<td>AmtCM1</td>
<td>15.0 ± 1.0</td>
<td>3.19 ± 0.48</td>
<td>4,700</td>
</tr>
<tr>
<td>AmtCM2</td>
<td>22.8 ± 5.7</td>
<td>6.79 ± 3.02</td>
<td>3,360</td>
</tr>
</tbody>
</table>

Table 1. Steady-state kinetic parameters of chorismate mutases from basal plants. Enzyme assays were performed using purified protein as described in the methods. Values shown are the mean ± standard deviation for n = 3.

Differential regulation of basal plant chorismate mutases by amino acid effectors

Because the two plastidial chorismate mutases in the model plant Arabidopsis thaliana are regulated by downstream aromatic amino acids, the potential role of all twenty amino acids as effector molecules for each of the five basal chorismate mutases was tested. In this initial screen, assays showing a minimum of 1.5-fold change in activity (either increase or decrease)
Table 2. Summary of amino acid effector regulation in chorismate mutases from basal plants.

Enzyme assays were performed using purified protein as described in the methods. EC₅₀ values for activators (tryptophan and histidine) and inhibition (tyrosine and phenylalanine) with specific activities for apoenzyme and bound forms shown as mean ± standard deviation for n = 3.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>EC₅₀, µM</th>
<th>Apoenzyme specific activity, µmol/min/mg protein</th>
<th>Bound Enzyme specific activity, µmol/min/mg protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tryptophan (activator)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PpCM1</td>
<td>20 ± 9.6</td>
<td>3.35 ± 0.46</td>
<td>10.2 ± 1.2</td>
</tr>
<tr>
<td>PpCM2</td>
<td>19 ± 8.8</td>
<td>2.44 ± 0.22</td>
<td>10.4 ± 0.2</td>
</tr>
<tr>
<td>SmCM</td>
<td>14 ± 5.7</td>
<td>2.81 ± 0.27</td>
<td>4.53 ± 0.82</td>
</tr>
<tr>
<td>AmtCM1</td>
<td>140 ± 37</td>
<td>3.41 ± 0.09</td>
<td>9.71 ± 1.29</td>
</tr>
<tr>
<td>AmtCM2</td>
<td>73 ± 14</td>
<td>3.25 ± 0.39</td>
<td>14.9 ± 0.8</td>
</tr>
<tr>
<td>Histidine (activator)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PpCM1</td>
<td>250 ± 82</td>
<td>2.72 ± 0.23</td>
<td>3.82 ± 0.06</td>
</tr>
<tr>
<td>AmtCM1</td>
<td>2,300 ± 980</td>
<td>2.52 ± 0.50</td>
<td>3.44 ± 0.31</td>
</tr>
<tr>
<td>Tyrosine (inactivator)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PpCM1</td>
<td>100 ± 25</td>
<td>3.28 ± 0.27</td>
<td>0.58 ± 0.12</td>
</tr>
<tr>
<td>PpCM2</td>
<td>240 ± 61</td>
<td>2.48 ± 0.23</td>
<td>1.64 ± 0.29</td>
</tr>
<tr>
<td>AmtCM1</td>
<td>170 ± 64</td>
<td>2.68 ± 0.53</td>
<td>0.84 ± 0.28</td>
</tr>
<tr>
<td>AmtCM2</td>
<td>62 ± 21</td>
<td>3.24 ± 0.10</td>
<td>2.14 ± 0.47</td>
</tr>
<tr>
<td>Phenylalanine (inactivator)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PpCM1</td>
<td>2,600 ± 590</td>
<td>3.02 ± 0.37</td>
<td>1.68 ± 0.28</td>
</tr>
<tr>
<td>PpCM2</td>
<td>7,400 ± 2,000</td>
<td>2.44 ± 0.25</td>
<td>1.40 ± 0.04</td>
</tr>
<tr>
<td>AmtCM2</td>
<td>82 ± 17</td>
<td>2.62 ± 0.11</td>
<td>0.16 ± 0.05</td>
</tr>
</tbody>
</table>

suggested that tryptophan and histidine were potential activators and that tyrosine and phenylalanine were negative effectors; however, the effect of these amino acids varied between each of the five enzymes examined. To determine how these amino acids affect the activity of PpCM1, PpCM2, SmCM, AmtCM1, and AmtCM2, EC₅₀ values for the effectors were determined (Table 2). Tryptophan enhanced the activity of all five chorismate mutases. The PpCM1 and PpCM2 enzymes showed 3.0- and 4.2-fold increases in activity, respectively, with comparable EC₅₀ values of ~20 µM. The lycophte enzyme displayed a 1.6-fold increase in rate
with an EC$_{50}$ = 14 µM. Tryptophan exhibited 2.9- and 4.6-fold increases in maximal rate of the two chorismate mutases from Amborella, which is comparable to the moss enzymes with EC$_{50}$ values 3.5- to 7-fold higher than the moss enzymes. Weak activation of PpCM1 (1.4-fold; EC$_{50}$ = 250 µM) and AmtCM1 (1.4-fold; EC$_{50}$ = 2,300 µM) by histidine was observed and required concentrations roughly 15-fold higher than tryptophan for each enzyme. Given the high EC$_{50}$ values and modest increases in maximal rate, histidine may not be a physiologically relevant effector for PpCM1 and AmtCM1.

As reported for the chorismate mutases from yeast, Arabidopsis, tobacco (*Nicotiana silvestris*), and potato (*Solanum tuberosum*), tyrosine and phenylalanine act as negative effectors of four of the five basal chorismate mutases examined here (Table 2) (Goers and Jensen, 1984; Kuroki and Conn, 1988; Eberhard et al., 1993; Eberhard et al., 1996; Schnappauf et al., 1998; Mobley et al., 1999; Westfall et al., 2014). Neither tyrosine nor phenylalanine had any effect on the enzyme from Selagenella. Both tyrosine and phenylalanine were negative effectors of PpCM1, PpCM2, and AmtCM2. For PpCM1, tyrosine was a more potent effector than phenylalanine, as it decreased activity by 5.2-fold with an EC$_{50}$ = 100 µM compared to a 1.8-fold effect and EC$_{50}$ = 2,600 µM for phenylalanine. Similarly, PpCM2 had EC$_{50}$ values of 240 µM with tyrosine and 7,400 µM with phenylalanine, and both amino acids lowered activity ~1.5-fold. Of the two chorismate mutases from Amborella, AmtCM1 was only inhibited by tyrosine with a 3.2-fold reduction in activity and EC$_{50}$ = 170 µM. In contrast, phenylalanine reduced AmtCM2 activity 16-fold with an EC$_{50}$ = 82 µM and tyrosine was the less effective molecule (1.5-fold lower turnover; EC$_{50}$ = 62 µM). Overall, although the chorismate mutases from *P. patens*, *Am. trichopoda*, and *S. moellendorffii* group phylogenetically and have similar
steady-state kinetic parameters, they display differences in which amino acids act as regulators and the degree to which those effectors either activate or inhibit.

**Overall structure, active site, and effector site of *P. patens* chorismate mutase 1**

To understand the evolution of regulation in the early plant chorismate mutases, the x-ray crystal structure of PpCM1 was determined in complex with the activator tryptophan at 2.0 Å resolution (**Table 3**). PpCM1 crystallized as a homodimer in the asymmetric unit with eight α-helices in each monomer (**Figure 3A**). The overall fold of PpCM1 is similar to that of the AtCM1 with a root mean square deviation of 1.02 Å for 208 Cα-atoms (Westfall et al., 2014). As observed in the Arabidopsis enzyme, four α-helices (α1, α2, α4, and α7) from one monomer interact with the same helices in the adjacent monomer to form the dimer interface. Clear electron density for two tryptophan molecules at the interface between the subunits was observed (**Figure 3B**). Each tryptophan binds in the effector site at the N-terminal ends of α2 and α4 (**Figure 3A**) with the indole ring braced against the two α-helices and the carboxyl group oriented toward the adjacent subunit (**Figure 3C**).

Although none of the plant chorismate mutases have been crystallized with a ligand bound in the catalytic site, structural comparison of PpCM1 to the yeast chorismate mutase co-crystallized with the 8-hydroxy-2-oxa-biocyclo[3.3.1]non-6-ene-3,5-dicarboxylic acid transition state analogue bound in the active site (Sträter et al., 1997). This indicates the position of the active site in a region in PpCM1 surrounded by the C-terminal portion of α4 and α1, α5, and α8 (**Figure 3A**). The yeast enzyme and PpCM1 also share a similar fold with root mean square deviation of 1.65 Å for 214 Cα-atoms. Within the catalytic site (**Figure 4A**), several basic residues— Arg77, Arg214, Lys225, and Lys304—are positioned around the site and based on
**Figure 3. X-ray crystal structure of the PpCM1•tryptophan complex.**
(A) The overall dimeric structure of PpCM1 with each monomer colored gold and white, respectively. The eight $\alpha$-helices forming the monomer fold are labeled in the gold monomer. Tryptophan binding defines the location of the effector sites at the dimer interface, as indicated. The open circle indicates the location of the active site, based on comparison to the structure of yeast chorismate mutase bound to a transition state analog. The bottom panel shows the dimer structure rotated 90° relative to the view in the top panel. (B) Electron density for tryptophan is shown as a 2F$_o$-F$_c$ omit map (1.5 $\sigma$). (C) Surface view of the dimer interface showing the two bound tryptophan molecules as stick models.

comparison to the yeast and bacterial chorismate mutases likely position the negatively-charged dicarboxylate substrate and stabilize the transition state leading to prephenate formation (**Figure 1B**) (Chook et al., 1994; Lee et al., 1995; Rippert et al., 2009).

Sequence comparison of PpCM1, PpCM2, SmCM, AmtCM1, AmtCM2 and the three AtCM isoforms indicates that residues corresponding to Arg77, Arg214, Val221, Lys225, Glu255, Thr303, and Lys304 of PpCM1 are invariant (**Figure 2**). Three other residues, Phe251, Met300, and Gln307, of PpCM1 are retained in PpCM2, AmtCM1, and AtCM1. The *Selaginella* enzyme has the most variations in other positions of the active site with an asparagine, isoleucine, and glutamate replacing the side-chains of Phe251, Met300, and Gln307, respectively. The isoleucine and glutamate substitutions for Met300 and Gln307 are also found
Table 3. Summary of crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th>Crystal</th>
<th>PpCM1•Trp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P22₁2₁2₁</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a = 57.76 Å, b = 87.41 Å, c = 126.0 Å</td>
</tr>
<tr>
<td>Data Collection</td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.979</td>
</tr>
<tr>
<td>Resolution (Å)</td>
<td>37.4 - 2.00 (2.03 - 2.00)</td>
</tr>
<tr>
<td>Reflections (total / unique)</td>
<td>131,509 / 38,326</td>
</tr>
<tr>
<td>Completeness (highest shell)</td>
<td>96.5% (97.0%)</td>
</tr>
<tr>
<td>&lt;I/σ&gt; (highest shell)</td>
<td>10.6 (2.1)</td>
</tr>
<tr>
<td>Rsym (highest shell)</td>
<td>5.9 (41.6%)</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>R&lt;sub&gt;cryst&lt;/sub&gt; / R&lt;sub&gt;free&lt;/sub&gt;</td>
<td>15.5% / 19.6%</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>3,859</td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>495</td>
</tr>
<tr>
<td>No. of ligand atoms</td>
<td>45</td>
</tr>
<tr>
<td>R.m.s. deviation, bond lengths (Å)</td>
<td>0.007</td>
</tr>
<tr>
<td>R.m.s. deviation, bond angles (°)</td>
<td>1.005</td>
</tr>
<tr>
<td>Avg. B-factor (Å&lt;sup&gt;2&lt;/sup&gt;), protein, water, ligand</td>
<td>26.1, 36.8, 27.6</td>
</tr>
<tr>
<td>Stereochemistry: favored, allowed, outlier</td>
<td>96.5, 3.5, 0%</td>
</tr>
</tbody>
</table>

in AtCM2. AmtCM1 and AtCM3 share a tyrosine substitution for Phe251. Overall, the catalytic site is highly conserved across these enzymes.

Binding of two tryptophan molecules in the dimer interface of PpCM1 identified the effector binding site for this enzyme near α2 of one monomer and α4 of the adjacent monomer (Figure 3). The location of the effector site is the same as in the previously described AtCM1 crystal structure, which is consistent with binding of positive and negative regulators at the same site in the plant enzymes (Westfall et al., 2014). Tryptophan binds in the PpCM1 effector site through multiple charge-charge, hydrogen bond, and van der Waals contacts (Figure 4B). Val133 and Arg135 provide surface contacts to the amino acid with the aliphatic portion of the arginine side-chain packing against the indole ring of the ligand. The guanidinium group of Arg134 (3.3 Å) of one monomer and the backbone nitrogens of Gly198 (2.8 Å) and Ser199 (2.9
Figure 4. Active site and allosteric effector sites of PpCM1.

(A) Residues in the PpCM1 active site are shown as stick models. Residues shown were identified by structural comparison to the yeast chorismate mutase in complex with a transition state analog. (B) Structure of the PpCM1 effector site. Amino acids from α2 of one monomer (gold) and amino acids from α4 of the adjacent monomer (white) form the allosteric effector binding site and are shown as stick models. Tryptophan (purple) and two water molecules (red spheres) are bound in the interface. (C) Targeted sequence alignment of allosteric effector site residues of the chorismate mutases from *P. patens* (PpCM1-2), *S. moellendorffii* (SmCM), *Am. trichopoda* (AmtCM1-2), and *A. thaliana* (AtCM1-3). Variable residues are shown in green. The glycine versus aspartate change that defines the activator profiles of AtCM1 versus AtCM3 is indicated in blue. The proline substitution that disrupts allosteric regulation in AtCM2 is shown in red.

Å) from the adjacent monomer provide charge-charge and hydrogen bond interactions with the carboxylate of the effector molecule. The side-chain hydroxyl of Ser199 (3.0 Å) formed a hydrogen bond with the tryptophan carboxylate group. The amine group of the tryptophan interacts through polar contacts with two water molecules (2.5 and 2.8 Å) and the amide oxygen of Asn196 (2.8 Å). The side-chain of Glu141 (3.3 Å), which is located on the α2-α3 loop, helps position the indole ring of the effector by a hydrogen bond to the ring nitrogen.

Comparison of residue in the effector binding site of PpCM1 to the chorismate mutases from moss, Amborella, Selaginella, and Arabidopsis reveals common and varied features (Figure 4C). Key interaction residues, Glu141, Asn196, and Ser199 of PpCM1, are invariant across these enzymes. Val133 is retained in most of the plant enzymes, but is replaced in AmtCM1 (methionine). The overall sequence in the SmCM region is not conserved, and
Arg135, while conserved in all other sequences, is a leucine; additionally, Val133 is an alanine in Selaginella. The residue at position 134 appears to change over evolutionary time, as the bryophyte and lycophyte chorismate mutases share an arginine, but Amborella, a later evolving basal angiosperm, and two of the Arabidopsis isoforms have a glycine at this position with AtCM3 having an aspartate. In ScCM, a phenylalanine is positioned between the PpCM1 residue 134 and 135. The difference between a glycine in AtCM1 versus an aspartate in AtCM3 was shown to account for the differences in positive effector profiles of these enzymes (Westfall et al., 2014). The yeast enzyme contains an asparagine for the residue corresponding to Gly195 of PpCM1 (Figure 2). Multiple variations occur at position 197. The second key glycine residue in the effector site (i.e., Gly198 in PpCM1) is retained in all of the basal plant chorismate mutases. The proline in AtCM2 disrupts the allosteric effect of ligand binding, whereas, the glycine allows for regulatory effects. This structural and sequence information provides insight on the varied effector profiles of the enzymes from basal plants. Moreover, the structure of PpCM1 in complex with an activator and the previously reported structures of AtCM1 in complex with either tyrosine or phenylalanine, which inhibit activity, provide snapshots of the structural changes associated with transition from T- (less active) and R- (more active) states of the plant chorismate mutases.

**DISCUSSION**

Chorismate is the precursor to the three aromatic amino acids and to a number of growth hormones in plants, as well as an essential metabolite (Gosset 2009; Maeda and Dudareva, 2012; Tohge et al., 2013; Tzin and Galili, 2010). At the chorismate branch point, the enzymes anthranilate synthase and chorismate mutase compete for chorismate to generate either
tryptophan or tyrosine and phenylalanine, respectively (Figure 1). Although the three *A. thaliana* chorismate mutases have been investigated (Bernasconi et al., 1994; Mobley et al., 1999; Westfall et al., 2014), questions remained as to the evolution of these highly regulated enzymes in higher plants. Previous phylogenetic analysis suggested that the early plant chorismate mutases from a representative bryophyte (*P. patens*), lycophyte (*S. moellendorffii*), and early angiosperm (*Am. trichopoda*) form a separate clade that might share properties distinct from the different forms of the enzyme found in higher plants. To examine the basal chorismate mutases in the green lineage, we determined the steady-state kinetic properties of five early plant chorismate mutases (Table 1), identified differences in positive and negative effectors for these enzymes (Table 2), and solved the three-dimensional structure of the PpCM1•tryptophan complex (Figures 3 & 4). Our data provides insight on the structure-activity relationship in the effector site and reveals structural changes associated with allosteric transition between activated and inhibited forms of the plant chorismate mutases.

Kinetic analysis of the purified early plant chorismate mutases provided insight into the biochemical and regulatory properties of these enzymes (Table 1). Although PpCM1, PpCm2, SmCM, AmtCM1, and AmtCM2 displayed higher $K_m$ values than isoforms from Arabidopsis, all of the plant chorismate mutases had comparable turnover rates of ~20 s$^{-1}$. This reflects the highly conserved structural similarity of the active site in these enzymes (Figure 4A). The overall sequence and structural similarity of the plant chorismate mutases indicates a common reaction mechanism in which the proper orientation of charged amino acid side-chains and groups on the substrate guide the pericyclic rearrangement that converts chorismate to prephenate (Figure 1B) (Lee et al., 1995). The increased (i.e., mM) $K_m$ values of the *P. patens, S.*
moellendorffii, and Am. trichopoda enzymes may correlate with higher local concentrations of chorismate within the chloroplasts compared to the cytosol of these plants (Rippert et al., 2009).

The basal plant chorismate mutases display a mix of allosteric effector responses that reflect the sequence variation in the regulatory site of these enzymes (Table 2). The structure of the PpCM1•tryptophan complex identifies the location of the effector site (Figures 3 & 4B), which combined with sequence comparisons (Figure 4C), provides insight on differential regulation in this clade of enzymes. Similar to AtCM1, both moss chorismate mutases (PpCM1 and PpCM2) and one of the two basal angiosperm enzymes (AmtCM2) were activated by tryptophan and inhibited by both tyrosine and phenylalanine. These enzymes share the highest level of effector site sequence similarity with AtCM1 and AtCM2; however, the proline replacement for the glycine corresponding to position 198 of PpCM1 disrupts allosteric control (Westfall et al., 2014). Thus, the glycine at this position is a key determinant of allosteric regulation of plant chorismate mutases.

Previous work identified two residues in AtCM1 and AtCM3 that control amino acid effector specificity (Westfall et al., 2014). Mutating AtCM1 Gly149 (Gly134 in PpCM1) to the corresponding aspartate of AtCM3 led to kinetic properties and effector regulation more similar of AtCM3, and the reverse was true of the reciprocal glycine mutation in AtCM3. Interestingly, the glycine is conserved in AmtCM1 and AmtCM2, but is an arginine in the enzyme from P. patens and a serine in the S. moellendorffii enzyme. This residue is also an arginine in the yeast chorismate mutase (Sträter et al., 1997; Schnappauf et al., 1998).

Of the basal plant enzymes, SmCM is most similar to the cytosolic, unregulated AtCM2 isoform, as there are no negative effectors, although tryptophan serves as an activator. The three amino acid substitutions in the effector site of SmCM may disrupt binding of the other aromatic
Figure 5. Allosteric structural changes in plant chorismate mutases.

(A) The effector site of PpCM1 with tryptophan (activator; purple) bound. Side-chains of key residues from the α2 monomer (white) and α4 monomer (gold) are shown as stick models. (B) The effector site of AtCM1 with phenylalanine (inhibitor; rose) bound. Side-chains of key residues from the α2 monomer (white) and α4 monomer (rose) are shown as stick models. In panels A and B, the positions of the left-hand monomers were overlaid to highlight the 3.5 Å shift in positions of the effector molecule and adjacent monomer between the two complexes. (C) Conformational changes in the transition from activated (R-state) to inhibited (T-state) forms of plant chorismate mutases. The three-dimensional structures of one monomer from the PpCM1•tryptophan complex and the AtCM1•phenylalanine complex were overlaid (not shown) with the adjacent monomers shown as ribbon diagrams with cylindrical α-helices (PpCM1 - gold with white labels and AtCM1 - rose with black labels). Space-filling models for the two tryptophans bound in the effector sites of PpCM1 are shown. The active site location, based on comparison to the yeast enzyme is shown as a circle. (D) Summary of allosteric effector action on the reaction catalyzed by PpCM1.

Of the five basal chorismate mutases, AmtCM1 falls between the other members of the clade, as it was activated by tryptophan and inhibited only by tyrosine. These results indicate
that the basal plant chorismate mutases provide functional adaptability as a foundation for evolution of the three distinct clades of chorismate mutases found in later evolving plants.

The x-ray crystal structure of PpCM1 in complex with the activator tryptophan reveals the common overall three-dimensional fold reported for the chorismate mutases from Arabiopsis and yeast (Figure 3) (Sträter et al., 1997; Westfall et al., 2014). Unlike bacterial chorismate mutases, the enzymes from yeast and Arabidopsis are allosterically regulated by aromatic amino acids, although the plant chorismate mutases use a wider range of effector molecules. As reported for the yeast chorismate mutase (Xue et al., 1994; Sträter et al., 1996; Sträter et al., 1997; Schnappauf et al., 1998; Hemstaedt et al., 2002), positive and negative effectors bind at the same sites in the dimer interface in the moss and Arabidopsis enzymes (Figure 3) (Westfall et al., 2014). Comparison of the chorismate mutases from moss and Arabidopsis provides snapshots of the conformational changes associated with the transition from the activated R-state to inactivated T-state forms of the plant enzymes (Figure 5). To highlight the differences in the effector binding site and the structural changes associated with activator versus inhibitor binding, one monomer of the PpCM1•tryptophan complex was aligned with one monomer of the AtCM1•phenylalanine complex (r.m.s.d. = 0.73 Å). As described above, binding of tryptophan identified the effector site at the interface of the two PpCM1 monomers with residues contributed from the α2-α3 loop of one monomer and α4 of the adjacent monomer (Figures 4B & 5A). The indole ring of tryptophan makes van der Waals contacts with the aliphatic portion of the Arg135 side-chain and orients the effector carboxylate group to interact with residues of α4. Likewise, binding of phenylalanine in the effector site of AtCM1 (Figure 5B) shows a similar set of interactions [8]; however, the position of phenylalanine is shifted by 3.3 Å away from the dimer interface relative to where tryptophan binds in PpCM1 (Figure 5A). In the yeast chorismate
mutase, the difference between R-state activation by tryptophan and T-state inhibition by tyrosine was suggested to result from an arginine that did not interact with the indole ring of tryptophan but formed a hydrogen bond to the tyrosine hydroxyl group to move the contact helix (Sträter et al., 1996); however, neither PpCM1 nor AtCM1 retain the arginine. As suggested in later work on the yeast enzyme (Westfall et al., 2014), steric interactions may lead to conformational changes in the plant chorismate mutases.

Comparison of the PpCM1•tryptophan and AtCM1•phenylalanine structures indicates that the 6-membered ring of phenylalanine in the AtCM1 structure binds in approximately the same space as the 5-membered ring of tryptophan in the PpCM1 structure (Figure 5A-B). Steric interactions with the indole ring of tryptophan may provide additional contacts that position the amino acid deeper into the effector site compared to phenylalanine. This would then alter the placement and orientation of the amino acid carboxylate between the two structures, which in turn shifts the position of the N-terminal region of α4 (i.e., Asn196-Ser199). The 3.5 Å shift in placement of a positive versus negative effector at the intersection of α2 and α4 leverages larger structural changes in the α-helices around the active site for the R- versus T-state transition (Figure 5C). For example, binding differences in the effector site result in a ~10° tilt of α4 and translation of active site residues, including Arg214, Val221, Lys225, and Glu255 (Figure 4A), on this helix by 10-12 Å toward the putative substrate binding site. This movement also shifts the positions of other helices (i.e., α1 and α7) that contribute residues to the catalytic site. As suggested for the yeast enzyme, movement of key catalytic residues into the chorismate mutase active site likely positions reactive groups to enhance the pericyclic reaction (Figure 1B) in the R-state with movement away from the active site favored by the T-state. Moreover, molecular dynamics simulations of the yeast chorismate mutase suggest that positive and negative effectors
influence global flexibility of the enzyme, which also mediates the transition between R- and T-
states (Kong et al., 2006). This is likely the case for the plant chorismate mutases, as well.

In seed plants, there are typically two or more chorismate mutase genes in the genome: one
that encodes for a cytosolic isoform and at least one that encodes for a plastid localized protein
(Westfall et al., 2014). Because aromatic amino acid synthesis is localized to the plastid, the
physiological role of the cytosolic chorismate mutases remains to be determined (Rippert et al.,
2009). The enzymes that did not contain a putative chloroplast transit peptide were assumed to
be cytosolic; however, localization data is needed to verify this. A prephenate dehydrogenase in
the cytosol has been recently identified in legumes (Leguminosae) and shown to have evolved
from an arogenate dehydrogenase and neofunctionalized in the cytosol to generate 4-
hydroxyphenylpyruvate for tyrosine biosynthesis (Schenck et al., 2015; Schenck and Holland et
al., 2017). 4-Hydroxyphenylpyruvate is also an important plant metabolite because it is
upstream of tocopherol (vitamin E), plastoquinone, isoquinoline alkaloids, cyanogenic
glycosides, and betalain pigments (Gosset 2009; Maeda and Dudareva, 2012; Tohge et al., 2013;
Tzin and Galili, 2010). Because a cytosolic prephenate dehydrogenase has not been identified in
these early plant lineages, the value of an extra-plastidial pool of prephenate remains to be
determined. In later evolving plants, cytosolic forms of chorismate mutases may provide a
source of prephenate for specialized metabolic pathways (Ishihara et al., 2008; Colquhoun et al.,
2010; Alvarez et al., 2016).

In summary, using a sequence-structure-function analysis, we have characterized chorismate
mutases from several evolutionary ancestors to modern flowering plants. These enzymes
retained some of the characteristics similar to the yeast enzyme, but also share some features
with the well-characterized Arabidopsis isoforms.
ACKNOWLEDGEMENTS

This work was supported by the National Science Foundation (MCB-1614539 to J.M.J.).

K.K. received summer research funding from the Washington University uSTAR Summer Program and a summer undergraduate research fellowship provided through a grant to J.M.J. at Washington University in St. Louis from the Howard Hughes Medical Institute Science Education Program. C.K.H. was supported by the NSF Graduate Research Fellowship Program (DGE-1143954). Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center of the Advanced Photon Source, a national user facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research (DE-AC02-06CH11357).
REFERENCES


Parker JB, Walsh CT (2012) Stereochemical outcome at four stereogenic centers during conversion of prephenate to tetrahydroxyproline by BacABGF in the bacilysin pathway. Biochemistry 51, 5622-5632


Chapter 5

Structural Basis for Substrate Recognition and Inhibition of Prephenate Aminotransferase from Arabidopsis

This chapter was previously published as Holland CK, Berkovich DA, Kohn ML, Maeda H, and Jez JM (2018) Structural basis for substrate recognition and inhibition of prephenate aminotransferase from Arabidopsis. Plant J. 94 (2), 304-14

Author contributions: C.K.H., H.M., and J.M.J. designed experiments; C.K.H. collected structural and kinetics data; C.K.H., D.A.B., and M.L.K. conducted ligand-binding experiments; C.K.H. and J.M.J. analyzed the data and wrote the manuscript with all authors providing input.
ABSTRACT

Aromatic amino acids are protein building blocks and precursors to a number of plant natural products, such as the structural polymer lignin and a variety of medicinally relevant compounds. Plants make tyrosine and phenylalanine by a different pathway from many microbes, which requires prephenate aminotransferase (PAT) as the key enzyme. PAT produces arogenate, the unique and immediate precursor for both tyrosine and phenylalanine in plants, and also has aspartate aminotransferase (AAT) activity. The molecular mechanisms governing the substrate specificity and activation or inhibition of PAT are currently unknown. Here we present the x-ray crystal structures of wild-type and various mutants of PAT from Arabidopsis thaliana (AtPAT). Steady-state kinetic and ligand binding analyses identified key residues, such as Glu108, that are involved in both keto acid and amino acid substrate specificities and likely contributed to the evolution of PAT activity among class Ib AAT enzymes. Structures of AtPAT mutants co-crystalized with either α-ketoglutarate or pyridoxamine 5′-phosphate (PMP) and glutamate further define the molecular mechanisms underlying keto acid and amino acid substrate recognition. Furthermore, cysteine was identified as an inhibitor of PAT from A. thaliana and Antirrhinum majus plants as well as Chlorobium tepidum bacterium, uncovering a potential new effector of PAT.
INTRODUCTION

Throughout evolution, fungi, bacteria, and plants maintained the machinery necessary to synthesize aromatic amino acids de novo (Maeda and Dudareva, 2012; Tzin and Galili, 2010; Toghe et al., 2013). Aromatic amino acids are precursors of multiple physiologically important metabolites, including phytohormones such as auxins and salicylic acid, the structural polymer lignin, flavonoids, vitamin E, and alkaloids (Boerjan et al., 2003; Hunter and Cahoon, 2007; Yu and Jez, 2008; Zhao, 2012; Hagel and Facchini, 2013; Geadow and Møller, 2014; Cook et al., 2016). Additionally, aromatic amino acid biosynthetic pathways are potential targets of antimicrobials and plant herbicides (Schönbrunn et al., 2001; Reichau et al., 2011).

In plants and microbes, phenylalanine and tyrosine biosynthesis both stem from the precursor prephenate, which is formed from chorismate (Romero et al., 1995; Westfall et al., 2014). In this pathway, prephenate dehydrogenase or prephenate dehydratase decarboxylate prephenate into either 4-hydroxyphenylpyruvate (4-HPP) or phenylpyruvate, respectively (Bentley, 1990). An aminotransferase next converts the resulting α-keto acid (i.e., 4-HPP or phenylpyruvate) to an amino acid (i.e., tyrosine or phenylalanine). Plants, and some microbes, use an alternative pathway to synthesize these amino acids via arogenate (Eberhard et al., 1996; Mobley et al., 1999; Colquhoun et al., 2010; Maeda et al., 2010). In this pathway, prephenate aminotransferase (PAT) catalyzes the first step by aminating prephenate into arogenate (Bonner and Jensen, 1985; Graindorge et al., 2010; Maeda et al., 2011) (Figure 1). Decarboxylation of arogenate by either arogenate dehydrogenase or arogenate dehydratase yields either tyrosine or phenylalanine, respectively (Jung et al., 1986; Cho et al., 2007; Rippert et al., 2009; Maeda et al., 2010).
Figure 1. Prephenate aminotransferase catalytic reaction.
Prephenate aminotransferase (PAT) catalyzes the PLP-dependent transfer of an amine from an amine acid (aspartate) to a keto acid substrate (prephenate) resulting in a new amino acid (arogenate) and a new keto acid (oxaloacetate - OAA).

Much of what is known about aromatic amino acid biosynthesis in plants is inferred from microbial investigations (Maeda and Dudareva, 2012), but the modulation of aromatic amino acid metabolism in plants is not well understood. Although PAT activity has been detected in various plants (Rubin and Jensen, 1979; Bonner and Jensen, 1985; Siehl et al., 1986; de la Torre et al., 2014; Kilpatrick et al., 2016), the gene encoding a plant PAT was only recently identified in Arabidopsis (i.e., AtPAT) (Graindorge et al., 2010; Maeda et al., 2011). PAT has a dual role in metabolism and can also function as an aspartate aminotransferase (AAT). The reaction catalyzed by PAT involves the reversible transfer of aspartate's α-amino group to a keto acid substrate, either α-ketoglutarate or prephenate, generating oxaloacetate and either glutamate or arogenate, respectively. Pyridoxal 5’-phosphate (PLP) serves as a cofactor to accept the amino group for transfer to a keto acid. In Arabidopsis, an embryo-lethal mutant identified in a screen was mapped to PAT (Pagnussat et al., 2005). RNAi suppression of PAT in petunia did not affect AAT activity, suggesting that PAT does not significantly contribute to the overall AAT activity in plants but is contributing to aromatic amino acids (Maeda et al., 2011). Similarly, virus-induced gene silencing of PAT in tobacco led to plants that were significantly reduced in stature.
and had decreased lignin content, reduced chlorophyll content, and less free asparagine, while aspartate and glutamate levels remained unchanged compared to wild-type (de la Torre et al., 2014).

Previous biochemical studies of plant PATs reveal a higher specific activity with prephenate versus arogenate (Maeda et al., 2011); however, detailed kinetic and ligand-binding analyses of a plant PAT remain to be performed. A phylolbiochemical investigation identified two residues (Thr84 and Lys169 in AtPAT) that were conserved among PATs, but were absent in AATs (Dornfeld et al., 2014). Corresponding residues are also critical for binding of dicarboxylates in the orthologous enzyme from *Thermus thermophilus*, which belongs to the class Ib AAT family (Nobe et al., 1998; Nakai et al., 1999; Ura et al., 2001; Wang and Maeda, 2017). The T84V and K169V mutants of AtPAT significantly reduce PAT activity (Dornfeld et al., 2014). Interestingly, the T84V/K169V mutant lost function with prephenate, but gained activity with alanine and tryptophan as amino donors and with 4-HPP as the keto acid. Without three-dimensional structural data, the positioning of these residues in the context of the active site remains unclear. Considering how critical PAT is for proper growth and development in plants, surprisingly little is known about the mechanisms governing its reaction chemistry. Here we present a structure/function analysis of AtPAT that yields new insight on keto and amino acid recognition and identifies cysteine as an inhibitor of AtPAT, as well as other plant and microbial homologs.

**METHODS**

Site-directed mutagenesis, protein expression, and protein purification
The AtPAT T84V, K169V, and T84V/K169V mutant expression constructs, along with AmPAT and CtPAT expression constructs, were previously described (Dornfeld et al., 2014). The AtPAT E108K, A168G, K169S, K306A, and R445G mutants were generated using the codon-optimized AtPAT template (see below) and oligonucleotides containing mutations using the QuikChange PCR method (Stratagene). Codon-optimized AtPAT lacking the chloroplast transit peptide was synthesized (Genewiz) and cloned into the pET28a expression vector. Each construct was transformed into E. coli Rosetta2 (DE3) cells (Novagen). Cells were cultured in Terrific broth until $A_{600nm} \sim 0.6–0.8$, at which time protein expression was induced using a final concentration of 1 mM IPTG. After overnight incubation at 16°C, cells were pelleted by centrifugation (5000 x g; 10 min) and resuspended in 50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, 10% glycerol, and 1% Tween-20. Following sonication, cell debris was removed by centrifugation (13,000 x g; 45 min) and the resulting lysate was passed over a Ni$^{2+}$-nitrilotriacetic acid (Qiagen) column equilibrated in the lysis buffer. The column was then washed (50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, and 10% glycerol) and bound His-tagged protein eluted (50 mM Tris (pH 8.0), 500 mM NaCl, 250 mM imidazole, and 10% glycerol). Protein concentration was determined by the Bradford method (Bio-Rad) with bovine serum albumin as standard.

**Protein crystallography**

For crystallization, HIs-tagged protein was further purified using a Superdex-75 26/60 HiLoad ÄKTA FPLC size-exclusion column (GE Healthcare) equilibrated with 25 mM Hepes (pH 7.5) and 100 mM NaCl. Purified wild-type AtPAT and the T84V, T84V/K169V and K306A mutant proteins were concentrated to 12.6, 18.7, 7.0, and 12.5 mg mL$^{-1}$, respectively, and
crystallized using the hanging-drop vapor-diffusion method with a 2-µL drop (1:1 concentrated protein and crystallization buffer). Diffraction quality crystals of the enzyme were obtained at 4°C with a crystallization buffer of 20% PEG-3350, 200 mM magnesium formate, and 1 mM PLP for the wild-type structure; 25% PEG-1500, 100 mM MIB buffer (sodium malonate dibasic monohydrate, imidazole, boric acid; pH 5.0), 1 mM PLP, and 1 mM Glu for the K306A mutant structure; 20% PEG-6000, 100 mM citric acid (pH 5.0), and 3 mM α-ketoglutarate for the T84V mutant; and 25% PEG-1500 and 100 mM MIB buffer (pH 5.0) for the T84V/K169V mutant structure. Crystals were flash-frozen in liquid nitrogen with mother liquor supplemented with 25% glycerol as a cryoprotectant. Diffraction data (100 K) was collected at the Argonne National Laboratory Advanced Photon Source 19-ID beamline. The data were indexed, scaled, and integrated with HKL3000 (Minor et al., 2006). Molecular replacement was implemented in PHASER (McCoy et al., 2007) using a SwissModel-generated dimer of AAT from Phormidium lapideum (PDB: 1J32; 44% amino acid sequence identity) as a search model. The resulting refined AtPAT structure was then used as a search model for each of the AtPAT mutant structures. Iterative rounds of manual model building and refinement used COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010), respectively. Data collection and refinement statistics are summarized in Table 1.

**Kinetic analysis**

PAT activity was determined using a coupled assay system measuring NADH oxidation (Arnold and Parslow, 1995). Assays were performed using 50 mM Hepes (pH 7.5), 25 mM NaCl, 1 mM NADH, 0.11 mM PLP, and 3 units of malate dehydrogenase in a 100 µL volume. Aspartate (0-8 mM) was varied with fixed α-ketoglutarate (3 mM) as the keto-acid. With fixed
aspartate (4 mM), either α-ketoglutarate (0-8 mM) or prephenate (0-2 mM) was varied in assays of wild-type and the T84V mutant. For the E108K and A168G mutants, 6 mM aspartate was used as the fixed substrate. All reactions were initiated by the addition of protein and conducted at 25°C. Initial velocity data were fit to the Michaelis-Menten equation using GraphPad Prism. Inhibition of AtPAT, AmPAT, and CtPAT was analyzed using 3-6 µg of protein with 4 mM aspartate and either 2 mM α-ketoglutarate or 0.5 mM prephenate in the presence of varied (0-5 mM) cysteine. IC\textsubscript{50} values were determined by fitting data to $y = \frac{\text{max}}{1 + (\text{I}_f / \text{IC}_{50})}$, where max is the maximum observed rate difference and $\text{I}_f$ is the inhibitor concentration, in Prism.

**Computational docking of PMP-cysteine forms**

Molecular docking of PMP-cysteine into the AtPAT structure was performed using AutoDock Vina (Trott and Olson, 2010) (ver. 1.1.2). The active site of the structure was used to guide docking with a grid box of 30 x 30 x 30 Å with the exhaustiveness set to 8.

**Amino acid binding analysis**

Binding of aspartate, glutamate, alanine, tryptophan, tyrosine, and cysteine were measured by observing the change in absorbance upon formation of the external aldimine with PLP using a Beckman Coulter DU800 UV/Vis spectrophotometer. Titrations were performed using 200 µg of protein with 0-12.8 mM amino acid in 0.1 M Heps (pH 7.5) at 25°C. Apparent $K_d$ values were calculated by fitting the absorbance data to a reversible two-state model of binding $\Delta A = (\Delta A_{\text{max}}[L]) / (K_d + [L])$, using Prism, where $\Delta A$ is the change in absorbance at a given wavelength in the presence of the amino acid ligand at concentration [L].
RESULTS

Crystal structures of wild-type and K306A mutant AtPAT

To understand the molecular determinants of substrate binding, the three-dimensional structure of AtPAT co-complexed with PLP was solved by molecular replacement and refined to 2.50 Å resolution (Table 1; Figure 2a). AtPAT crystallized as a homodimer. Each monomer of AtPAT consists of 15 α-helices and 9 β-strands divided between two structural domains. The N-terminal domain (Lys115-Leu353) contains two sets of three α-helices surrounding six parallel and one anti-parallel β-strand. Two additional α-helices in the PLP-binding pocket, as well as the α-helix connecting the N- and C-terminal domains, complete the N-terminal domain. The smaller C-terminal domain contains part of the N-terminal region (Ser71-Pro114) and residues Gly354 through Leu469, totaling five α-helices and two β-strands. The N-terminal flexible loop (Ser71-Ser82) and features of the N-terminal domain form the dimer interface.

PLP is bound in the active site of each chain in the wild-type structure with clear electron density (Figure 3a) and defines key residues in the active site (Figure 2b). A catalytic lysine forms a Schiff base with the PLP through its ε-nitrogen, accounting for the active form of the resting enzyme (Malashkevich et al., 1993). In the AtPAT crystal structure, PLP is covalently linked to the ε-nitrogen of Lys306 to form the internal aldime (i.e., Schiff base). Trp193 and Ile274 position the ring of PLP through π-π stacking and van der Waals interactions, respectively. The pyridine ring nitrogen of PLP forms a charge-charge interaction with the side-chain of Asp272, an interaction that is characteristic of aminotransferases, such as PAT and AAT (Toney, 2014). The hydroxyl of Tyr275 and the side-chain amide nitrogen of Asn243 orient the PLP ring hydroxyl group to direct the reactive aldehyde toward the catalytic Lys. The phosphate-
<table>
<thead>
<tr>
<th>Crystal</th>
<th>AtPAT•PLP</th>
<th>AtPAT K306A•PMP•Glu</th>
<th>AtPAT T84V•α-ketoglutarate</th>
<th>AtPAT T84V/K169V•malate•borate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P2₁</td>
<td>P1</td>
<td>C2</td>
<td>C2</td>
</tr>
<tr>
<td>Cell dimensions</td>
<td>a = 80.34 Å, b = 111.6 Å, c = 141.8 Å; b = 90.4°</td>
<td>a = 51.83 Å, b = 60.12 Å, c = 74.8°; a = 76.7°; γ = 83.9°</td>
<td>a = 117.0 Å, b = 67.72 Å, c = 51.56 Å; b = 103.8°</td>
<td>a = 116.5 Å, b = 64.89 Å, c = 51.93 Å; b = 105.9°</td>
</tr>
</tbody>
</table>

**Data collection**

| Wavelength (Å) | 0.98 | 0.98 | 0.98 | 0.98 |
| Resolution range (Å) (highest shell) | 33.3 - 2.50 | 39.4 - 2.10 | 30.7 - 2.00 | 28.0 - 1.40 |
| Reflections (total/unique) | 322,037 / 86,420 | 59,550 / 39,241 | 90,885 / 25,739 | 288,795 / 73,134 |
| Completeness (highest shell) | 99.8% | 90.0% | 96.8% (100.0%) | 99.8% (98.3%) |
| I/σ (highest shell) | 15.4 (1.92) | 11.4 (2.0) | 32.4 (5.6) | 27.3 (1.5) |
| R_sym (highest shell) | 11.1% (93.4%) | 6.1% (27.1%) | 7.5% (25.3%) | 6.8% (58.1%) |

**Model and refinement**

| R cryst/R free | 20.2% / 27.4% | 16.1% / 21.4% | 16.7% / 21.5% | 16.3% / 17.7% |
| No. protein atoms | 18,210 | 6,119 | 3,049 | 3,130 |
| No. waters | 286 | 495 | 147 | 533 |
| No. ligand atoms | 90 | 68 | 10 | 11 |
| Root mean square deviation, bond lengths (Å) | 0.012 | 0.008 | 0.007 | 0.004 |
| Root mean square deviation, bond angles (°) | 1.86 | 1.28 | 1.03 | 0.80 |
| Avg. B-factor (Å²), protein, ligand, water | 87.3, 73.6, 61.2 | 46.8, 72.8, 46.2 | 20.8, 71.7, 34.8 | 29.0, 53.6, 35.4 |
| Stereochemistry: favored, allowed, outlier | 87.6, 8.5, 3.9% | 98.0, 1.8, 0.2% | 98.0, 1.2, 0.8% | 98.0, 2.0, 0.0% |

Table 1. Summary of AtPAT crystallographic statistics.
Figure 2. Three-dimensional structures of AtPAT wild-type and K306A mutant.
(a) Overall structure of AtPAT is shown as a cartoon with each monomer colored white and gold, respectively. Pyridoxal-5’-phosphate (PLP) in each monomer is depicted as a space-filling model (green). Portions of the N- and C-terminal domains are indicated in the gold monomer with the position of the a1 helix noted. (b) AtPAT active site. PLP (green) is covalently-linked as a PLP-Schiff base to Lys306. Protein-ligand interactions are depicted by dashes drawn to specific residues. Tyr132 (gold) is from the adjacent subunit. (c) Comparison of the AtPAT wild-type (gold) and K306A (white) structures. Pyridoxamine-5’-phosphate (PMP) (green) is bound in each monomer of the homodimer. (d) Comparison of the AtPAT (gold) and AtPAT K306A (white) active sites. Side-chains in the active site, along with PMP (green), and glutamate (rose) in the AtPAT K306A structure, are shown as stick models. Binding interactions are highlighted by dashed lines.

binding pocket is formed by polar contacts to Ser305 and the side-chain hydroxyl of Tyr132 from the other monomer. Charge-charge interactions between the PLP phosphate and the side-chain of Arg314 and the backbone amines of Ala168 and Lys169 complete the PLP-binding pocket.

Based on the wild-type structure, Lys306 was postulated to be the catalytic Lys in AtPAT. To capture ligands in the AtPAT active site, Lys306 was mutated to an alanine and the
Figure 3. Electron densities of ligands in AtPAT structures.
Electron density for each ligand is shown as a $2F_o-F_c$ omit map (1.5 $\sigma$). Density for (a) PLP (green), (b) glutamate (rose), (c) $\alpha$-ketoglutarate (purple), and (d) malate (gold) and borate (green) from the wild-type AtPAT, K306A mutant, the T84V mutant, and the T84V/K169V double mutant structures, respectively, are shown.

K306A mutant was crystalized and solved by molecular replacement at 2.1 Å resolution (Table 1). The overall structure is similar to that of the wild-type AtPAT structure (root mean square deviation (r.m.s.d. of 1.35 Å$^2$ for 368 C$_\alpha$-atoms) (Figure 2c). In the AtPAT K306A crystal structure, the alanine substitution of Lys306 prevents Schiff base formation with the cofactor. Instead, both PMP and a ligand, which was modeled as a Glu in the electron density (Figure 3b), were observed in the active site. In the AtPAT K306A active site, the interaction network with PMP is similar to that of wild-type. There is a 5 Å shift in the position of a1 into the active site in the mutant structure compared to wild-type (Figure 2d). Movement of a1 toward the bound ligand potentially allows the hydroxyl of Thr84 to interact with either the keto or amino acid ligands. In the closed state, the side-chain hydroxyl of Thr84 is 3.8 Å away from the backbone amine of the glutamate ligand. The carboxylate side-chain of the glutamate forms charge-charge interactions with the side-chain nitrogen of Lys169 and hydrogen bonds to a water molecule. The backbone oxygen of the amino acid is stabilized by interactions with the side-chain nitrogen of Arg445 and the backbone nitrogen of Gly107. A polar contact between the backbone carboxylate and the side-chain hydroxyl of Tyr275 complete the ligand-protein interactions. Thus, the
structure of the K306A mutant identified the residues critical for ligand binding in the presence of the cofactor.

Figure 4. Sequence comparison of AtPAT and Arabidopsis AATs.
The sequence alignment was generated using full-length sequences in Clustal Omega. Residues highlighted in green are involved in the PLP binding pocket with lighter shades of green representing variant residues and darker green shading representing conserved amino acids. Residues that contact both the PLP and bound ligand, and are conserved, are shaded in gold. Other active site residues are shaded purple with darker and lighter shading indicating conservation and variation, respectively. α-Helices and β-sheets above the sequence alignment were generated using the AtPAT structure as a reference. Residues colored blue are conserved in all six AAT sequences. The thinner line above the N-terminal section represents residues in the predicted localization sequence, while the thicker black lines representing residues in the crystal structure.
Biochemical analysis of active site mutants

To identify residues important for prephenate specificity, active site residues of AtPAT (Figure 2) was compared to closely-related class Ib AAT and more distantly-related class Ia AAT (Figure 4). As expected, the residue corresponding to the catalytic Lys306 is conserved in all of the sequences. Similarly, PLP-binding residues, such as Gly107, Tyr132 and Arg314, are conserved in each protein. In the putative substrate-binding pocket, residues corresponding to Asn243 and Tyr275 of AtPAT are conserved in all AATs. The Ile274 of AtPAT is an Ala in Ia AATs, but is conserved in Ib AATs. Previously identified active site residues Thr84 and Lys169 are only conserved among PAT, and the region containing Thr84 at the start of a1 is absent in the class Ia AATs. Other residues in the PLP- (i.e. Ala168) and substrate-binding sites (i.e., Glu108, and Arg445) vary in the AATs (Figure 4).

<table>
<thead>
<tr>
<th>protein</th>
<th>substrates: varied/fixed</th>
<th>$k_{cat}$ (min$^{-1}$)</th>
<th>$K_m$ (mM)</th>
<th>$k_{cat} / K_m$ (mM$^{-1}$ min$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtPAT</td>
<td>Asp, aKG</td>
<td>560 ± 34</td>
<td>1.9 ± 0.3</td>
<td>295</td>
</tr>
<tr>
<td></td>
<td>Prephenate, Asp</td>
<td>230 ± 11</td>
<td>0.17 ± 0.03</td>
<td>1350</td>
</tr>
<tr>
<td></td>
<td>aKG, Asp</td>
<td>610 ± 25</td>
<td>1.1 ± 0.1</td>
<td>555</td>
</tr>
<tr>
<td>T84V</td>
<td>Asp, aKG</td>
<td>110 ± 11</td>
<td>1.6 ± 0.3</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>Prephenate, Asp</td>
<td>71 ± 3</td>
<td>0.55 ± 0.06</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td>aKG, Asp</td>
<td>110 ± 4</td>
<td>0.46 ± 0.06</td>
<td>239</td>
</tr>
<tr>
<td>E108K</td>
<td>Asp, aKG</td>
<td>500 ± 15</td>
<td>2.8 ± 0.2</td>
<td>179</td>
</tr>
<tr>
<td></td>
<td>Prephenate, Asp</td>
<td>160 ± 8</td>
<td>&gt;2</td>
<td>&lt;80</td>
</tr>
<tr>
<td></td>
<td>aKG, Asp</td>
<td>500 ± 45</td>
<td>1.0 ± 0.3</td>
<td>500</td>
</tr>
<tr>
<td>A168G</td>
<td>Asp, aKG</td>
<td>290 ± 21</td>
<td>3.8 ± 0.5</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>Prephenate, Asp</td>
<td>400 ± 35</td>
<td>0.48 ± 0.10</td>
<td>830</td>
</tr>
<tr>
<td></td>
<td>aKG, Asp</td>
<td>470 ± 76</td>
<td>1.2 ± 0.3</td>
<td>392</td>
</tr>
</tbody>
</table>

Table 2. Summary of steady-state kinetics of wild-type and mutant AtPAT.
Assays were performed and kinetic parameters determined as described in the Methods. Values are averages ± SD (n = 3). aKG = α-ketoglutarate.
To experimentally test the function of these active site residues, the AtPAT E108K, A168G, K169S, and R445G mutants were generated for functional analysis. The mutants together with previously generated AtPAT T84V, K169V, and T84V/K169V mutants were expressed and purified as described for wild-type enzyme. Using a coupled assay system, the steady-state kinetic properties of wild-type and mutant AtPAT were determined (Table 2). With varied aspartate concentrations and a fixed concentration of α-ketoglutarate, the $K_m^{\text{Asp}}$ is 1.9 mM with a turnover rate ($k_{\text{cat}}$) of 560 min$^{-1}$. For the keto acid substrates, assayed at a fixed aspartate concentration, the catalytic efficiency ($k_{\text{cat}}/K_m$) for prephenate is 3-fold higher than that of α-ketoglutarate.

Of the eight AtPAT mutants analyzed, only the T84V, E108K, and A168G proteins were active with Asp as an amino donor (Table 2). The T84V mutant had a $K_m^{\text{Asp}}$ comparable to wild-type enzyme, but the E108K and A168G mutations modestly increased the $K_m$ for this substrate and reduced $k_{\text{cat}}/K_m$ compared to wild-type. The AtPAT T84V and A168G mutants displayed 2-fold higher catalytic efficiency with α-ketoglutarate than with prephenate. In assays where prephenate was varied for the AtPAT E108K mutant, the velocity versus substrate concentration plots were linear to a concentration of 2 mM, indicating that the E108K substitution greatly reduced affinity for prephenate; however, the enzyme was still active with α-ketoglutarate and displayed a catalytic efficiency similar to wild-type. This result indicates that the E108K mutation is functional as an AAT, but has substantially reduced activity with prephenate. Thus, Glu108 plays a critical role in prephenate substrate specificity of PAT.

**AtPAT amino donor binding analysis**
Given that five of the eight mutants analyzed lacked full aminotransferase activity, we evaluated amino acid binding to AtPAT and the site-directed mutants using UV/Vis spectroscopy. Titration of PLP-dependent AATs with amino acids allows for conversion of the pyridoxal (i.e., PLP) form ($\lambda_{\text{max}} = 390$ nm) to the pyridoxamine (i.e., PMP) form ($\lambda_{\text{max}} = 330$ nm), which can be monitored to determine apparent $K_d$ values for the added ligand (Kirsch et al., 1984; Ziak et al., 1990; Karsten and Cook, 2009). For example, AtPAT displays a mixture of E-PLP ($A_{330\text{nm}}$ peak) and E-PMP ($A_{330\text{nm}}$) forms (Figure 5, red line), but with aspartate titration a decrease in PLP signal and increase in PMP signal was observed (Figure 5, purple line).

![Figure 5. Amino acid titration of AtPAT PLP signal.](image)

**Figure 5. Amino acid titration of AtPAT PLP signal.** UV/Vis absorption spectra for the conversion of the PLP form of AtPAT ($A_{390\text{nm}}$) to the PMP form ($A_{330\text{nm}}$) were measured as described in the Methods. Addition of aspartate (0-16 mM) leads to saturation, as indicated by increased $A_{330\text{nm}}$ signal and decreased $A_{390\text{nm}}$ signal.

For comparison of wild-type and mutant AtPAT, titrations of each protein using aspartate, glutamate, alanine, tryptophan, and tyrosine were performed. Wild-type AtPAT and all of the mutants except for the K169S and K306A mutants bound Asp (Table 3). The T84V and K169V mutants had $K_d$ values similar to wild-type, but the E108K and A168G mutants bound aspartate with 3- to 6-fold tighter affinity. In contrast, the R445G and T84V/K169V mutants decreased
affinity up to 3-fold. With titrations using glutamate, only wild-type, E108K, and A168G AtPAT showed changes in the spectra with the E108K substitution leading to a 22-fold decrease in binding affinity compared to AtPAT. The T84V, A168G, K169V, K169S, R445G, and T84V/K169V PAT mutants bound alanine to varying degrees, whereas, AtPAT wild-type did not bind alanine. In earlier work (Dornfeld et al., 2014), the amino donor for AtPAT was changed from aspartate to tryptophan in the T84V/K169V double mutant.

To analyze this specificity change further, the binding of tryptophan and tyrosine was assessed (Table 3). Surprisingly, the wild-type enzyme had the highest binding affinity (3.6 mM) for tryptophan, although the $K_D$ was still higher than that of either aspartate or glutamate. Except for the A168G, K306A, and R445G mutants, all of the AtPAT variants also retained tryptophan binding, but with decreased affinities. Using tyrosine, only the E108K and K169V mutants exhibited binding. Overall, titration of wild-type and mutant AtPAT with selected amino acids highlights the flexibility of ligand binding and the contributions of individual residues to amino donor binding.

<table>
<thead>
<tr>
<th>protein</th>
<th>$K_d^{\text{Asp}}$ (mM)</th>
<th>$K_d^{\text{Glu}}$ (mM)</th>
<th>$K_d^{\text{Ala}}$ (mM)</th>
<th>$K_d^{\text{Trp}}$ (mM)</th>
<th>$K_d^{\text{Tyr}}$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtPAT</td>
<td>1.5 ± 0.2</td>
<td>0.19 ± 0.03</td>
<td>--</td>
<td>3.6 ± 0.6</td>
<td>--</td>
</tr>
<tr>
<td>T84V</td>
<td>1.3 ± 0.1</td>
<td>--</td>
<td>1.0 ± 0.4</td>
<td>19.0 ± 2.4</td>
<td>--</td>
</tr>
<tr>
<td>T84V/K169V</td>
<td>5.1 ± 1.6</td>
<td>--</td>
<td>0.14 ± 0.08</td>
<td>22.0 ± 4.0</td>
<td>--</td>
</tr>
<tr>
<td>E108K</td>
<td>0.53 ± 0.05</td>
<td>4.2 ± 0.1</td>
<td>--</td>
<td>15.0 ± 2.0</td>
<td>1.1 ± 0.2</td>
</tr>
<tr>
<td>A168G</td>
<td>0.24 ± 0.04</td>
<td>0.29 ± 0.04</td>
<td>10.0 ± 1.4</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>K169V</td>
<td>1.3 ± 0.6</td>
<td>--</td>
<td>0.22 ± 0.05</td>
<td>24.0 ± 2.8</td>
<td>--</td>
</tr>
<tr>
<td>K169S</td>
<td>--</td>
<td>--</td>
<td>0.42 ± 0.15</td>
<td>10.0 ± 1.9</td>
<td>7.4 ± 1.0</td>
</tr>
<tr>
<td>K306A</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>R445G</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>

Table 3. Summary of $K_D$ values for selected amino acids binding to wild-type and mutant AtPAT.

UV/Vis absorption spectra for the conversion of the PLP to PMP form of each protein during amino acid titrations were measured and $K_d$ values calculated as described in the Methods. Values are averages ± SD (n = 3). Dashed lines indicate no significant changes were observed in the titrations.
Cysteine inhibits plant and microbial PATs

Because of the relaxed specificity in amino acid binding with AtPAT, the effect of all 20 proteogenic amino acids was analyzed to determine if any act as either positive or negative effectors of PAT activity (Figure 6). Surprisingly, cysteine reduced AtPAT activity with \( \alpha \)-ketoglutarate (IC\(_{50} \) = 0.56 mM) and with prephenate (IC\(_{50} \) = 0.75 mM) (Figure 7a). To our knowledge, metabolite effectors of a PAT have not been previously reported from either a microbial or plant source.

**Figure 6. Effect of amino acids on AtPAT activity.**
AtPAT activity was assayed as described in the methods using \( \alpha \)-ketoglutarate as the keto acid and aspartate as the amino donor in the absence (D only) or presence of 5 mM of the indicated amino acid (single letter abbreviations). Specific activities are shown as averages ± SD (n = 3).

Earlier work suggests the evolutionary basis for PAT acquisition in plants as lateral gene transfer from Chlorobi/Bacteroidetes (Dornfeld et al., 2014). To test if cysteine also inhibited PAT homologues in other plants and in microbes, two previously characterized PAT from *Antirrhinum majus* (snapdragon) and *Chlorobium tepidum* (a Gram-negative, thermophilic green sulfur bacterium) were assayed for potential inhibition (Figure 7a). Cysteine inhibited both
AmPAT and CtPAT with similar IC₅₀ values (1.3 mM). All of the active site residues of AtPAT are highly conserved in the *A. majus* and *C. tepidum* enzymes except for Ala168 and Ile274, which are Gly106 in CtPAT and Met211 in AmPAT, respectively.

**Figure 7. Inhibition of plant and microbial PAT by cysteine.**

(a) Inhibitory concentrations of prephenate aminotransferases with cysteine from Arabidopsis (AtPAT), *A. majus* (AmPAT), and *C. tepidum* (CtPAT) were determined. Values are averages ± SD (n = 3). (b) Molecular docking of non-cyclized PMP-cysteine (green) into the AtPAT active site. The surface of the AtPAT active is shown with key residues indicated. The docking solution for a non-cyclized PMP-cysteine is shown as a stick model. (c) Molecular docking of cyclized PMP-cysteine (green) into the AtPAT active site.

<table>
<thead>
<tr>
<th>Protein, substrate</th>
<th>IC₅₀ (mM Cys)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtPAT, aKG</td>
<td>0.56 ± 0.06</td>
</tr>
<tr>
<td>AtPAT, Prephenate</td>
<td>0.75 ± 0.15</td>
</tr>
<tr>
<td>AmPAT, Prephenate</td>
<td>1.3 ± 0.30</td>
</tr>
<tr>
<td>CtPAT, Prephenate</td>
<td>1.3 ± 0.40</td>
</tr>
</tbody>
</table>

Table 4. Summary of $K_d$ values for cysteine binding to wild-type and mutant PAT.

UV/Vis absorption spectra for the conversion of the PLP to PMP form of each protein during amino acid titrations were measured and $K_d$ values calculated as described in the Methods. The PAT from *A. majus* (AmPAT), and *C. tepidum* (CtPAT) were also examined. Values are averages ± SD (n = 3). Dashed lines indicate no significant changes were observed in the titrations.

<table>
<thead>
<tr>
<th>protein</th>
<th>$K_d^{\text{Cys}}$ (mM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AtPAT</td>
<td>0.61 ± 0.10</td>
</tr>
<tr>
<td>T84V</td>
<td>0.51 ± 0.13</td>
</tr>
<tr>
<td>T84V/K169V</td>
<td>0.48 ± 0.12</td>
</tr>
<tr>
<td>E108K</td>
<td>0.61 ± 0.08</td>
</tr>
<tr>
<td>A168G</td>
<td>0.70 ± 0.04</td>
</tr>
<tr>
<td>K169V</td>
<td>--</td>
</tr>
<tr>
<td>K169S</td>
<td>0.61 ± 0.04</td>
</tr>
<tr>
<td>K306A</td>
<td>0.96 ± 0.33</td>
</tr>
<tr>
<td>R445G</td>
<td>--</td>
</tr>
<tr>
<td>AmPAT</td>
<td>1.30 ± 0.15</td>
</tr>
<tr>
<td>CtPAT</td>
<td>0.90 ± 0.08</td>
</tr>
</tbody>
</table>
To determine if cysteine reacted with PLP, wild-type and mutant AtPAT were titrated with cysteine and the shift in spectra from PLP to PMP forms measured. With wild-type AtPAT, changes in the spectra between forms were observed and yielded a $K_d = 0.61$ mM (Table 4). Most of the AtPAT mutants bound cysteine with $K_d$ values similar to wild-type. The T84V mutant displayed a 3-fold weaker binding affinity. Notably, neither the K169V mutant nor R445G mutant exhibited changes in signal upon addition of the amino acid. Binding of cysteine to CtPAT AmPAT displayed $K_d$ values similar to AtPAT. The $K_d$ values for cysteine with each enzyme (Table 4) were also comparable to the IC$_{50}$ values determined from enzymatic assays (Figure 7a).

**Crystal structures of AtPAT T84V and T84V/K169V mutants**

Sequence comparisons of known functional PATs and biochemical analysis identified Thr84 and Lys169 of AtPAT as determinants of PAT versus AAT activity (Dornfeld et al., 2014). To rationalize the biochemical data presented here and in the earlier study, the x-ray crystal structures of AtPAT T84V in complex with α-ketoglutarate and AtPAT T84V/K169V with malate and borate bound (both from the crystallization buffer) were determined to 2.0 Å and 1.4 Å resolution, respectively (Table 1).

The T84V mutant is structurally similar to wild-type enzyme (r.m.s.d. = 0.66 Å$^2$ for 350 C$_\alpha$-atoms) (Figure 8a). Clear electron density for α-ketoglutarate (Figure 3c) was observed with the ligand making similar contacts as described for glutamate in the AtPAT K306A structure (Figure 8b). In the absence of cofactor, the phosphate-binding residue Arg314 shifts to partially fill the site where PLP would bind. The hydroxyl on the β-carbon of the ligand interacts with the ring nitrogen of Trp193. The α-carbon of α-ketoglutarate forms ionic interactions with the side-
Figure 8. Three-dimensional structures of the AtPAT T84V and AtPAT T84V/K169V mutants. (a) An overlay of the AtPAT (white), AtPAT T84V (blue), and AtPAT T84V/K169V (purple) crystal structures shows the degree of conservation among the structures. PLP (green) from the wild-type structure is highlighted for reference. (b) Active site comparison of the wild-type AtPAT and AtPAT T84V structures shows the differences between the open (AtPAT) and closed (AtPAT T84V) form structures when α-ketoglutarate (purple), a keto acid substrate, is bound. (c) Active site comparison of the wild-type AtPAT and AtPAT T84V/K169V shows the binding interactions that occur in a closed form structure. Malate (gold) and borate (green) are bound. Dashed lines indicate binding interactions.

Chain nitrogens of Arg445 and Asn243, along with the main-chain amide of Gly107. The α1 helix, where residue 84 is found, moves 5 Å toward the catalytic lysine; however, this could be an artifact of the absence of the cofactor and the presence of a small ligand.

The AtPAT T84V/K169V structure is more similar to the AtPAT T84V structure (r.m.s.d. = 0.358 Å² for 384 Cα-atoms) than to wild-type (r.m.s.d. = 0.753 Å² for 346 Cα-atoms). Like the AtPAT T84V structure, α1 shifts into the active site (Figure 8c). The positioning of the α4 helix that contains the K169V mutation remains unchanged. Malate and borate were bound in the AtPAT double mutant active site (Figure 8d). The carboxylates of malate form hydrogen bond interactions with Tyr275 and charge-charge interactions with Gly107, Trp193, Asn243, and Arg445. Borate binds in the PLP-binding pocket and forms a hydrogen bond with Asp272. The position of malate in the T84V/K169V structure approximates the binding of glutamate in the AtPAT K306A structure and α-ketoglutarate in the AtPAT T84V structure.
Both structures show that mutations reducing enzymatic activity with prephenate do not alter the
global architecture of the active site. The AtPAT T84V and T84V/K169V mutant structures
suggest that substitution of side-chains that provide hydrogen bond and/or ionic interactions with
apolar residues leads to altered substrate preference, potentially from the loss of key interactions
that allow for prephenate binding.

DISCUSSION

When plants moved onto land, the need for lignin, which is the second most abundant
organic compound on Earth and makes up roughly 30% of plant biomass, intensified the
evolutionary pressures to maintain the aromatic amino acid biosynthesis pathways, as
phenylalanine is a precursor to lignin biosynthesis (Boerjan et al., 2003). Similarly, aromatic
amino acids became critical for phenylpropanoids needed for protection against UV light,
herbivore and pathogen defense; for floral pigmentation; and the synthesis of aroma compounds
(Dixon and Paiva 1995; Corea et al., 2012). Historically, PAT activity had been detected in
plants, but only recently was the gene encoding PAT identified (Graindorge et al., 2010; Maeda
et al., 2011). Here we examined the molecular mechanisms for substrate recognition and
inhibition of the plant PAT.

Using x-ray crystallography, we captured AtPAT in an open active site conformation with
PLP linked to the catalytic lysine as a Schiff base (Figure 2). In the active site, multiple residues
contact PLP; however, two residues are key for the aminotransferase chemistry. Lys306, which
formed an internal aldimine with the bound PLP, clearly serves as the Schiff base during the
reaction sequence (Figure 9, E+PLP and E-PLP steps). In addition, the interaction between
Asp272 and the pyridine nitrogen is a hallmark of aminotransferases, as this interaction
maintains a protonated form of the PLP that allows for the ensuing aminotransferase reaction and not other types of chemistry, such as decarboxylation, condensation, and elimination reactions (Toney, 2014).

Although PAT and AAT share common chemistry, these enzymes differ in substrate preference. Amino acid sequence comparison of AtPAT with other AATs highlights conservation of the PLP-binding site (Figure 4), which is consistent with the common reaction chemistry of these enzymes. The crystal structure of the AtPAT K306A mutant in complex with

Figure 9. Summary of AtPAT reaction sequence and roles of active site residues.
The chemical reaction mechanism of AtPAT is shown to highlight key intermediates. Multiple arrows are used to simplify mechanistic details. Residues involved in PLP (red), aspartate (blue), and prephenate (black) interactions are shown with thin and thick dashed lines indicating hydrogen bond and charge-charge interactions respectively. Interactions shown in the E + PLP step are maintained throughout the reaction sequence. Interactions shown for aspartate are also conserved in prephenate binding.

PMP and glutamate provides insight on the substrate binding site (Figure 2). In this structure, the amino acid backbone of the ligand is oriented toward Gly107, Glu108, and Arg445, which provides a counter ion to the backbone carboxylate group. This positions the amino acid amine toward the PMP in an orientation that would facilitate formation of an external aldimine in the
reaction sequence (Figure 9, E+PLP-Asp step). In addition, Lys169 forms an ionic interaction with the side-chain carboxylate of the ligand and movement of the a1 helix into the active site places Thr84 opposite the side-chain carbon to provide van der Waals contacts.

The reaction mechanism of aminotransferases involves a ping-pong type reaction sequence and requires the use of a shared binding site for both amino acid and keto acid substrates. Although the AtPAT T84V mutant was crystallized without cofactor, the position of α-ketoglutarate in the active site (Figure 8b) reveals a similar set of protein-ligand interactions compared to the AtPAT K306A•PMP•glutamate complex (Figure 2c-d). The α-ketoglutarate backbone carboxylate is oriented toward Gly107, Glu108, and Arg445, as observed with glutamate bound. Although Lys169 does not directly interact with α-ketoglutarate, it is possible that with prephenate bound that this residue provides additional contacts with the substrate's bridging carboxylate group. Moreover, similar interactions with malate were observed in the AtPAT T84V/K169V structure (Figure 8c).

Biochemical analyses of mutants targeting active site residues identified key residues and molecular mechanisms responsible for reaction chemistry and substrate preference of PAT (Tables 2-3). Mutations of Lys306, Arg445, and Lys169 yielded inactive enzymes and the K306A and R445G mutants did not bind amino acid ligands. The x-ray crystal structure of the K306A mutant and the resulting effect on enzymatic activity and ligand binding is consistent with a role for this Lys as the Schiff base in the reaction mechanism, as discussed above. The position of Arg445 in the active site suggests that this residue serves as a counter-ion to the backbone carboxylate group of the amino acid and keto acid substrates (Figs. 2c-d & 8b). As no signal change was observed in amino acid titrations, elimination of a key ionic interaction provided by Arg445 likely alters binding affinity to prevent catalysis. Similarly, the K169S and
K169V mutants were inactive; however, the K169S mutant lost aspartate binding with the K169V mutant retaining a binding affinity similar to wild-type (Table 3). Overall, these results indicate that Lys169 and Arg445 are key for substrate binding in AtPAT. We propose that these residues facilitate both amino acid and keto acid binding and formation of the external aldimine form of AtPAT, which that allows for the donor amine from aspartate to react with the prephenate to form arogenate (Figure 9).

Previous studies identified Thr84 and Lys169 of AtPAT as determinants of PAT activity (Dornfeld et al., 2014); however, the molecular basis for substrate preference was not addressed. To evaluate the contributions of these residues to substrate preference, we re-examined the effect of mutating Thr84 and Lys169 on PAT activity and ligand binding (Tables 2-3) and determined the crystal structures of the AtPAT T84V and T84V/K169V mutants (Figure 8). Previous work indicated that the T84V mutation reduced PAT activity and that the double mutant abolished both PAT and AAT activity to yield an enzyme that used tryptophan as an amine donor and 4-HPP as a keto acid. Consistent with that study, the T84V mutant displayed reduced catalytic efficiency compared to AtPAT with the largest effect (10-fold) observed with prephenate (Table 2). In amino acid titrations, the $K_d$ of this mutant for aspartate was comparable to wild-type; however, this mutation introduced alanine binding and decreased affinity for tryptophan (Table 3). The active sites of the wild-type and T84V AtPAT proteins are structurally similar (Figure 8b) with only the loss of a potential hydrogen bond donor affecting the $K_m$ for prephenate. As mentioned above, the K169 mutations lost PAT activity with aspartate (Table 2). Examination of amino acid binding showed that the K169V mutant, like the T84V mutant, could bind alanine. The K169S mutant was unable to bind aspartate, which explains why the enzyme was inactive in the coupled assay. Interestingly, the K169S mutant had a higher affinity for tyrosine and
tryptophan. This data suggests that the T84V and K169V mutants more readily accommodate smaller ligands, such as α-ketoglutarate, alanine, and aspartate, but filter out larger molecules like prephenate and tryptophan. This is also consistent with conservation of these residues as valine in the Arabidopsis AAT. Differences that lead to the new activity of the double mutant protein are difficult to assess; however, the presence of side-chains that provide either hydrogen bond (Thr84) or charge-charge (Lys169) interactions are key for PAT function.

Structure-guided sequence comparisons (Figure 4) also identified other changes in the active sites, including substitution of Ala168 and Glu108 of AtPAT with glycine and lysine, respectively, of AATs. Kinetic and ligand binding analyses of the AtPAT A168G mutant indicates that a change at this position modestly impacts biochemical function (Tables 2-3). In contrast, the E108K mutant altered substrate preference of AtPAT (Table 2). The resulting AtPAT E108K mutant still used prephenate, which suggests that multiple residues contribute to this gain-of-function mutation in an AAT. AtPAT displays a 2-fold preference for prephenate versus α-ketoglutarate. Mutation of Glu108 to a lysine shifts the $k_{cat}/K_m$ for α-ketoglutarate more than 8-fold compared to prephenate. It should be noted that the AAT versus PAT activity of the E108K mutant is likely greater than this, because the enzyme did not display saturation kinetics with prephenate. Moreover, ligand-binding analysis suggests that the E108K mutant is more promiscuous than AtPAT (Table 3). The E108K mutant has a higher affinity for aspartate and a lower affinity for glutamate, and can bind tyrosine and tryptophan. In the crystal structures of AtPAT, the side-chain of Glu108 is oriented away from the active site with protein-ligand contacts provided by the peptide backbone. It is possible that changes in the side-chain structure may reposition the peptide backbone and influence how ligands fit in the active site. Relaxed ligand binding preference would be a step toward altering substrate selection. Interestingly,
Glu108 is also conserved among class Ib AAT enzymes not having PAT activity (Dornfeld et al., 2014). Thus, the presence of glutamate at position 108 among the class Ib AATs, but not class Ia AATs, might have contributed to the evolution of PAT function.

The ability of AtPAT to bind a range of amino acids led us to examine if these molecules could serve as potential effectors of the enzyme and identified cysteine as an inhibitor of AtPAT (Figs. 6 & 8a). Ligand binding analysis indicates that cysteine forms an external aldimine with PLP (Table 4). This suggests that the Cys-PMP aldimine likely prevents the keto acid substrate from reacting with PLP. Two of the active site mutants (K169V and R445G) did not form an external aldimine with cysteine, which suggests that these key substrate interaction residues are also important for binding of this inhibitor molecule (Table 4). Previous work on PLP-dependent aspartate decarboxylases elucidated a potential mechanism of cysteine binding that leads to irreversible inhibition (Liu et al., 2013). In these enzymes, cysteine enters the active site and reacts with PLP to form a cysteine-PMP aldimine through nucleophilic addition by the sulphydryl group and subsequent cyclization. In PAT, we presume that cysteine forms a permanent external aldimine that sterically blocks the active site, but further studies are needed to confirm the irreversible inhibition. To visualize this possible reaction in the AtPAT active site, molecular docking was used to predict the binding of PMP-cysteine in both non-cyclic and cyclic forms (Figure 8b-c). Either molecule fit in the AtPAT active site and is proximal to substrate binding and catalytic residues.

PATs from other organisms, including snapdragon and a gram-negative, thermophilic green sulfur bacterium from Chlorobi, were also inhibited by cysteine (Figure 8a; Table 4). While it was maybe not surprising that another plant PAT is sensitive to Cys, finding that PAT in a bacterium is inhibited by cysteine was unexpected. This finding also suggests that either the
ancestral enzyme that was acquired in plants was sensitive to cysteine or that both plant and microbial enzymes have independently evolved sensitivity to the same amino acid. Cysteine may play a role as an effector molecule across different metabolic pathways in plants, as this amino acid has been identified as an effector for several other plant amino acid biosynthesis enzymes, including aspartate kinase (Curien et al., 2005), homoserine dehydrogenase (Rognes et al., 2003; Paris et al., 2003) 3-phosphoglycerate dehydrogenase (Okamura and Hirai, 2017), and chorismate mutase (Westfall et al., 2014). Given that plants need to coordinate a variety of biochemical pathways connected to carbon, nitrogen, and sulfur metabolism (Shin et al., 2011), the use of a biochemical effector to modulate activity across those pathways may contribute to flux between tyrosine/phenylalanine biosynthesis and tryptophan/chorismate-derived metabolite biosynthesis.

ACCESSION NUMBERS

Coordinates and structure factors for AtPAT•PLP (PDB: 5WMH), AtPAT K306A•PMP•glutamate complex (PDB: 5WML), AtPAT T84V•α-ketoglutarate complex (PDB: 5WMI), and AtPAT T84V/K169V•malate•borate complex (PDB: 5WMK) were deposited in the RCSB Protein Data Bank. Sequences used for alignments are from NCBI (AtAAT1: NP_001031394.1; AtAAT2: NP_197456.1; AtAAT3: NP_196713.1; AtAAT4: NP_564803.1; AtAAT5: NP_001031767.1; AtPAT: ADM67558.1).

ACKNOWLEDGEMENTS

We thank Timothy Wencewicz for providing prephenate. This work was supported by the NSF (MCB-1614539 to J.M.J. and IOS-1354971 to H.M.). C.K.H. was an NSF Graduate
Research Fellow (DGE-1143954). D.A.B. received supported through St. Louis Students and Teachers as Research Scientists (STARS) program. M.L.K. was supported by a grant to J.M.J. from the Howard Hughes Medical Institute Science Education Program. Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center of the Advanced Photon Source, a national user facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research (DE-AC02-06CH11357).
REFERENCES


Cook SD, Nicholas DS, Smith J, Chourey PS, McAdam EL, Quittenden L, Ross JJ (2016) Auxin biosynthesis: are the indole-3-acetic acid and phenylacetic acid biosynthesis pathways mirror images? *Plant Physiol.* 171, 1230-1241


Chapter 6

Fidelity in substrate specificity of *Arabidopsis thaliana* GH3 acyl-acid amido synthetases

Cynthia Holland\(^1\) and Joseph Jez\(^1\)

\(^1\)Department of Biology, Washington University, St. Louis, MO 63130, USA
ABSTRACT

As a way of regulating the concentration of a given plant hormone within the cell, plants contain a class of enzymes known as GH3 proteins that conjugate amino acids to acyl acid hormones to activate or inactivate the hormone molecule. Grouping of the GH3 proteins by substrate preference shows that the largest class of GH3 proteins use the growth-promoting hormone auxin (indole-3-acetic acid, or IAA), and the second largest class uses the defense hormone jasmonic acid; however, some of the auxin-specific GH3 proteins, notably Arabidopsis thaliana GH3.5, have been found to use multiple acyl acid substrates, including the auxins IAA and phenylacetic acid as well as benzoic acid, with roughly the same catalytic efficiencies. Despite the promiscuity in acyl acid substrate preference, the active site residues that are in contact with a ligand in the crystal structure of AtGH3.5 do not vary from those found in the AtGH3.2 active site. Using single-turnover kinetics, the data suggests that both proteins prefer IAA as the substrate. In conducting these experiments, we found that the adenylated reaction intermediate was being hydrolyzed into acyl acid and AMP in the absence of the amino acid, a typical feature of pre-transfer editing. Here we show that a non-cognate acyl acid-adenylate intermediate is more easily hydrolysable than the cognate acyl acid-adenylate, likely due to a slowed structural switch that provides a checkpoint for fidelity before the full reaction proceeds.
INTRODUCTION

Plant responses to developmental cues and the environment are primarily modulated by the activity and concentration of hormones, including auxins, jasmonates, and benzoates (Woodward and Bartel, 2005; Loake and Grant, 2007; Browse, 2009; Santner et al., 2009; Westfall et al., 2013). The primary form of auxin, indole-3-acetic acid (IAA), is involved in cell growth and division, elongation, and proliferation, while the jasmonates (JA) are involved in pathogen responses, root growth, seed germination, and fertility (Staswick et al., 2002; Staswick and Tiryaki, 2004; Santner et al., 2009). Fluctuation in hormone concentration within the plant cell and modulations of the chemical structure of a hormone affect receptor interactions and therefore alter developmental responses (Hagen et al., 1984; Staswick et al., 2005; Terol et al., 2006; Chen et al., 2010; Westfall et al., 2013).

GH3 proteins, or acyl acid-amido synthetases, are responsible for maintaining hormone levels in plants by conjugating acyl acid hormones to amino acids (Staswick et al., 2002; Staswick et al., 2005). This family of proteins generates conjugates through a two-step mechanism that involves both an adenylation and transfer reaction (Staswick et al., 2002; Staswick et al., 2005; Terol et al., 2006; Chen et al., 2010; Westfall et al., 2010; Okrent and Wildermuth, 2011). In the first half-reaction, the acyl acid substrate is adenylated and pyrophosphate is released (Chen et al., 2010; Westfall et al., 2010); this is followed by a transferase reaction where the nucleophilic amine group of an amino acid displaces AMP to form an acyl acid-amino acid conjugate.

Genome and transcriptome data has identified 19 putative GH3 genes in the model plant Arabidopsis thaliana (thale cress) (Hagen and Guilfoyle, 2002). The physiological role of the hormone-amino acid conjugate depends upon both the hormone and the amino acid. For
example, the conjugation of amino acids to IAA leads to an inactive form of the molecule, and adding an aspartate or glutamate to IAA leads to hormone degradation, while alanine- and valine-IAA conjugates are hydrolysable and are considered IAA storage forms (LaClere et al., 2002). With other hormones, such as JA, the acyl acid is activated by the addition of the amino acid, isoleucine, to form the molecule that is recognized by the hormone receptor, JA-isoleucine (Staswick et al., 2002; Fonseca et al., 2009; Meesters et al., 2014). Other less studied GH3 acyl acid substrates include benzoic acid (BA) and phenylacetic acid (PAA) (Westfall and Sherp et al., 2016). BA is a precursor to many plant metabolites, including the defense hormone salicylic acid (SA, 2-hydroxy-BA) (Widhalm and Dudareva, 2015), aromatic cytokinins (Werbrouck et al., 1996; Strnad, 1997; Tarkowska et al., 2003; Mutui et al., 2012), the cofactor ubiquinone (Block et al., 2014; Tohge et al., 2014), folic acids (vitamin B9) (Basset et al., 2004a; Basset et al., 2004b; Hanson and Gregory, 2011), and numerous specialized metabolites involved in attracting pollinators and seed dispersers (Wang and De Luca, 2005; Goff and Klee, 2006; Dudareva et al., 2006; Schwab et al., 2008; Dudareva et al., 2013). On the other hand, phenylacetic acid (PAA) is a non-indolic active auxin that is present at physiologically relevant levels within plants (Korasick et al., 2013). PAA is widely distributed in both vascular and non-vascular plants and regulates the same signaling pathways as IAA (Sugawara et al., 2015).

GH3.2, or YDK1, from A. thaliana is an auxin-using GH3 protein that conjugates aspartate to IAA and PAA, as well as BA to a lesser extent (Westfall and Sherp et al., 2016). The ydk 1-D mutant in Arabidopsis has been characterized as a dwarf mutant with a short hypocotyl in both light and darkness, and the YDK1 gene is expressed in roots and flowers and is a negative regulator of auxin signaling (Takase et al., 2004). GH3.5, or WES1, from Arabidopsis is similar to GH3.2 in that it too conjugates aspartate to IAA, PAA, and BA (Westfall and Sherp et al.,
Unlike GH3.2, GH3.5 is more of a generalist and uses each of these hormones with a similar catalytic efficiency with little known preference for one over another. Arabidopsis *wes1* T-DNA insertional mutants are phenotypically indistinguishable from wild-type plants under normal growth conditions but are more susceptible to abiotic and biotic stresses than wild-type (Park et al., 2007). Studies involving the *wes1* loss-of-function mutant and two gain-of-function mutants (*wes-1D* and *gh3.5-1D*) show that this gene contributes to both IAA and SA responses and is implicated in SA-linked pathogen responses (Park et al., 2007a; Zhang et al., 2007; Park et al., 2007b; Zhang et al., 2008; Chen et al., 2013).

Unlike GH3.2 and GH3.5, GH3.11 (JAR1) from Arabidopsis is a JA-conjugating enzyme that forms the active hormone, JA-isoleucine (Browse, 2009; Staswick et al., 2002; Sheard et al., 2010). The *jar1* mutant in Arabidopsis does not produce JA-isoleucine and is blocked in JA-mediated responses, including defense against herbivores, fungi, and pathogens (Staswick et al., 2002; Staswick and Tiryaki, 2004; Howe and Jander, 2008).

Several crystal structures of Arabidopsis and grapevine (*Vitus vinifera*) have provided insight on the sequence-structure-function relationship among the GH3 family of proteins (Peat et al., 2012; Westfall and Zubieta et al., 2012; Round et al., 2013; Westfall and Sherp et al., 2016; Chen et al., 2017). GH3 proteins share a common tertiary structure containing two domains—a large N-terminal domain and a smaller C-terminal domain characteristic of the adenylating firefly luciferase (ANL) enzyme superfamily—with the active site at the interface of the two domains (Westfall and Zubieta et al., 2012). A flexible hinge loop connects the two domains and pivots the C-terminal domain during the reaction. In the open conformation, ATP, Mg\(^{2+}\), and the acyl acid bind for the adenylation reaction to occur, and a solvent-accessible channel allows for pyrophosphate release. Upon adenylation and pyrophosphate release, the C-terminal domain
rotates about the hinge loop and closes the active site, repositioning residues for the transferase reaction to occur (Figure 1).

Figure 1. The GH3 enzymatic reaction.
GH3 proteins proceed through a two-step reaction, an adenylation and transfer reaction. In the first half-reaction, an acyl acid, Mg\(^{2+}\), and ATP bind the open form of the protein, the acyl acid is adenylated and pyrophosphate is released. In the second half-reaction, the enzyme closes through a pivoting of the C-terminal domain (gold) for the transfer reaction to proceed. The final acyl acid-amino acid conjugate is released, and the enzyme resets.

From Arabidopsis, structures of GH3.5, GH3.11, and GH3.12 have been determined, and the substrates of uncharacterized GH3 proteins can be predicted based on analysis of the active site residues (Westfall and Zubieta et al., 2012). For example, GH3.2 was predicted to use auxin (IAA) as an acyl acid substrate, and this is the case (Westfall and Sherp et al., 2016). Based on amino acid sequence alignments, all of the residues in contact with a ligand in the structure of AtGH3.5 are conserved in AtGH3.2; these two proteins also use the same amino acid, aspartate, for the second half-reaction (Westfall and Sherp et al., 2016). Despite these similarities, the steady state kinetics suggests the two proteins have different catalytic efficiencies, and likely preferences, for different acyl acid substrates. While structural data has allowed for the separation of JA-using from IAA-using GH3 proteins, questions remain as to how enzymes with
the same active site residues have differential preferences for similar substrates, such as IAA, PAA, and BA, that have distinct physiological roles in the plant.

Figure 2. Visualization of the first half and full reactions of GH3.2, GH3.5, and GH3.11 with cognate substrates.
TLC reactions are shown for all three proteins with IAA for GH3.2 and GH3.5 and with JA for GH3.11. The first half-reaction is shown as (-), meaning no amino acid was added, whereas the full reaction is shown in the lanes with a (+), meaning amino acid was added. Acyl-AMP forms (marked with a red arrow) in the first half-reaction, while AMP accumulates in the full reaction.

To address the question of substrate preference in GH3 proteins, we used a single-turnover kinetics approach. The rate of the first half-reaction, $k_1$, was determined for GH3.2, GH3.5, and GH3.11 with various acyl acid substrates, including IAA, BA, and PAA for the auxin-using enzymes and JA for GH3.11 (Figure 2). In conducting these experiments, we noticed that AMP accumulated in the first half-reaction in the absence of amino acid, suggesting that the enzyme may be hydrolyzing intermediates that contain non-cognate acyl acids. To understand the influence of distal active site residues on substrate specificity, chimeras of GH3.2 and GH3.5, which use the same amino acid substrate and use IAA, BA, and PAA to varying extents, were generated by swapping the C-terminal domains of the two proteins. Steady state and single-turnover kinetics for these chimeras showed that both domains are critical for activity and substrate selection. These data broaden our understanding of the reaction chemistry of an
important class of plant hormone modifying enzymes and reveals the mechanisms by which the enzyme selects a substrate.

METHODS

Protein Expression and Purification

AtGH3.2, AtGH3.5, and AtGH3.11 cloning, expression, and purification have been described previously (Westfall and Zubieta et al., 2012; Westfall and Sherp et al., 2016). Codon-optimized AtGH3 N2/C5 and AtGH3 N5/C2 were synthesized (Genewiz) and cloned into the pET28a expression vector, which was transformed into E. coli Rosetta2 (DE3) cells (Novagen). Cells were cultured in Terrific broth until A_{600nm} 0.6–0.8 was obtained, at which time a final concentration of 0.75 mM isopropyl-β-D-1-thiogalactopyranoside was added to induce protein expression. After a 16-hour incubation at 16°C, cells were pelleted by centrifugation (5000 x g; 10 min) and resuspended in lysis buffer (50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, 10% (v/v) glycerol, and 1% (v/v) Tween-20). Following sonication, cell debris was removed by centrifugation (13,000 x g; 45 min). The resulting lysate was passed over a Ni^{2+}-nitrilotriacetic acid (Qiagen) column equilibrated in the lysis buffer. The column was then washed (50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, and 10% (v/v) glycerol) and bound His-tagged protein eluted (50 mM Tris (pH 8.0), 500 mM NaCl, 250 mM imidazole, and 10% (v/v) glycerol). Protein concentration was determined by the Bradford method (Protein Assay, Bio-Rad) with bovine serum albumin as standard.
Thin-layer Chromatography Assays

Individual reactions contained 50 mM Tris (pH 8.0), 1 mM MgCl₂, 5 µM ATP with 1 µCi of [α-³²P]ATP, 0.1 mM acyl-acid substrate, and 50 µM protein. The specificities of AtGH3.2, AtGH3.5, AtGH3 N2/C5, and AtGH3 N5/C2 were tested with indole-3-acetic acid, benzoic acid, and phenylacetic acid, while the specificity of AtGH3.11 was tested with jasmonate. To visualize the two half-reactions for these enzymes, 0.8-1 mM acyl-acid substrate, 0.1 mM ATP, 1 µCi of [α-³²P]ATP, in Tris and MgCl₂ were used in the first half-reaction, and 10 mM amino acid (glutamate or isoleucine) was added to the reaction buffer for the second half-reaction. Reactions were allowed to proceed for 1-2 minutes before they were quenched with formic acid, after which 1 µl was spotted onto a cellulose thin-layer chromatography plate, which was developed in isobutyrate: ammonium hydroxide:water (66:1:33) (Christensen and Cronan, 2009). TLC plates were analyzed by phosphorimaging and quantified using Bio-Rad Quantity One software. Reactions were repeated in triplicate, and rate data were determined by fitting the data to a one-phase decay equation using GraphPad Prism software.

Steady-state Enzyme Assays

AtGH3 N2/C5 and N5/C2 activities were determined using a coupled assay system with a standard reaction buffer of 50 mM Tris (pH 8.0) and 1 mM MgCl₂ in a 100 µL reaction volume, as previously described (11). Kinetic analysis of acyl acid substrates used ATP (1 mM) and aspartate (5 mM) with varied concentrations of acyl acid (0 to 1 mM). Reactions were initiated by the addition of the enzyme (10 µg) and were conducted at 25 °C. The data then were fit to the Michaelis–Menten equation using GraphPad Prism (version 7.0c for Mac, GraphPad Software).
RESULTS

Visualization of the First Half-reaction and Full Reaction of GH3s using TLC

Despite recent molecular and biochemical investigations of the GH3 family of acyl-acid amid synthetases in plants, the basis of substrate specificity in these hormone-conjugating proteins remains to be understood. Previous steady state kinetics data suggests that AtGH3.2 is specific for indole-3-acetic acid (IAA) while AtGH3.5 does not have a preference for IAA, PAA (phenylacetic acid), or BA (benzoic acid) (Westfall and Sherp et al., 2016). While structural studies have aided our understanding of the classification of GH3 substrate preference based on residues found in the acyl acid binding pocket, the subtleties that differentiate two auxin-using GH3s, for example, remain unknown (Westfall and Zubieta et al., 2012). Moreover, the acyl acid binding pockets for GH3.2 and GH3.5 are nearly identical, although the catalytic efficiencies with different acyl acids vary (Westfall and Sherp et al., 2016).

To begin to unravel the mechanisms by which GH3 proteins differentiate among acyl acid substrates with similar chemical structures, we first designed and tested a TLC method for visualizing the first and second half-reactions of GH3.2, GH3.5, and GH3.11 with their cognate substrates (Figure 2). In these reactions, the substrate, [α-32P]ATP, and the enzyme were mixed and the reaction was quenched before being spotted on a TLC plate, allowing us to follow the formation of a labeled acyl-AMP intermediate after the first half-reaction and the release of labeled AMP in the second half-reaction once the amino acid was added to the reaction mixture.

Single-turnover Kinetics for First Half-reaction of AtGH3.2, AtGH3.5, and AtGH3.11

Using the TLC plates with [α-32P]ATP in the enzyme reaction mixture, the rate of the first
Figure 3. Single turnover kinetics of GH3.2 with IAA, GH3.5 with IAA, and GH3.11 with JA.

Time course reactions are shown (right) for each of the proteins for the first half-reaction for (a) GH3.2 with IAA, (b) GH3.5 with IAA, and (c) GH3.11 with JA. Data were quantified to calculate the reaction rate (left).

Half-reaction ($k_1$) of GH3.2, GH3.5, and GH3.11 was measured using single-turnover kinetics (Figure 3). In these experiments, GH3.2 and GH3.5 were assayed using IAA, PAA, and BA as acyl acid substrates, while GH3.11 was assayed with JA. Single-turnover kinetics were determined using a 10:1 molar ratio of enzyme over substrate. Reactions were initiated by the addition of enzyme, and data were fit to a first-order exponential curve to determine the $k_1$ for each of the three GH3 proteins with each of the acyl acid substrates. GH3.2 forms IAA-AMP, the product of the first half-reaction, at a rate of 21.5 min$^{-1}$, while the rate for IAA-AMP
formation with GH3.5 is 0.834 min\(^{-1}\) (Table 1). GH3.11 forms JA-AMP at a rate of 1.75 min\(^{-1}\).

With GH3.5, BA-AMP and PAA-AMP both form at rate of 0.186 min\(^{-1}\), while BA-AMP (1.40 min\(^{-1}\)) forms at a lower rate than PAA-AMP (9.06 min\(^{-1}\)) with GH3.2. These rates suggest that GH3.2 forms IAA-AMP intermediates faster than PAA-AMP with BA-AMP as the slowest reaction of the three substrates. With GH3.5, the rate of IAA-AMP is highest and is 4.6 times higher than that of either PAA-AMP or BA-AMP. These results are similar to what would be expected based on the steady state kinetics of these enzymes.

<table>
<thead>
<tr>
<th>Hormone</th>
<th>AtGH3.2 (min(^{-1}))</th>
<th>AtGH3.5 (min(^{-1}))</th>
<th>AtGH3.11 (min(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAA</td>
<td>(k_1) 21.5 ± 2.78</td>
<td>0.834 ± 0.120</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(k_{\text{AMP}}) 0.042 ± 0.006</td>
<td>0.066 ± 0.012</td>
<td>--</td>
</tr>
<tr>
<td>BA</td>
<td>(k_1) 1.40 ± 0.390</td>
<td>0.186 ± 0.042</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(k_{\text{AMP}}) 0.072 ± 0.012</td>
<td>0.054 ± 0.006</td>
<td>--</td>
</tr>
<tr>
<td>PAA</td>
<td>(k_1) 9.06 ± 2.12</td>
<td>0.186 ± 0.030</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>(k_{\text{AMP}}) 0.108 ± 0.018</td>
<td>0.042 ± 0.012</td>
<td>--</td>
</tr>
<tr>
<td>JA</td>
<td>(k_1) --</td>
<td>--</td>
<td>1.75 ± 0.120</td>
</tr>
<tr>
<td></td>
<td>(k_{\text{AMP}}) --</td>
<td>--</td>
<td>0.120 ± 0.006</td>
</tr>
</tbody>
</table>

Table 1. Single turnover kinetics data for AtGH3.2, AtGH3.5, and AtGH3 N2/C5 for IAA, BA, and PAA first half-reactions (n=3).

\(k_1\) is depicted as the \(k\) value when an acyl acid-AMP is the product, and \(k_{\text{AMP}}\) is shown as the \(k\) value when AMP is the product for each of the acyl acid substrates.

Rates of AMP Accumulation in the First Half-reaction of AtGH3.2, AtGH3.5, and AtGH3.11

In these time course reactions, we noticed that not only was acyl-AMP accumulating over time, but AMP was also accumulating in the first half-reaction. We could visualize AMP release in the absence of the amino acid, similar to the BioW acyl-coenzyme A synthetase/ligase from *Bacillus subtilis* and the aminoacyl-tRNA synthetases (Yadavalli and Ibba, 2012; Manandhar and Cronan, 2013). This accumulation was visible in all three GH3 proteins with their cognate
substrates, and a rate of AMP release could be measured for each of the first half-reactions, although the rates lagged behind the rates of acyl-AMP formation (Table 1). The rate of AMP release was also measured for the cognate substrates in addition to PAA and BA. Interestingly, the rates of AMP release (k<sub>AMP</sub>) vary based on the acyl acid substrate. For GH3.2, the rate of AMP release with PAA (0.108 min<sup>-1</sup>) is about 2 times faster than that of that of the AMP released with IAA (0.042 min<sup>-1</sup>) and 1.5 times that of the AMP released with BA (0.072 min<sup>-1</sup>). The rates of AMP release in the first half-reaction with GH3.5 do not vary as much among the acyl acid substrates (0.066, 0.054, and 0.0042 min<sup>-1</sup> for IAA, BA, and PAA, respectively). The rate of AMP release with GH3.11 and JA is 0.120 min<sup>-1</sup>.

**Steady state Kinetics of Chimeric AtGH3.2 and AtGH3.5**

Despite having differences in catalytic efficiencies with various acyl acid substrates, GH3.2 and GH3.5 use the same amino acid (aspartate), and all of the residues in their acyl acid binding pocket that form the active site are identical (Westfall and Sherp et al., 2016). The residues that vary between the two proteins were not in direct contact with a ligand in the AtGH3.5 x-ray structure. To ascertain whether or not there is a structural component to the enzyme’s substrate promiscuity, chimeric proteins were generated by swapping the N-terminal domain of one GH3 with the other (Figure 4). The resulting proteins have been named GH3 N2/C5 (N-terminal domain of GH3.2 (1-442) and C-terminal domain of GH3.5 (442-605)) and GH3 N5/C2 (1-449; 450-556). The two domains are connected through a hinge loop region, which is where the site of the domain swap was introduced.

The chimeric proteins were first assayed for activity using steady state kinetics. While GH3 N5/C2 is inactive, GH3 N2/C5 have activity with IAA, PAA, and BA (Table 2). AtGH3.2 has a
catalytic efficiency (kcat/Km) of 556 M⁻¹s⁻¹ for IAA, 125 M⁻¹s⁻¹ for PAA, and 77 M⁻¹s⁻¹ for BA, while AtGH3.5 has a catalytic efficiency of 314 M⁻¹s⁻¹ for IAA, 253 M⁻¹s⁻¹ for PAA, and 338 M⁻¹s⁻¹ for BA (Westfall and Sherp et al., 2016). The substrate with the highest turnover for both of these enzymes is PAA (19.5 min⁻¹ for GH3.2 and 28 min⁻¹ for GH3.5). Unlike these two enzymes, the GH3 N2/C5 enzyme has the highest turnover (74 min⁻¹) with BA. The chimeric protein has the highest catalytic efficiency with IAA (480 mM⁻¹min⁻¹), followed by PAA (360 mM⁻¹min⁻¹) (Table 2). Despite the high turnover with BA as a substrate, a high Km value (440 µM) leads to a low catalytic efficiency (170 mM⁻¹min⁻¹). Compared to the native enzymes, the
chimeric enzyme follows the same pattern as GH3.2 for catalytic efficiency, with IAA having the highest and BA having the lowest.

<table>
<thead>
<tr>
<th>Hormone</th>
<th>Km (µM)</th>
<th>E/Ei (nmol min⁻¹mg⁻¹)</th>
<th>kcat (min⁻¹)</th>
<th>kcat/Km (mM⁻¹min⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAA</td>
<td>54 ± 17</td>
<td>400 ± 53</td>
<td>26 ± 3.5</td>
<td>480</td>
</tr>
<tr>
<td>BA</td>
<td>440 ±68</td>
<td>1100 ± 85</td>
<td>74 ± 5.6</td>
<td>170</td>
</tr>
<tr>
<td>PAA</td>
<td>150 ±28</td>
<td>800 ± 51</td>
<td>52 ± 3.3</td>
<td>360</td>
</tr>
</tbody>
</table>

**Table 2. Steady state kinetics for GH3 N2/C5.**
Values were determined by fitting the data to the Michaelis-Menten equation (n=3).

**Single-turnover Kinetics for GH3 N2/C5**

In the single-turnover kinetics assays, the highest $k_1$ for GH3 N2/C5 is with IAA as a substrate (13.3 min⁻¹), followed by PAA (1.59 min⁻¹), and then BA (0.414 min⁻¹) (**Table 3**). The $k_{\text{AMP}}$ for the chimera is highest with PAA (0.312 min⁻¹) and lowest with IAA (0.042 min⁻¹), with BA in between (0.102 min⁻¹). These results suggest that the chimera is more likely to form IAA-AMP intermediates than BA- or PAA-AMP intermediates and will degrade PAA-AMP three times faster than BA-AMP or about 7.5 times faster than IAA-AMP.

<table>
<thead>
<tr>
<th>Hormone</th>
<th>$k_1$ (min⁻¹)</th>
<th>$k_{\text{AMP}}$ (min⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>IAA</td>
<td>13.3 ± 2.59</td>
<td>0.042 ± 0.006</td>
</tr>
<tr>
<td>BA</td>
<td>0.414 ± 0.036</td>
<td>0.102 ± 0.018</td>
</tr>
<tr>
<td>PAA</td>
<td>1.59 ± 0.322</td>
<td>0.312 ± 0.012</td>
</tr>
</tbody>
</table>

**Table 3. Single turnover kinetics data for AtGH3 N2/C5 first half-reactions (n=3).**
$k_1$ is depicted as the $k$ value when an acyl acid-AMP is the product, and $k_{\text{AMP}}$ is shown as the $k$ value when AMP is the product for each of the acyl acid substrates.
DISCUSSION

GH3 proteins, or acyl acid-amido synthetases, regulate the fluctuations of active hormone concentrations within a plant cell using amino acid conjugation to generate either a storage or degradation form or an active form of a hormone (Korasick et al., 2013). Many acyl acid plant hormones, including the growth hormone indole-3-acetic acid (auxin) and the defense hormone jasmonate, are substrates of these proteins. Although most GH3 proteins are specific for a single hormone, several of these proteins have relaxed substrate preferences and can use similar acyl acids with almost equal catalytic efficiencies (Westfall and Sherp et al., 2016). In the case of non-specific GH3 proteins, such as AtGH3.5, the physiological roles of IAA-, PAA-, and BA-aspartate conjugates may greatly vary, so it seems that the protein either accepts multiple substrates as a way of controlling several metabolites simultaneously, or one is preferred over another for a more targeted physiological response. Despite intensive studies on the physiology, biochemistry, and three-dimensional structures of these proteins, our understanding of the mechanisms by which GH3 proteins select a substrate out of the milieu of acyl acids in the cell with which to form an amino acid conjugate escapes our current knowledge.

After developing a TLC assay to visualize the GH3 first half-reaction and full reaction, we addressed the question of substrate preference using single-turnover kinetics experiments. These reactions measure the rate ($k_1$) of the first half-reaction of two auxin-using GH3s (GH3.2 and GH3.5) with a panel of substrates that included IAA, PAA, and BA. The data show that AtGH3.2 forms the cognate IAA-AMP intermediate at a rate much higher than with BA or PAA as acyl acid substrates with a ratio of about 20:9:1 for IAA:PAA:BA (Table 1). From a physiological standpoint, it does make sense that the enzyme may have evolved the ability to generate PAA conjugates since this hormone is able to interact with the TIR1/AFB auxin
receptor that IAA binds and can regulate the same genes as free IAA (Sugawara et al., 2015). The physiological roles of PAA-aspartate conjugates are not well understood in plants.

For AtGH3.5, a generalist enzyme based on steady state kinetics, the reaction rates were much slower than those of AtGH3.2, but nonetheless there was a clear difference among the acyl acid substrates (Table 1). IAA-AMP intermediates were formed at a rate that was about four times that of the rate of BA- or PAA-AMP intermediates. It seems that the enzyme does have clear preference for IAA as the acyl acid hormone substrate.

To our surprise, we noticed an accumulation of AMP in the first half-reaction for these proteins with these substrates in the absence of the amino acid (Figure 3; Table 1). In searching the literature, we found that others have reported similar findings with adenylating enzymes, including an example in a bacterial enzyme in biotin biosynthesis (Manandhar and Cronan, 2013). In this bacterial enzyme, pimeloyl-coenzyme A (CoA) ligase (BioW) from *Bacillus subtilis*, there was an accumulation of AMP in the absence of the final product synthesis when a non-cognate substrate was assayed with the enzyme. In both BioW and the GH3 proteins, the adenylate intermediate is synthesized and then cleaved into free acyl acid and AMP, which is considered to be a means of enzymatic proofreading to prevent non-cognate substrates from going through the complete reaction. The classic example of AMP accumulation with non-cognate substrates, a hallmark for enzymatic editing (proofreading), is by tRNA synthetases (Martinis and Bonlecki, 2010; Yadavalli and Ibba, 2012).

We were able to calculate the rate of AMP accumulation in the first half-reaction of all of the GH3 proteins analyzed in this study (Table 1). For AtGH3.2, $k_{AMP}$ is highest with PAA with a ratio of approximately 11:7:4 for PAA: BA: IAA, respectively (Table 1). The ratio of the rates of AMP accumulation were about equal with AtGH3.5 with a ratio of about 4:5:7 for PAA: BA:
IAA, respectively. Despite the rate of AMP accumulation being highest for IAA with GH3.5, there is more IAA-AMP being made compared to the degradation of the intermediate. For every roughly 13 IAA-AMP molecules that are formed, one will be degraded; whereas with BA and PAA about one in every four or five acyl-AMP molecules will be degraded to free acyl acid and AMP. This suggests that despite the generalist nature of this enzyme in steady state kinetics assays, GH3.5 does prefer IAA as its cognate substrate.

Ultimately, this data led us to ask whether the acyl acid intermediate is hydrolyzed in the active site of the enzyme or in solution. We do know that there is an enzymatic component because we are able to calculate rates for acyl-AMP hydrolysis (Table 1). There are, however, two possible scenarios. One is that the enzyme is able to hydrolyze the substrate using reactive residues, such as arginine(s) in the case of BioW, in the active site, and the free acyl acid and AMP then diffuses away from the enzyme’s active site (Estrada et al., 2017). An alternative hypothesis is that when a non-cognate substrate binds, the enzyme is unable to close, keeping the active site open for the adenylate to be released into solution where the acyl-AMP bond would be cleaved by a hydroxide ion (Manandhar and Cronan, 2013). When the cognate substrate does bind, the enzyme closes and awaits the amino acid substrate for the transferase reaction to occur. Based on structural studies of the GH3 family of proteins, it is unlikely that there is a specific proofreading domain, as is the case for many of the tRNA synthetases (Martinis and Bonlecki, 2010; Yadavalli and Ibba, 2012); instead, it is more likely that the enzyme edits for fidelity in the absence of such a domain.

To understand the structural component to pre-transfer substrate proofreading, we generated chimeras of two GH3s that used similar acyl acid substrates and the same amino acid substrate, aspartate (Figure 4). While the residues in contact with ligands in the structure of GH3.5 are
exactly conserved, distal active site residues vary, and the C-terminal domain of GH3.5 is larger than that of GH3.2 (Westfall and Sherp et al., 2016). The GH3 N5/C2 enzyme was inactive in steady state and single-turnover assays, which suggests either that the smaller C-terminal domain of GH3.2 is insufficient for activity with the N-terminal domain of GH3.5 or that the active site of the enzyme did not form properly (Table 2). GH3 N2/C5, however, was active in both steady state and single-turnover assays. Based on the Michaelis-Menten kinetics, the enzyme functioned similarly to GH3.2 with IAA as the preferred substrate, followed by PAA (Table 2). The chimeric protein actually has a higher turnover with IAA (26 min\(^{-1}\)) than either of the native enzymes from which it was generated (15 min\(^{-1}\) for AtGH3.5 and 17 min\(^{-1}\) for AtGH3.2) (Westfall and Sherp et al., 2016). These results were confirmed with single-turnover experiments that showed a 32:4:1 ratio of the \(k_1\) for IAA: PAA: BA substrates, respectively (Table 3). Similar to the native enzymes, \(k_{\text{AMP}}\) could be measured for all three substrates in the first half-reaction with a ratio of about 7:2:1 for PAA-AMP: BA-AMP: IAA-AMP formed, also similar to GH3.2.

Although the single-turnover kinetics suggest that IAA is a preferred substrate for the auxin-using GH3s assayed here, the enzymes were able to form PAA and BA conjugates. In the plant cell, there is likely a flux of concentration of these hormone metabolites depending on many biotic and abiotic factors, and thus at higher concentrations, these molecules could outcompete IAA for binding auxin-using GH3s. Still, the physiological roles of PAA- and BA-conjugates remain to be investigated (Chong et al., 2001).

Taken together, our data suggests that the GH3 family of acyl acid amido synthetases in plants are able to proofread substrates in the first half-reaction to ensure hormone substrate fidelity (Figure 5). The work of others has shown that movement of the C-terminal domain in the first half-reaction occurs and allows for the reaction to proceed (Westfall and Zubieta et al.,
It is clear that substrates bind the enzyme in the open conformation, and the transfer reaction takes place in the second half-reaction when the enzyme is in a closed conformation. The enzyme then reopens to release the conjugate and bind more substrates. Based on data generated here, the pivoting of the C-terminal domain to conform to the closed state

![Figure 5. GH3 reaction mechanism, including proofreading of adenylated intermediates.](image)

In the first half-reaction, the acyl acid hormone is adenylated; if the adenylated intermediate is non-cognate, the intermediate is hydrolyzed to AMP and acyl acid ($k_{\text{AMP}}$). Occurs more quickly when the cognate substrate binds and is adenylated. When non-cognate substrates bind, the pivoting, or closing, of the active site is delayed, and the adenylated intermediate is degraded. This opening and closing of the enzyme allows for the enzyme to proofread substrates before the transfer reaction to ensure the fidelity of the reaction, which prevents the enzyme from the expensive synthesis of potentially inactive molecules.

**ACKNOWLEDGEMENTS**

We thank Hani Zaher for providing [$\alpha$-$^{32}$P] ATP. This work was supported by the NSF (MCB-1614539 to J.M.J.). C.K.H. was supported by the NSF Graduate Research Fellowship Program (DGE-1143954).
REFERENCES


Staswick PE, Tiryaki I, Rowe ML (2002) Jasmonate response locus JAR1 and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. Plant Cell 14(6), 1405-1415


184


Chapter 7

Characterization of a Chorismate-Conjugating Class of 
Brassicaceae GH3 Proteins

Cynthia K. Holland¹, Corey S. Westfall¹, Alejandro De Santiago-Perez¹, Chloe Zubieta², Sophie 
Alvarez³, and Joseph M. Jez¹*

¹Department of Biology, Washington University, St. Louis, MO 63130, USA
²European Synchrotron Radiation Facility, 38000 Grenoble, France
³Department of Agronomy and Horticulture, University of Nebraska, Lincoln, NE 68588

Author contributions: CKH, CSW, CZ, and JM designed experiments; CKH and CSW 
conducted kinetic analysis, and CZ determined the x-ray structure; SA contributed mass- 
spectrometry data; CKH generated the gh3.7/gh3.12 double mutant Arabidopsis line, and CKH 
and ADSP conducted in vivo experiments. CKH, CSW, and JM analyzed the data and wrote the 
manuscript.
ABSTRACT

As a way of modulating chemicals in response to a developmental or environmental cue, plants have GH3 proteins that conjugate an amino acid onto a plant hormone to regulate its concentration and downstream transcriptional responses. There are 19 GH3 proteins in the model plant Arabidopsis, but only 9 of those proteins have confirmed substrates. One clade of GH3 proteins is predicted to use benzoate as substrate and includes AtGH3.7 and AtGH3.12. Previously, AtGH3.12 was identified as a 4-hydroxybenzoic acid-glutamate synthetase and was found to influence pathogen defense responses through the plant defense hormone salicylic acid. Here, we screened 15 plant hormones with 20 amino acids to find that AtGH3.7 is a chorismate-cysteine synthetase. Rescreening of AtGH3.12 found that it too uses chorismate and functions as a chorismate-glutamate synthetase. These results were confirmed using a combination of x-ray crystallography, site-directed mutagenesis, single-turnover kinetics, and mass spectrometry. Because chorismate is a precursor to the three aromatic amino acids in plants and a precursor to salicylic acid, these proteins are modulating the concentration of a pathway intermediate. Interestingly, gh3.7 mutants did not have the same pathogen susceptibility phenotype as gh3.12 mutants, but the gh3.7/gh3.12 double mutant was more susceptible to pathogen infection than gh3.12 alone. This study reveals how GH3 proteins, which traditionally regulate plant hormones, can also use hormone intermediates to regulate multiple hormones.
INTRODUCTION

As sessile organisms, plants have adapted and evolved a number of chemicals to respond to the world around them (Santer et al., 2009). By modifying common growth and defense hormones, plants are able to activate or inactivate hormone molecules rapidly in response to environmental cues (Westfall et al., 2013; Korasick et al., 2013). One family of hormone modifying enzymes is the GH3 family of acyl acid amido synthetases, an enzyme family responsible for conjugating amino acids to acyl acids, notably the plant hormones indole-3-acetic acid and jasmonate, to alter the plants’ physiology (Hagen and Guilfoyle, 1985; Westfall et al., 2013). In some cases, the amino acid-conjugated hormone is a storage or degradation form of the molecule, but in other cases, the conjugate is the active form of the molecule that binds the hormone receptor (Feys et al., 1994; Xie et al., 1998; LeClere et al., 2002).

While GH3 proteins are found in all plants, the roles of many of these proteins remains to be investigated. In Arabidopsis, there are 19 GH3 proteins, and to date, only 8 of these have been biochemically characterized and have an identified substrate—GH3.1, GH3.2, GH3.3, GH3.5, GH3.11, GH3.12, GH3.15, and GH3.17 (Staswick et al., 2002; Staswick et al., 2005; Westfall and Zubieta et al., 2012; Westfall and Sherp et al., 2016; Sherp et al., 2018). Using the x-ray crystal structures of AtGH3.11 and AtGH3.12, the acyl acid substrate binding pocket was compared to determine the active site residues across all GH3 proteins (Westfall and Zubieta et al., 2012). With this information, the function of a given GH3 protein can be predicted based on the known function of GH3 proteins within that clade.

The largest clade of GH3 proteins are predicted to use the auxin indole-3-acetic acid, and the well-characterized AtGH3.1, AtGH3.2, AtGH3.5, and AtGH3.17 fall into this clade (Westfall and Zubieta et al., 2012). The second largest clade contains AtGH3.11, or JAR1, the jasmonyl-
isoleucine synthetase, and it is predicted that other proteins in this family also use jasmonate as an acyl acid substrate. Interestingly, AtGH3.15 was recently found to use a hormone precursor, indole-3-butyric acid, as a substrate, suggesting a role for GH3 proteins outside of the canonical plant hormones (Sherp et al., 2018).

Another clade of GH3 proteins with a characterized homolog is Brassicaceae-specific and has two members in Arabidopsis—AtGH3.7 and AtGH3.12 (Westfall and Zubieta et al., 2012). AtGH3.12 was identified in a mutant screen, named pbs3-1 (avrPphB susceptible 3) (Warren et al., 1999), for increased susceptibility to both avirulent and virulent *Pseudomonas syringae* strains and was previously found to conjugate 4-hydroxybenzoic acid to glutamate (Okrent et al., 2009) These knockout plants were deficient in the biosynthesis of salicylate (SA), a plant pathogen defense hormone, and showed decreased accumulation of pathogen responsive gene transcripts and SA-glucosides (SAG) after infection (Jagadeeswaran et al., 2007; Nobuta et al., 2007; Okrent et al., 2009). While SA biosynthesis is predicted to occur in the chloroplast (Wildermuth et al., 2001; Strawn et al., 2007; Garciaion et al., 2008), SAG can be actively transported to the vacuole where it presumably functions as a storage form of SA (Dean and Mills, 2004; Dean et al., 2005; Hennig et al., 1993). Interestingly, treatment of *pbs3-I* plants with exogenous SA leads to normal induction of PR transcripts and SA-glucosides after infection and the SA treated *pbs3-I* plants are as resistant to *P. syringae* infection as wild-type (Nobuta et al., 2007). From this, it is believed that PBS3 must work upstream of SA, maybe in the SA biosynthesis pathway. AtGH3.7 has escaped in *vivo* and biochemical characterization.

SA can be synthesized via two routes in plants, both of which use the aromatic amino acid branchpoint metabolite chorismate as a precursor (Vlot et al., 2009). In one pathway, the chorismate-derived aromatic amino acid phenylalanine is converted into SA by benzoate
intermediates or coumaric acid (Mauch-Mani et al., 1996). The second pathway converts chorismate into SA using two enzymes—isochorismate synthase and a predicted, but unidentified, isochorismate pyruvate lyase (Serino et al., 1995; Wildermuth et al., 2001). There are two *isochorismate synthase* genes in Arabidopsis, but the *ics1/ics2* double mutant line still contains SA, suggesting that this pathway is not the only source of SA in plants (Garcion et al., 2008).

Here, we were interested in identifying an acyl acid substrate for AtGH3.7 and determining if it played a redundant role *in planta*. Along the way, we identified the substrate for AtGH3.7 and found that it was also the most logical substrate for AtGH3.12. Using a combination of steady-state kinetics, single-turnover kinetics, x-ray crystallography, and mass spectrometry, AtGH3.12 was determined to be a chorismate-glutamate synthetase, while AtGH3.7 is a chorismate-cysteine synthetase. The *in vivo* role of AtGH3.7 was analyzed through *Pseudomonas syringae pv. tomato* DC3000 infection assays, and a double mutant Arabidopsis line was also generated to determine the combined role of this unique class of chorismate-conjugating GH3 proteins in the Brassicaceae family.

**METHODS**

**Protein Expression and Purification**

The pET-28a:AtGH3.12 construct was used from previous work (Westfall and Zubieta et al., 2012), and codon-optimized pET-28a:AtGH3.7 was ordered from Genewiz. pET-28a:AtGH3.12 and pET28a:AtGH3.7 were transformed into the Rosetta2 (DE3) cells (Novagen), and protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside for 15-18 hours at 18°C. Cells were lysed by sonication, and proteins were purified using a Ni²⁺-nitrilotriacetic acid
column equilibrated in the lysis buffer (50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, 10% glycerol, and 1% Tween-20). The column was then washed with 50 mM Tris (pH 8.0), 500 mM NaCl, 20 mM imidazole, and 10% glycerol, and bound His-tagged protein was eluted with wash buffer containing 250 mM imidazole. Protein concentration was determined by the Bradford method (Bio-Rad) with bovine serum albumin as standard. For crystallography, GH3.12 was then further purified by size-exclusion chromatography using a Superdex-75 26/60 HiLoad ÄKTA FPLC size-exclusion column (GE Healthcare) equilibrated with 50 mM Tris (pH 8.0) and 100 mM NaCl. Mutants were either from previous work or generated using the Quikchange PCR method (Strategene) with expression and purification as above.

**Enzyme Assays**

Activity assays were performed using an enzyme coupled system containing 10 units of myokinase, pyruvate kinase, and lactate dehydrogenase, 1 mM phosphoenolpyruvate, and 1 mM NADH which couples AMP generation to conversion of NADH to NAD\(^+\), which leads to a decrease in absorbance at 340 nm (Chen et al., 2010). Reaction mixtures contain ATP (2 mM), MgCl\(_2\) (2 mM), amino acid (0-10 mM), and purified AtGH3.12 (1 µg) or AtGH3.7 (10 µg). For determination of kinetic parameters, reactions were performed with 4-hydroxybenzoic acid or chorismate varied from 10-1,000 µM. All data were fit to the Michaelis-Menten equation using GraphPad Prism. Inhibition of GH3.7 and GH3.12 was analyzed using 10 µg of protein with 1 mM chorismate and 10 mM amino acid (glutamate or cysteine for GH3.12 or GH3.7, respectively) in the presence of varied (0-10 mM) salicylic acid. IC\(_{50}\) values were determined by fitting data to \(y = \frac{\text{max}}{1 + (I_L / \text{IC}_{50})}\), where max is the maximum observed rate difference and
I_f is the inhibitor concentration, in GraphPad Prism.

**Pyrophosphate Release Assays**

Adenylation assays were performed using the pyrophosphate reagent to couple pyrophosphate release to NADH depletion, with ATP (2 mM), MgCl₂ (2 mM), and purified AtGH3.12 (30 µg) with either 4-hydroxybenzoic acid or chorismate (1 mM) in the absence or presence of glutamate (10 mM). Assays were performed at 25 °C and were initiated by the addition of enzyme.

**Mass Spectrometry**

Reaction assays were performed at 25 °C with 1 mM ATP, 1 mM MgCl₂, 1 mM chorismate, 5 mM glutamate, 50 mM Tris pH 8.0, and 100 µg recombinant AtGH3.12 in 500 µL. Control reactions without AtGH3.12 were also performed. Reactions were quenched with 50% acetonitrile and 0.1% acetic acid. Samples were run through LC-MS/MS as previously described (Chen et al., 2009).

**X-ray Crystallography**

Crystals of AtGH3.12 in complex with AMP and chorismate were grown by the vapor diffusion method in hanging drops of a 1:1 mixture of protein (15 mg/mL) and crystallization buffer (15% PEG-3350, 0.25 M ammonium acetate, 100 mM sodium acetate pH 4.4, 5 mM tris(2-carboxyethyl)phosphine, 5 mM MgCl₂, 10 mM chorismate, and 4 mM AMP). Crystals were flash frozen with mother liquor supplemented with 30% PEG-3350 as a cryoprotectant. Diffraction data was collected at ESRF ID23-2 and were indexed and integrated with XDS and
scaled with XSCALE (Kabsch, 2010; Sheldrick, 2008). Molecular replacement was performed with Phaser using the previously solved AtGH3.12 structure (PDB: 4EQL) for the search model (de la Fortelle and Bricogne, 1997; Westfall and Zubieta et al., 2012). Structures were refined with Buster, and manual model building was done in COOT (Blanc et al., 2004; Emsley and Cowtan, 2004). Data collection and refinement statistics are shown in Table 1.

Rapid Chemical-Quench Kinetics of AtGH3.12

Single-turnover measurements were performed to determine $k_1$ of the AtGH3.12 reaction using a rapid chemical-quench flow instrument (RQF-3, KinTek Instruments, University Park, PA) at saturating substrate concentrations (1 mM MgCl$_2$, 1 mM chorismate or 4-hydroxybenzoic acid) with the same conditions as in the steady state reaction. The temperature was maintained at 25 °C, and a molar ratio of AtGH3.12 to ATP, including 1 µCi of [$\alpha$-$^{32}$P]ATP, was maintained at 5:1 for the reactions. The enzyme and its substrates were rapidly mixed from two sample loops (20 µL) into a reaction loop of specified dimensions based on the time of the reaction. Reactions were quenched with formic acid, after which 1 µl was spotted onto a cellulose thin-layer chromatography plate and was developed in isobutyrate: ammonium hydroxide:water (66:1:33) (Christensen and Cronan, 2009). TLC plates were analyzed by phosphorimaging and quantified using Bio-Rad Quantity One software. Reactions were repeated in triplicate, and rate data were determined by fitting the data to a one-phase decay equation using GraphPad Prism software.

Plant and pathogen materials and growth conditions

All Arabidopsis thaliana transgenic lines and mutants used in this study were in the Col-0 background. The sid2-2 mutant (Wildermuth et al., 2001) was obtained from Barbara Kunkel.
Plants were grown on soil in a growth chamber with a short-day photoperiod (8 h light/16 h dark) at 21°C with a relative humidity of 75% and a light intensity of ~130 µEinstiens sec⁻¹ m⁻¹. *Pseudomonas syringae* strain DC3000 (Cuppels, 1986) were grown on Nutrient Yeast Glycerol Medium (NYG) at 28°C (Daniels et al., 1984). The gh3.7/gh3.12 double mutant was generated using gh3.7 SALK_106726 and gh3.12 SALK_018225C lines as parents in the crosses (Alonso et al., 2003; ABRC). To confirm the phenotype of gh3.7, the SAIL_755_E09 line was also used in plant infection experiments.

**P. syringae** inoculation and quantification of bacterial growth

Five-week old *A. thaliana* plants were infected by syringe infiltrations using a solution containing 10⁵ cells mL⁻¹ in 10 mM MgCl₂ and a 1-mL needleless syringe. To quantify bacterial growth in the plant, whole leaves were sampled at days zero, two, and four post-inoculation, weighed to determine leaf mass (mg), ground in 10 mM MgCl₂, and then plated in serial dilutions on NYG media supplemented with rifampicin. Between four and six leaves were sampled per treatment.

**RESULTS**

**AtGH3.7 is a chorismate-cysteine synthetase**

Prior to this study, AtGH3.7 had escaped biochemical or genetic investigation, while GH3.12 had been determined to play a role in pathogen defense (Jagadeeswaran et al., 2007; Nobuta et al., 2007). To screen GH3.12 for activity, a previously published coupled assay was used to measure AMP release in the second half-reaction (Westfall and Zubieta et al., 2012).

A panel of 15 acyl acid hormones and hormone intermediates were screened for activity with
all 20 amino acids: 4-hydroxybenzoic acid, benzoic acid, indole-3-acetic acid, phenylacetic acid, salicylic acid, indole-3-butyric acid, gibberilic acid, abscisic acid, anthranilate, shikimate, chorismate, prephenate, 4-hydroxyphenylpyruvate, t-cinnamate, and p-coumarate (Figure 1a). Surprisingly, chorismate had the highest activity, while none of the other substrates showed activity (Table 1). Of the 20 amino acids, cysteine, serine, and glutamine had the highest activity, but further analysis by Michaelis-Menten kinetics showed that the preferred amino acid substrate is cysteine (Figure 1b). Serine and glutamine kinetics were linear to 10mM, suggesting that they have a low binding affinity and are not physiologically relevant substrates. Cysteine has a Km value of 8.4 mM, while chorismate has a Km value of 280 uM with cysteine as the amino acid substrate and a catalytic efficiency of 2600 M⁻¹ sec⁻¹ (Table 1).

<table>
<thead>
<tr>
<th>Substrates (varied, constant)</th>
<th>Km (mM)</th>
<th>kcat (min⁻¹)</th>
<th>kcat/Km (M⁻¹ sec⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chorismate, Cys</td>
<td>0.28 ± 0.06</td>
<td>45 ± 4.1</td>
<td>2600</td>
</tr>
<tr>
<td>Chorismate, Ser</td>
<td>0.23 ± 0.06</td>
<td>19 ± 2.0</td>
<td>1300</td>
</tr>
<tr>
<td>Chorismate, Gln</td>
<td>0.08 ± 0.03</td>
<td>8.4 ± 0.87</td>
<td>1700</td>
</tr>
<tr>
<td>Cys, Chorismate</td>
<td>8.4 ± 1.4</td>
<td>55 ± 5.1</td>
<td>110</td>
</tr>
<tr>
<td>Ser, Chorismate</td>
<td>&gt;10</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Gln, Chorismate</td>
<td>&gt;&gt;10</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>ATP, Chorismate &amp; Cys</td>
<td>0.3 ± 0.5</td>
<td>37 ± 1.6</td>
<td>2000</td>
</tr>
</tbody>
</table>

Table 1. Summary of steady-state kinetics of AtGH3.7.
Assays were performed and kinetic parameters determined as described in the Methods. Values are averages ± SD (n = 3).

**Kinetic reevaluation of AtGH3.12 reveals it is a chorismate-glutamate synthetase**

AtGH3.12 had been identified as a 4-hydroxybenzoic acid-glutamate synthetase using an adenylolation assay that measured pyrophosphate release in the first half-reaction of the enzyme.
Figure 1. AtGH3.7 is a chorismate-cysteine synthetase.
(a) GH3.7 was screened with 1 mM of each of 15 benzoate precursors, benzoates, and plant hormones with 10 mM each of all 20 amino acids to identify the acyl acid substrate and amino acid substrate pair. The highest activity of all was with chorismate and cysteine, so the cysteine kinetics for all 15 substrates is shown. (b) Knowing the acyl acid substrate led us to screen all 20 amino acids in triplicate with 1 mM chorismate. Cysteine still had the highest activity, followed by serine, and glutamate. Abbreviations: 4-HBA (4-hydroxybenzoic acid), PAA (phenylacetic acid), SA (salicylic acid), BA (benzoic acid), 4-HPP (4-hydroxyphenylpyruvate), IAA (indole-3-acetic acid), IBA (indole-butyric acid), GA (gibberellic acid), and ABA (abscisic acid).
(Okrent et al., 2009). We now know that this first half-reaction is likely the limiting step in catalysis and is not a true estimate of the kinetic parameters of the enzyme (Holland and Jez, Chapter 5). To reexamine AtGH3.12 substrate specificity, recombinantly purified AtGH3.12 was used to determine the kinetic parameters for 4-hydroxybenzoic acid and chorismate. Kinetic analysis revealed similar Km values for 4-hydroxybenzoic acid (270 µM) and chorismate (170 µM) but the kcat for chorismate is ~10-fold higher than 4-hydroxybenzoic acid (43 ± 3 min⁻¹ and 4.5 ± 0.4 min⁻¹, respectively) (Table 2). This leads to chorismate having a ~20-fold higher catalytic efficiency (kcat/Km) than 4-hydroxybenzoic acid (4200 M⁻¹ sec⁻¹ and 280 M⁻¹ sec⁻¹, respectively).

<table>
<thead>
<tr>
<th>Substrates (varied, constant)</th>
<th>Km (mM)</th>
<th>kcat (min⁻¹)</th>
<th>kcat/Km (M⁻¹ sec⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-HBA, Glu</td>
<td>0.27 ± 0.05</td>
<td>4.5 ± 0.4</td>
<td>280</td>
</tr>
<tr>
<td>Chorismate, Glu</td>
<td>0.17 ± 0.03</td>
<td>43 ± 3.0</td>
<td>4200</td>
</tr>
<tr>
<td>Glu, Chorismate</td>
<td>3.1 ± 0.52</td>
<td>220 ± 14</td>
<td>1200</td>
</tr>
</tbody>
</table>

Table 2. Summary of steady-state kinetics of AtGH3.12.
Assays were performed and kinetic parameters determined as described in the Methods. Values are averages ± SD (n = 3). 4-HBA= 4-hydroxybenzoic acid.

For good measure, all 20 amino acids were rescreened with chorismate as the acyl acid substrate to ensure that glutamate was indeed the correct amino acid substrate. Glutamate had the highest specific activity (2400 nmol min⁻¹ mg⁻¹), and was 4.5-fold higher than the second highest specific activity (leucine; 520 nmol min⁻¹ mg⁻¹) (Figure 2). Glutamate had a Km value of 3.1 mM with chorismate but was linear to 10 mM with 4-hydroxybenzoic acid as the acyl acid substrate.

The true substrate of AtGH3.12 was missed in previously published results because the assay only measured the first half-reaction, which is the rate-limiting step. Previous work with the
adenylation assay showed lower activity with chorismate than with 4-hydroxybenzoic acid (Okrent et al., 2009). To understand the failure of the adenylation assay in detecting the chorismate activity, we repeated the assay in the absence or presence of glutamate. Using this assay in the absence of amino acid, AtGH3.12 shows much higher activity with 4-hydroxybenzoic acid than with chorismate (120 ± 4 nmoles min⁻¹ mg⁻¹ and 24 ± 1.2 nmoles min⁻¹ mg⁻¹, respectively). If glutamate is added to the reaction mixture, the 4-hydroxybenzoic acid rate has a modest increase, 140 ± 1.5 nmoles min⁻¹ mg⁻¹, but the rate with chorismate increases drastically in the presence of glutamate, ~35-fold to 620 ± 14 nmoles min⁻¹ mg⁻¹ (Figure 3).

**Figure 2. AtGH3.12 is a chorismate-glutamate synthetase.**

AtGH3.12 was screened with 10 mM of each of the 20 amino acids with 0.5 mM chorismate (n=3).

Formation of the chorismate-glutamate conjugates were confirmed by LC-MS/MS after the assay was quenched. Indeed, a peak with a m/z of 372.1 was observed when AtGH3.12 was present but not in the control. Chorismate-glutamate is expected to have a mass of 371.1 g/mol,
Figure 3. Adenylation activity of AtGH3.12.
The adenylation rate was measured with either 4-hydroxybenzoic (4-HBA) acid or chorismate in the absence or presence of glutamate. While glutamate has minimal effect on the adenylation rate of 4-hydroxybenzoic acid, glutamate leads to a large increase in the adenylation rate of chorismate. Values shown are mean ± SD (n=3).

Figure 4. Mass spectra of chorismate-glutamate.
Reaction mixture of 1 mM chorismate, 1 mM ATP, and 5 mM glutamate with 100 µg AtGH3.12 was reacted before LC-MS analysis.
so it may be that the low pH of the method protonated the amide nitrogen to give the resulting m/z (Figure 4).

To confirm the differences in rates in the first half reaction ($k_1$), single-turnover kinetics using rapid chemical quenching was used. In this assay, the substrates were separate from the protein and were rapidly mixed before being quenched at a specified timepoint, for a total of 10 timepoints. The ratio of acyl-AMP formed to total adenosine moieties (AMP, ATP, and acyl-AMP) were plotted over time to generate the single-turnover rates with both chorismate and 4-hydroxybenzoic acid (Figure 5). The $k_1$ is 0.062 sec$^{-1}$ when 4-hydroxybenzoic acid is the substrate and is 0.989 sec$^{-1}$ when chorismate is the substrate. This confirms what was seen in the pyrophosphate release assay where about 16 molecules of 4-hydroxybenzoic acid are adenylated to every one chorismate adenylate that is formed. However, in the full reaction, the turnover rate with chorismate and glutamate is 10 times higher (43 min$^{-1}$) compared to that of 4-hydroxybenzoic acid and glutamate (4.5 min$^{-1}$).

Figure 5. Single-turnover kinetics of the first half-reaction of GH3.12.
The left panel shows the rate of the first half-reaction with 4-hydroxybenzoic acid as the acyl acid substrate, and the left panel shows the rate of the first half-reaction with chorismate as a substrate (n=3).
Salicylic acid inhibits formation of both chorismate conjugates

Previously published GH3.12 kinetic data showed that SA was a competitive inhibitor in the reaction with 4-hydroxybenzoic acid (Okrent et al., 2009). With the new data showing that 4-hydroxybenzoic acid is not the preferred substrate of GH3.12, we wanted to confirm that the reaction was still inhibited with by SA when chorismate was the acyl acid substrate. As expected, SA also inhibits AtGH3.12 when forming chorismate-glutamate, with an IC$_{50}$ of 7 µM, which is similar to what was found previously with 4-hydroxybenzoic acid (IC$_{50}$ = 15 µM) (Table 3).

Because chorismate is a precursor to SA and GH3.7 also uses chorismate as a substrate, we hypothesized that chorismate-cysteine conjugate formation by GH3.7 would also be inhibited by SA. SA is an inhibitor of GH3.7 too, but to a lesser degree. The IC$_{50}$ value is 2.4 mM, which is 340-fold higher than the IC$_{50}$ value of SA for GH3.12 with chorismate as an acyl acid substrate. This suggests that GH3.7 has relaxed SA inhibition compared to GH3.12 and is only effected by SA at higher concentrations.

<table>
<thead>
<tr>
<th>Protein, Substrate</th>
<th>IC$_{50}$ (µM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>GH3.12, Chorismate</td>
<td>7 ± 1.4</td>
</tr>
<tr>
<td>GH3.7, Chorismate</td>
<td>2400 ± 200</td>
</tr>
</tbody>
</table>

Table 3. Summary of AtGH3.7 and AtGH3.12 inhibition kinetics with salicylic acid.
Assays were performed and kinetic parameters determined as described in the Methods. Values are averages ± SD (n = 3). 4-HBA= 4-hydroxybenzoic acid.

Structural analysis of AtGH3.12 co-crystalized with chorismate in the active site

Prior to this study, three-dimensional x-ray crystal structures of AtGH3.12 had been solved in complex with AMP and with AMP and SA (PDB: 4EQL) (Westfall and Zubieta et al., 2012). Interestingly, the active site pocket seemed much larger than SA and had a shape more similar to
that of chorismate (Okrent et al., 2009; Westfall and Zubieta et al., 2012). To gain an understanding of how chorismate fits in the acyl acid binding pocket of AtGH3.12, the 1.94 Å structure was solved with chorismate and AMP (Table 4). The overall structure was the same as the previously solved structures in the closed conformation (Figure 6). Clear electron density for AMP shows that the adenylate binding site remains unchanged (Figure 7).

![Figure 6. Active site of AtGH3.12 with chorismate and AMP bound.](image)

Chorismate (green) and AMP (purple) are shown interacting with key active site residues.

Chorismate was bound in the structure with low electron density (Figure 7); however, the molecule was found in the same area as SA in the previous structures. The low electron density is likely due to low occupancy, high B-factors, or the inherent propensity of chorismate to degrade. Compared to other hormones or hormone analogues that have been co-crystalized with GH3 proteins, including jasmonyl-isoleucine, a non-hydrolyzable IAA-AMP analogue, SA, and IAA, chorismate contains many polar and charged side chains (Westfall and Zubieta et al., 2012; Round et al., 2013; Westfall and Sherp et al., 2016). Chorismate contains two carboxylate moieties, one of which binds deep in the pocket.

In the GH3.12 acyl-binding site, chorismate is oriented so that the carboxylate off of the ring is positioned closest to the phosphate of the bound AMP molecule (Figure 6); therefore, this
carboxylate is likely the position where the adenylation and subsequent amino acid transfer occurs. In the structure, this carboxylate forms charge-charge interactions with the terminal side chain nitrogen of Arg123; the molecule is further stabilized by hydrogen bonding interactions with Tyr120 and a water molecule. The hydrophobic residue Ile217 is positioned below the chorismate ring and fits into the ring pucker. The distal chorismate carboxylate is forms charge-charge interactions with Arg213 and hydrogen bonding interactions with Tyr181.

Although it does not directly interact with chorismate, Tyr178 forms charge-charge interactions with the side chain nitrogens of Arg123 to position it to interact with the ring carboxylate of chorismate. While the ring carboxylate of chorismate is 11 Å away from the AMP α-phosphate, chorismate is more likely to bind the enzyme when Mg-ATP is bound. All residues that were found to be critical in chorismate binding in the AtGH3.12 structure (Y120, R123, Y178, Y181, R213, and I217) are conserved in AtGH3.7.

Figure 7. Electron density for chorismate (left) and AMP (right) in the AtGH3.12 active site using 2F_o-F_c maps contoured to 1.0 σ.
Site-directed Mutant Analysis

To gain a better understanding of the role of active site residues in chorismate binding and catalysis, several site-directed mutants that were previously generated were re-analyzed with chorismate, including Y112F, Y120F, R123K, V159A, T161S, I217A, and F218Y (Table 5). Removing the hydrogen bond from the tyrosine hydroxyl increased the Km (290 μM) while maintaining wild-type levels of turnover, and the other residue that interacts with the chorismate ring carboxylate, Arg123, had both increased turnover and Km (70 min⁻¹ and 1300 μM) when
mutated to a lysine, which ultimately decreased the catalytic efficiency 4.6-fold (900 M⁻¹sec⁻¹).

The isoleucine (I217) that interacts with the chorismate ring was mutated to a smaller hydrophobic residue, alanine, which decreased the catalytic efficiency of the enzyme nearly 50-fold (86 M⁻¹sec⁻¹) due to the high Km value (1100 µM) and low turnover rate (9.2 min⁻¹).

With chorismate bound in the structure, several residues that were not previously analyzed emerged as being important for substrate binding and catalysis. At the distal carboxylate of chorismate, the two tyrosines (Y178 and Y181) were individually mutated to phenylalanines. The Y178F mutation, which removes the charge-charge interaction with Arg213, is inactive even though it does not directly contact the ligand in the structure. The Y181F mutant is active but has an increased Km value (500 µM) and turnover (74 min⁻¹). Finally, the R213K mutant is inactive too, suggesting that the combination of both Tyr178 and Arg213 are critical for both chorismate binding and catalysis.

<table>
<thead>
<tr>
<th>Mutant</th>
<th>Km (mM)</th>
<th>kcat (min⁻¹)</th>
<th>kcat/Km (M⁻¹ sec⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y112F</td>
<td>0.29 ± 0.03</td>
<td>42 ± 2</td>
<td>2400 ±</td>
</tr>
<tr>
<td>Y120F</td>
<td>2.3 ± 0.45</td>
<td>27 ± 4</td>
<td>200 ±</td>
</tr>
<tr>
<td>R123K</td>
<td>1.3 ± 0.10</td>
<td>70 ± 5</td>
<td>900 ±</td>
</tr>
<tr>
<td>V159A</td>
<td>0.24 ± 0.04</td>
<td>69 ± 4</td>
<td>4800 ±</td>
</tr>
<tr>
<td>T161S</td>
<td>0.28 ± 0.04</td>
<td>28 ± 1</td>
<td>1700 ±</td>
</tr>
<tr>
<td>Y178F</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Y181F</td>
<td>0.50 ± 0.01</td>
<td>74 ± 7</td>
<td>2500 ±</td>
</tr>
<tr>
<td>R213K</td>
<td>--</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>I217A</td>
<td>1.1 ± 0.16</td>
<td>9.2 ± 0.5</td>
<td>86 ±</td>
</tr>
<tr>
<td>F218Y</td>
<td>0.65 ± 0.09</td>
<td>9.2 ± 0.5</td>
<td>240 ±</td>
</tr>
</tbody>
</table>

**Table 5. Summary of steady-state AtGH3.12 mutant kinetics with chorismate.** Assays were performed and kinetic parameters determined as described in the Methods. Values are averages ± SD (n = 3).
Arabidopsis *gh3.7* mutants show wild-type pathogen response when infected with *PstDC3000*

Because GH3.7 and GH3.12 use the same acyl acid substrate, we hypothesized that the two proteins could have redundant functions *in planta*. Previous physiological investigations of *gh3.12* (or *pbs3*) mutants showed that the gene was involved in pathogen defense, and the knockout showed enhanced disease susceptibility to virulent and avirulent strains of *Pseudomonas syringae* pv. DC3000 (*PstDC3000*) (Warren et al., 1999). Also, *gh3.12* mutants are compromised in pathogen-induced SAG (SA-O-β-glucoside) accumulation, and this accumulation could be restored by exogenous application of SA, which also enhanced resistant to virulent DC3000 (Nobuta et al., 2007).

![PstDC3000 Infection of Arabidopsis Leaves](image)

**Figure 8.** Arabidopsis *gh3.7-1* does not have a pathogen infection phenotype.
Leaves from infected plants were sampled 0, 2, and 4 days post-inoculation for quantifying *Pseudomonas syringae* pv. tomato DC3000 growth in the tissue (CFUs/mg leaf). This graph depicts three biological replicates with 4-6 leaves per replicate per genotype. Statistical analyses were conducted using Tukey’s multiple comparisons test. * represents a p-value of < 0.05, and **** represents a p-value of < 0.0001 compared to wild-type.
Here we infected *gh3.7-1* mutants by leaf syringe infiltration with wild-type *PstDC3000* and found no statistically significant difference between wild-type and mutant plants (Figure 8). The *iscochorismate synthase* mutant, *sid2-2*, was included as a positive control because it shows enhanced disease symptoms and is more susceptible to pathogen infection (Wildermuth et al., 2001). This confirmed that GH3.7 has a different role than GH3.12 than in aerial defense against pathogens. There was no visible growth phenotype in the *gh3.7-1* plants compared to wild-type.

**Arabidopsis gh3.7/gh3.12 double mutants are more susceptible to infection than wild-type**

While *gh3.7-1* had no disease resistance or susceptibility phenotype, a double mutant *Arabidopsis thaliana* line was generated by crossing *gh3.7-1* and *gh3.12* mutant lines. Overall, there were no noticeable growth phenotypes between the *gh3.7-1/gh3.12* double mutant and the *gh3.12* or *gh3.7* single mutant plants, suggesting there is no additive effect (Figure 9). While chorismate-glutamate conjugates are important for pathogen resistance, chorismate-cysteine appears to not be involved. Also, to confirm that *gh3.7-1* knockout lines show wild-type pathogen responses, a second allele was included in these experiments, *gh3.7-2*. Plants were infected by syringe infiltration using wild-type *PstDC3000*, and *sid2-2* was included as a positive control.

In these experiments, *gh3.7-2* was also not significantly more susceptible or resistant to *PstDC3000* than wild-type plants, confirming that GH3.7 does not play a role in aerial pathogen defense. Therefore, GH3.7 and GH3.12, and thus chorismate-cysteine and chorismate-glutamate conjugates, do not have redundant roles in pathogen responses in infected leaf tissue.

Interestingly, the *gh3.7-1/gh3.12* double mutant was more susceptible to infection than either *gh3.12* or *sid2-2* alone. This suggests that while GH3.7 alone does not function in defense in leaf
tissues, the combination of both GH3.7 and GH3.12 does indeed aid plants in pathogen response.

**Figure 9. Arabidopsis gh3.7-1/gh3.12 double mutant lines are more susceptible to pathogen infection than wild-type or sid2-2.**

Leaves from infected plants were sampled 0, 2, and 4 days post-inoculation for quantifying *Pseudomonas syringae* pv. tomato DC3000 growth in the tissue (CFUs/mg leaf). This graph depicts three biological replicates with 4-6 leaves per replicate per genotype. Statistical analyses were conducted using Tukey’s multiple comparisons test. * represents a p-value of < 0.05, and **** represents a p-value of < 0.0001 compared to wild-type.

**DISCUSSION**

The GH3 family of proteins found in all plants is critical for modulating the concentration of acyl acid hormones in plants by conjugating them to amino acids (Westfall et al., 2013; Korasick et al., 2013). While this conjugation is activating in some instances, like with jasmonates, some amino acid-conjugated hormones are inactive storage or degradation forms of the molecule, as with indole-3-acetic acid (IAA) conjugates. Interestingly, the Arabidopsis GH3 protein AtGH3.12, or PBS3, was previously found to conjugate glutamate to the plant metabolite 4-hydroxybenzoic acid, which is not a plant hormone (Okrent et al., 2009). With this knowledge, we screened aromatic and aromatic precursor metabolites as substrates for the previously uncharacterized AtGH3.7, a homolog of AtGH3.12, and found that the enzyme is a chorismate-
cysteine synthetase (*Table 1; Figure 1*).

When GH3.12 was first discovered in plants, it was identified as having an SA-deficiency phenotype that could be rescued by the application of exogenous SA, leading the authors to speculate that GH3.12 must work upstream of SA signaling in plants (Nobuta et al., 2007). Follow-up work to that study found that AtGH3.12 had the highest adenylation activity with 4-hydroxybenzoic acid (Okrent et al., 2009). Despite the role of GH3.12 in defense against the plant pathogen *Pseudomonas syringae*, the previously identified GH3.12 substrate, 4-hydroxybenzoic acid, while known to exist in plants, has no known role in pathogen defense. Rescreening of AtGH3.12 revealed that it had higher activity with chorismate than with 4-hydroxybenzoic acid and is a chorismate-glutamate synthetase (*Table 2*). Interestingly, despite the high sequence similarity and use of the same acyl-acid substrate, SA inhibited GH3.12 with about a 340-fold lower IC$_{50}$ (7 µM) than that of GH3.7 (2400 µM) (*Table 3*).

GH3 proteins follow a two-step ping-pong mechanism where the acyl acid substrate binds and is adenylate and pyrophosphate is released, followed by a transfer step where the amino acid is added to the acyl acid to form the conjugate and AMP is released (Chen et al., 2010; Westfall et al., 2010). 4-Hydroxybenzoic acid was identified as the substrate of GH3.12 using an assay that only measured pyrophosphate release, or the first half reaction (Okrent et al., 2009). We presume this assay was used to simplify the screening process, as each acyl acid substrate must be screened with all 20 amino acids to determine the correct acyl acid and amino acid substrate combination. The assay used to identify chorismate as a substrate measured AMP release, or the full reaction.

We believe that 4-hydroxybenzoic acid was misidentified as the substrate of GH3.12 because of the pyrophosphate release assay and the rate of turnover of the adenylated intermediate.
Depending on the acyl acid substrate, the molecule will either bind the active site until the amino acid is added, or the incorrect adenylated substrate is released from the active site more quickly, giving a higher turnover value in a pyrophosphate release assay (Figure 3). Based on our results, it seems that the protein adenylates both correct and incorrect substrates, but the correct substrate is held in the active site while the incorrect substrate is released. These results were further corroborated by mass-spectrometry and single-turnover kinetics (Figures 4 & 5).

Having chorismate bound in a structure provided further evidence that it is the preferred acyl acid substrate of AtGH3.12 (Figure 6). Compared to the x-ray crystal structures of other GH3 proteins that use either IAA or JA, VvGH3.1 and AtGH3.11, respectively, key differences among the structures reveal how the AtGH3.12 active site recognizes chorismate (Westfall and Zubieta et al., 2012; Peat et al., 2012). The active site of AtGH3.11, or JAR1, has only one polar residue, a histidine, that interacts with the JA ketone through a water molecule, and VvGH3.1 has two polar residues, both tyrosines, that hydrogen bond with the amine of IAA. Unlike these two structures, the AtGH3.12 active site contains two arginines and two tyrosines, with one of each positioned near each chorismate carboxylate. As confirmed by kinetic analysis of site-directed mutants, the charge-charge interactions between the arginines and the carboxylates are critical for positioning chorismate in the active site for the adenylation and amino acid transfer reactions to occur (Table 5).

Considering that AtGH3.12 and AtGH3.7 are 79% identical based on amino acid sequence comparisons, we wondered if they would have physiologically redundant roles in planta. Because AtGH3.12 has an SA-deficiency response and both proteins use chorismate as an acyl-acid substrate, we hypothesized that the chorismate-glutamate and chorismate-cysteine conjugates may have the same role in pathogen defense. To determine the role of AtGH3.7 in
Figure 10. Chorismate is a precursor to salicylate and the three aromatic amino acids in plants and is synthesized in both the plastid and the cytosol, where GH3 proteins are found. Abbreviations: phosphoenolpyruvate (PEP); erythrose-4-phosphate (E4P); chorismate mutase (CM); isochorismate synthase (ICS); prephenate aminotransferase (PAT); arogenate dehydratase (ADT); arogenate dehydrogenase (ADH); prephenate dehydrogenase (PDH); prephenate dehydratase (PDT); hydroxyphenylpyruvate aminotransferase (HPP-AT); phenylpyruvate aminotransferase (PPY-AT).

Pathogen defense, knockout lines were infected with *P. syringae* and were not more susceptible to infection than wild-type plants, suggesting that these proteins are not redundant in Arabidopsis (Figure 8). Next, we wondered if the roles of AtGH3.7 and AtGH3.12 were additive and generated an Arabidopsis double mutant. The *gh3.7-1/gh3.12* double mutants were even more susceptible to pathogen infection that *gh3.12* mutants alone or than the positive control, the *iscochorismate synthase* knockout line, which is deficient in SA biosynthesis, *sid2-2* (Figure 9). This confirmed that having both functional proteins is important in Brassicaceae species for defense against pathogenic microorganisms.

Knowing the substrate of AtGH3.12 presents several possibilities for the role of this metabolite *in planta*. In the absence of this protein, plants are more susceptible to pathogen
attack and were found to have less SAG accumulation after infection (Jagadeeswaran et al., 2007; Nobuta et al., 2007). This data, as suggested by others who have studied this protein, suggests that the metabolite is upstream of SA biosynthesis. One possibility is that chorismate-glutamate is siphoning chorismate away from other biosynthetic pathways, including that of the auxins indole-3-acetic acid and phenylacetic acid, and funneling it towards SA biosynthesis; however, a hydrolase that would release free chorismate remains to be identified (Figure 10). Another possibility is that chorismate-glutamate is a substrate for isochorismate synthetase and is an intermediate metabolite in SA biosynthesis. A third option is that chorismate-glutamate is a signaling molecule that regulates SA biosynthetic genes or proteins. At this point, all of these possibilities are purely speculative.

AtGH3.12 and AtGH3.7 are cytoplasmic proteins, while chorismate is produced in the plastid of plants, as is SA (Figure 10). A chorismate transporter on the chloroplast membrane has yet to be identified. Another possibility is that the amino acid-conjugated chorismate is recognized by a dipeptide transporter and imported into the chloroplast during infection, the amino acid is cleaved, and chorismate is then converted into SA.

Interestingly, AtGH3.12 and AtGH3.7 are not the only cytosolic enzymes that use chorismate as a substrate; chorismate mutase 2 in Arabidopsis is found in the cytosol and converts chorismate into prephenate, the precursor to tyrosine and phenylalanine via the arogenate pathway (Eberhard et al., 1996; Mobley et al., 1999; Colquhoun et al., 2010; Westfall et al., 2014). The chorismate-conjugates could be removing chorismate from the cytosolic chorismate mutase that would convert it into prephenate. Or, one form of the conjugate, such as chorismate-cysteine, for example, could be involved in preventing cytosolic chorismate into prephenate, while the other conjugate is involved in transport of chorismate into the plastid. For now, there
is no way of confirming any of these hypotheses, and future studies are needed to decipher the role of these metabolites in pathogenesis in Brassicaceae.

In conclusion, we have identified chorismate, a hormone biosynthetic intermediate, as the substrate of the Brassicaceae-specific clade of GH3 proteins that contains AtGH3.7 and AtGH3.12. Because chorismate is upstream of a number of plant hormones, including indole-3-acetic acid (auxin), phenylacetic acid, and salicylic acid, conjugating chorismate to an amino acid may act to regulate the biosynthesis of these downstream hormones simultaneously. While AtGH3.12 was previously shown to be involved in pathogen defense responses, the role of the chorismate-glutamate conjugate in defense remains unknown. Also, the physiological role of AtGH3.7 remains to be understood outside of its role alongside AtGH3.12 in pathogen defense response. Regulation and modification of chorismate, the aromatic amino acid and salicylic acid precursor, plays a bigger role in plant biotic stress responses than previously thought.

ACKNOWLEDGEMENTS

This work was supported by NSF grant MCB-1614539 to J.M.J. C.K.H. was supported by the NSF Graduate Research Fellowship Program (DGE-1143954), and C.S.W. was supported by a U.S. Department of Agriculture–National Institute of Food and Agriculture pre-doctoral fellowship (MOW-2010-05240). Portions of this research were carried out at the European Synchrotron Radiation Facility and the Proteomics and Mass Spectrometry Facility at the Donald Danforth Plant Science Center.
REFERENCES


Dean JV, Mills JD (2004) Uptake of salicylic acid 2-O-β-D-glucose into soybean tonoplast vesicles by an ATP-binding cassette transporter-type mechanism. Physiol Plant. 120, 603–612


Feys BJF, Benedetti CE, Penfold CN, Turner JG (1994) Arabidopsis mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. *Plant Cell* 6, 751-759


Staswick PE, Tiryaki I, Rowe ML (2002) Jasmonate response locus JAR1 and several related *Arabidopsis* genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. *Plant Cell* 14, 1405-15


217
Chapter 8

Conclusions and Future Directions
ABSTRACT

The overarching goal of my thesis research was to understand how plants regulate the biosynthesis of aromatic amino acids, which are essential to both protein synthesis and the production of downstream secondary metabolites, including phytohormones. This was accomplished by structural and biochemical studies of prephenate dehydrogenase, arogenate dehydrogenase, chorismate mutase, and prephenate aminotransferase. Also, biochemical assays were used to unveil the basis of substrate fidelity in the GH3 proteins. Finally, the role of aromatic amino acid biosynthesis pathway intermediates was further explored when AtGH3.7 and AtGH3.12 were found to conjugate amino acids to the aromatic branchpoint metabolite chorismate. This research as a whole has advanced our understanding of the regulation of plant aromatic amino acid and hormone metabolic pathways, showing how plants have evolved specific regulatory mechanisms and networks that are more complex than previously thought.
Chapter 2: Summary of the regulation of soybean prephenate and arogenate dehydrogenases

Despite its role in protein synthesis and as a precursor to many plant secondary metabolites, the evolution of the alternative route for tyrosine biosynthesis in legumes remained unknown prior to this study. The enzyme in the alternate tyrosine biosynthetic pathway, prephenate dehydrogenase (PDH), was only discovered four years ago (Schenck et al., 2014). This enzyme catalyses the tyrosine-insensitive oxidative decarboxylation of prephenate to form 4-hydroxyphenylpyruvate (4-HPP), a precursor to tyrosine. PDH is legume-specific, cytosolic, and uninhibited by tyrosine, unlike the other dehydrogenase enzyme in the tyrosine biosynthetic pathway, arogenate dehydrogenase (ADH).

To unravel the molecular mechanisms underlying the evolution of substrate specificity and tyrosine regulation, a combination of x-ray crystallography, sequence analysis, and biochemistry were used to investigate PDH and ADH enzymes (Schenck and Holland et al., 2017). The three-dimensional structure of PDH from soybean (GmPDH1) was solved and refined to 1.7 Å with the cofactor NADP⁺ bound. When comparing the active site residues from GmPDH1 to the residues that would be found in the active site of other PDHs or in ADHs, two residues stood out as being conserved in ADHs and variable in PDHs, Met219 and Asn222. Asn222 in GmPDH1, was an aspartate in the ADHs, while Met219 was a threonine.

Site-directed mutagenesis of the PDH residue to the ADH residue (N222D) was sufficient to switch the substrate preference of the enzyme from prephenate to arogenate and to introduce tyrosine inhibition. The reciprocal mutation in the soybean ADH enzyme (D218N) introduced prephenate activity and relieved tyrosine inhibition.
Three-dimensional structures of GmPDH1 N222D and N222D/M219T mutants showed that the structural architecture was conserved; the only changes occurred at the single residue level. In the GmPDH1 N222D/M219T mutant structure, tyrosine was bound in the active site. This showed that the backbone amine of tyrosine was hydrogen bonding with the carboxylate oxygen of Asp in the 222 position. This structure was used for molecular modeling to dock arogenate into the active site. Based on the docking results, the amine in arogenate would form dominant ionic interactions with the carboxylate of Asp. The amine in arogenate is a carbonyl in prephenate, which would hydrogen bond to the amide of Asn in the wild-type GmPDH1 structure.

This study demonstrated that alteration of an acidic residue, Asp218, in an ADH enzyme to an uncharged residue was critical for the neofunctionalization of the duplicated gene into a PDH. Not only is this residue essential in differentiating prephenate from arogenate, it is also involved in substrate inhibition. By removing the ability to form a charge-charge interaction with the substrate in the aspartate to asparagine mutation, tyrosine sensitivity was also relieved. Therefore, a single residue in the active site of a plant ADH was critical in the evolution of tyrosine-insensitive PDH activity in legumes.

**Chapter 2: Future directions**

Now that we understand the molecular foundation underlying the evolution of the alternative tyrosine biosynthetic pathway, many questions remain concerning the physiological role of PDH *in planta*. What is the advantage in legumes for having an alternative route to tyrosine biosynthesis, or is there any advantage at all? To address this, knockout and/or knockdown lines of PDH should be generated in a legume and analyzed for changes in growth, tyrosine content,
and downstream tyrosine-derived metabolites. Similarly, legume overexpression lines should also be used to identify phenotypes associated with PDH.

Because PDH is specific to legumes, which form root nodules to associate with symbiotic nitrogen-fixing bacteria, and tyrosine biosynthesis requires the transfer of an amine group onto a tyrosine precursor, there may be a role for PDH in legume-rhizobia interactions. Phenotyping knockout and overexpression of PDH in legumes should include a careful analysis of nodule formation and colonization by rhizobia.

One idea for the function of cytosolic PDH in plants is that it is responsible for producing 4-hydroxyphenylpyruvate (4-HPP) as a precursor to specialized metabolites and is not necessarily producing tyrosine, although tyrosine can be made from 4-HPP. 4-HPP-derived metabolites include tocopherols, like vitamin E, and the electron carrier plastoquinone (Schenck et al., 2014). Metabolite and transcript analysis of PDH knockout lines compared to the wild-type legume would give insight into the role of PDH in specialized metabolite biosynthesis. Additionally, gene expression analysis throughout the plant life cycle and with biotic and abiotic stressors may provide further insight into the role(s) of PDH in legumes.

Currently the transcriptional and post-transcriptional regulation of PDH in plants is not known. Studying the promoter sequence and comparing it to the promoters of other genes in the aromatic amino acid pathway and other secondary metabolite gene promoters may give rise to information about the regulation of PDH expression under normal physiological growth conditions and under biotic and abiotic stress conditions. It may also be possible to mine existing transcriptomic and proteomic datasets in legumes to see if PDH is regulated in published studies.

Having a highly efficient, deregulated enzyme for increasing tyrosine and tyrosine-derived metabolites could be useful from a metabolic engineering or synthetic biology standpoint. This
enzyme could be heterologously expressed in microorganisms or other plants to generate metabolites that use tyrosine as a precursor, such as the alkaloid morphine. I expect that introducing PDH into a plant that does not already contain a functional PDH would greatly increase the pool of tyrosine, especially if using a constitutive promoter.

Chapter 3: Summary of the enzymatic mechanism of soybean prephenate dehydrogenase

The legume-specific, cytosolic alternative tyrosine biosynthetic pathway involves the key enzyme prephenate dehydrogenase (PDH). While aromatic amino acid biosynthesis was traditionally thought to mirror microbial pathways, the recent discovery and structural characterization of PDH in plants highlights how this pathway in legumes evolved from a plant arogenate dehydrogenase instead of from a gene transfer from a microorganism (Schenck and Holland et al., 2017). Additionally, microbial PDHs are inhibited by tyrosine, while plant PDHs are insensitive (Sun et al., 2006; Sun et al., 2009). Enzymatically, PDH is highly efficient with a catalytic efficiency of $60,500 \text{ M}^{-1}\text{s}^{-1}$ and turnover of $980 \text{ min}^{-1}$. Taken together, this suggested that PDH could be a desirable engineering target for herbicide resistance to 4-HPP dioxygenase-targeted herbicides.

The previous crystal structure of GmPDH1 from soybean (*Glycine max*) shed light on the mechanisms regulating substrate specificity and inhibition by tyrosine, and a single active site residue (Asn222) was found to mediate both. This residue is an aspartate in plant ADHs, which use arogenate as a substrate and are inhibited by tyrosine. Outside of Asn222, the roles of the rest of the active site residues in both binding and catalysis remained to be investigated. Computational docking of prephenate into the active site of the previously solved GmPDH1 structure guided us to potential residues to probe for active site contributions.
In this study, nineteen active site point mutants were generated to assess the mechanism of soybean PDH1 (S101A, H124A, H124N, H124Q, Q130A, Q130D, Q130E, Q130S, T131A, T131S, Q184A, Q184D, H188A, H188N, H188Q, T206A, T206S, N222A, and N222Q) (Holland and Jez, 2018). Mutants were generated by site-directed mutagenesis, expressed in E. coli, and purified for enzymatic comparison to the wild-type enzyme. Several mutants, including H124A, H124Q, Q184D, H188A, H188N, H188Q, T206A, and N222A, were inactive, highlighting the critical roles of these active site residues in binding and catalysis.

Overall, the active site mutants were substantially impaired in catalytic efficiency, ranging from 0.02 to 14% of the catalytic efficiency of the wild-type enzyme and 4.3 to 890-fold reductions in turnover. The catalytic histidine, H124N had substantially decreased catalytic efficiency compared to the wild-type enzyme, which was unsurprising as this residue is positioned to transfer a hydride from prephenate to NADP⁺.

The degree of conservation among PDHs in plants was further analyzed by comparing the soybean PDH1 to common bean (Phaseolus vulgaris) and alfalfa (Medicago truncatula). While key catalytic residues were conserved (Ser101, His124, Gln184, His188, and Thr206), residues varied in the carboxylate-binding pocket (Gln130 and Thr131). Comparisons between microbial arogenate dehydrogenases (ADHs) (Ochrobactrum intermedium LMG 3301, Sediminspirochaeta smaragdinae DSM 11293, and Methanosaeta harundinacea) and soybean PDH1 also revealed that several residues were conserved (Ser101, His124, Gln184, and His188), while carboxylate-binding pocket residues (Gln130 and Thr131) varied.

Our site-directed mutagenesis and steady-state kinetic analysis of point mutants targeting each active site residue that contacts a ligand was useful in proposing a catalytic mechanism for the efficient conversion of prephenate to 4-HPP. His124 facilitates hydride transfer, while
His188, Gln184, and Thr206 assist with prephenate decarboxylation at C1. Gln130 and Thr131 stabilize the pyruvyl group carboxylate to orient the carboxyl towards the residue responsible for substrate selectivity and tyrosine inhibition, Asn222. Ser101 contributes to the coordination of the cofactor, the substrate, and the His124 hydrogen-bonding network to aid in hydride transfer.

Overall, these findings suggest that the active site of GmPDH1 has been highly conserved over the course of evolution, as small changes lead to large disruptions in catalytic efficiency. As for engineering applications, nature has already generated the best version of this enzyme, and our mutagenesis approach did not enhance enzyme activity.

Chapter 3: Future directions

A mechanistic perspective of the recently discovered prephenate dehydrogenase from soybean was presented using prephenate docked into the previously solved crystal structure of PDH1. However, a three-dimensional structure of GmPDH1 with prephenate bound in the active site remains to be solved. Despite this, the focus of the study was on the mechanism of the enzyme, which has now been resolved.

A future direction of this project would be to investigate a microbial PDH, which evolved separately from the plant PDH, and conduct similar mechanistic studies. While the catalytic residues in microbial PDHs have been identified, outside of the catalytic histidine and an arginine involved in prephenate decarboxylation (a histidine in plant PDHs), the contributions of other active site residues remain to be explored (Sun et al., 2006; Sun et al., 2009).

Another consideration is that our investigation only tested the mutants’ ability to use prephenate as a substrate; however, in the case of the GmPDH1 N222D mutant, arogenate was the preferred substrate (Schenck and Holland et al., 2017). Arogenate can be generated using
prephenate aminotransferase and purified from the enzyme using HPLC. Because this study focused on the enzymatic mechanism, arogenate was not included as a substrate. We expect that the mutants would not have activity with arogenate, but this was never confirmed. The ability to use other pathway keto acids, such as 4-hydroxyphenylpyruvate,phenylpyruvate, or chorismate, was also not tested in this investigation. Future work into engineering this enzyme may focus on the transcriptional regulation or tissue-specific localization rather than on metabolic engineering of the enzyme itself. There is great potential for introducing PDH into a plant outside of legumes to increase the pool of tyrosine and 4-HPP-derived metabolites.

Chapter 4: Summary of the regulation and evolution of chorismate mutases from early land plants

At the interface of plant primary metabolism and aromatic amino acid biosynthesis is the shikimate pathway, which generates the precursor to all three aromatic amino acids, chorismate (Maeda and Dudareva, 2012; Tzin and Galili 2010). To generate prephenate, the precursor to phenylalanine and tyrosine, the enzyme chorismate mutase (CM) catalyzes the pericyclic Claisen rearrangement of chorismate (Sträter et al., 1997). In Arabidopsis, there are three CMs, two of which are plastidial and allosterically regulated (AtCM1 and AtCM3) and one of which is cytosolic and unregulated (AtCM1) (Westfall et al., 2014).

In the biochemical study of the Arabidopsis CMs, a phylogenetic analysis revealed a basal lineage of plant enzymes homologous to AtCM1 based on sequence similarity, including chorismate mutases from the bryophyte (moss) Physcomitrella patens, the lycophyte Selaginella moellendorffii, and the basal angiosperm Amborella trichopoda (Westfall et al., 2014). To understand the origins and molecular basis of allosteric regulation in the plant chorismate
mutases, enzymes from *P. patens* (PpCM1 and PpCM2), *Am. trichopoda* (AmtCM1 and AmtCM2), and *S. moellendorffii* (SmCM) were characterized using steady-state kinetics and x-ray crystallography (Kroll and Holland et al., 2017).

All five enzymes from early plants had turnover rates (15 to 22.8 s\(^{-1}\)) similar to those of CM1 and CM3 from Arabidopsis (16 s\(^{-1}\) and 13 s\(^{-1}\), respectively). The catalytic efficiencies of these five basal plant CMs were comparable to AtCM3 (\(k_{\text{cat}}/K_m = 11,800 \text{ M}^{-1} \text{s}^{-1}\)) and were lower than either AtCM1 or AtCM2 (29,300 and 258,000 M\(^{-1}\) s\(^{-1}\), respectively). Additionally, the role of all twenty amino acids as effectors were tested for each of the five basal chorismate mutases. Tryptophan was an activator of all five CMs, while histidine was a weak activator for PpCM1 and AmtCM1. Tyrosine and phenylalanine were negative effectors of the CMs tested except for SmCM, and AmtCM1 was only inhibited by tyrosine and not phenylalanine.

To further our understanding of the evolution of regulation in these early plant chorismate mutases, the three-dimensional x-ray structure of PpCM1 was solved in complex with tryptophan at 2.0 Å. This portion of the work was completed by an undergraduate mentee, Kourtney Kroll, who was co-first author on this manuscript. Overall, the all α-helical structure was similar to AtCM1, and analysis of the active site and effector sites at the dimer interface and comparisons between the plant amino acid sequences in these sites revealed both common and varied features. Because PpCM1 was solved in complex with an activator, and the previous structures of AtCM1 were in complex with inhibitors, this structure provided a snapshot of the structural changes associated with transitioning from the T-(less active) to the R-(more active) states of the plant CMs.

The use of a sequence-structure-function analysis allowed us to characterize chorismate mutases from several evolutionary ancestors to modern angiosperms. While some characteristics
were conserved between the basal CMs and the yeast enzyme, some features were also shared with the Arabidopsis isoforms.

Chapter 4: Future directions

Despite our new knowledge on the evolution of allostERIC regulation of CMs in basal plants, many questions remain as to the physiology of these enzymes. Although these enzymes have now been characterized \textit{in vitro}, an \textit{in vivo} analysis should be undertaken in the future to confirm the predicted localizations. Seed plants typically have two or more chorismate mutates genes in the genome, with one cytosolic and one plastidial isoform. We predicted the localization of these enzymes based on the presence or absence of a chloroplast transit peptide, but that does not give us definitive evidence as to where these enzymes function in the cell. Localization data would be necessary to confirm the chloroplast or cytosolic location of these enzymes.

Similarly, these experiments raise the long-held question of why plants have a cytosolic CM if aromatic amino acids are synthesized in the chloroplast of plants. The downstream enzyme, prephenate aminotransferase, has only been found in the chloroplast of plants (Dornfeld et al., 2014). Additionally, it is unclear how chorismate makes it to the cytosol; is it transported from the chloroplast to the cytosol? A transporter has yet to be identified. Then, is prephenate transported back into the chloroplast to be used as a substrate for the next aromatic amino acid biosynthetic enzyme? While a plastidial tyrosine and phenylalanine transporter has been identified in plants, a prephenate transporter has not been identified (Widhalm et al., 2015).

Interestingly, there is some evidence for 3-dehydroquinate dehydratate/shikimate dehydrogenase-2 from \textit{Nicotiana tabacum}, a second isoform of the third and fourth steps of the shikimate pathway, being a cytosolic enzyme (Ding et al., 2007). Also, isoforms of 3-deoxy-D-
arabino-heptulosonate 7-phosphate synthase and 5-enolpyruvylshikimate 3-phosphate synthase (ESPS), the first and sixth steps of the shikimate pathway, respectively, lack N-terminal plastid transit peptide sequences (d’Amato et al., 1984; Mousdale and Coggins, 1985). Both ESPS synthase and shikimate kinase, the last enzyme in the shikimate pathway, have been found to maintain activity even in the absence of transit peptides (Della-Cioppa et al., 1986; Schmid et al., 1992), which some suggest could mean that these two enzymes could be constituents of a cytosolic shikimate pathway (Toghe et al., 2013).

If there is an alternative cytosolic shikimate pathway in plants, then chorismate would be synthesized in the cytosol for use by chorismate mutase to generate prephenate. Aromatic amino acid biosynthetic enzymes that localize to the cytosol include the legume-specific prephenate dehydrogenase (Schenck et al., 2014), tyrosine:phenylpyruvate aminotransferase (Yoo et al., 2013) and phenylalanine:4-hydroxyphenylpyruvate aminotransferase (Bedewitz et al., 2014). It may also be that alternative cytosolic pathways are unique to some plants but not all. How and if plants synthesize aromatic amino acids in the cytosol remains an open question.

**Chapter 5: Summary of the regulation of prephenate aminotransferase**

Prior to this study, the only investigations of prephenate aminotransferase (PAT) in plants involved the evolution of this gene from a lateral gene transfer from Chlorobi/Bacteroidetes microorganisms and the results of knock-downs in petunia and tobacco (Maeda et al., 2011; de la Torre et al., 2014; Dornfeld et al., 2014). Another recent study identified two residues in the Arabidopsis PAT that introduced a gain of function when mutated to valines—Thr84 and Lys169 (Dornfeld et al., 2014). The T84V/K169V PAT double mutant not only had altered keto acid preference for 4-hydroxyphenylpyruvate over prephenate, but it also preferred tryptophan over
aspartate as an amino donor. These altered substrate specificities are what initially led us to investigate this enzyme in greater detail.

To begin to understand the architectural changes behind the changes in substrate specificity, the wild-type PAT structure from *A. thaliana* was solved and refined to 3.0 Å with the cofactor pyridoxal-5’-phosphate bound, which led us to the active site of the enzyme. Further crystallization of the T84V mutant and the T84V/K169V double mutant showed the dynamics of the enzyme; the mutant structures were in a closed conformation, while the wild-type structure was in an open conformation. A structure of a mutant where the catalytic lysine was mutated to an alanine (K306A) was also solved and was in the closed conformation. The only difference between these two conformations was the movement of the α1 helix, where Thr84 was located.

Ligands bound in the active sites of these structures included pyradoxamine-5’-phosphate and glutamate in the PAT K306A structures, α-ketoglutarate in the PAT T84V structure, and borate and malate from the crystallization buffer bound in the PAT T84V/K169V structure. Having both keto and amino acid substrates bound in the active site allowed for the identification of residues that bind substrates and propose a reaction mechanism. Also, the wild-type structure was used to dock arogenate, the amino acid product of PAT, into the active site.

Because PAT is bifunctional and can accept α-ketoglutarate as a keto acid substrate with aspartate as the amino donor, we compared residues found in the PAT active site to those found in the active sites of the five aspartate aminotransferases (AATs) found in *A. thaliana*. Of the active site residues, only five varied between PAT and AATs: Thr84, Glu108, Ala189, Lys169, and Arg445. The PAT residues were individually mutated to the AAT residues (E108K, A168G, K169S, and R445G), with the exception of Thr84 which had been previously characterized and was predicted to be absent in the AATs. Of all of the mutants, only wild-type, T84V, E108K, and
A168G showed activity in the steady-state kinetic assay. One mutant in particular, E108K, stood out as having wild-type catalytic efficiency with α-ketoglutarate and virtually no activity with prephenate. Therefore, this single residue was critical in the evolution of prephenate activity in PAT.

Because the previously generated PAT mutants had altered amino acid preferences, all 20 proteogenic amino acids were tested as effectors. Surprisingly, L-cysteine was an inhibitor of PAT with an IC\(_{50}\) of 0.75 mM, a physiologically relevant concentration for the plant cell. Because cysteine is synthesized in the chloroplast, mitochondria, and cytosol in plants, it makes sense that cysteine would be found in millimolar concentration in the chloroplast where PAT is found. Cysteine also inhibited PAT activity in snapdragon (\textit{Antirrhinum majus}) and a Gram-negative thermophilic green sulfur bacterium (\textit{Chlorobium tepidum}) with IC\(_{50}\) values of 1.3 mM. This led us to conclude that cysteine is an effector of both plant and microbial PATs.

Cysteine was previously found to upregulate chorismate mutase 3 in \textit{A. thaliana} as a positive effector, and this is the second occurrence of cysteine as a regulator of aromatic amino acid biosynthesis. Overall, the structure-function-sequence analysis of PAT from plants and microorganisms led to key findings that are critical to our understanding of amino acid metabolism and plant metabolic regulation.

\textbf{Chapter 5: Future directions}

Because an embryo-lethal mutant in a screen mapped to PAT, the only genetic investigations of this gene used either RNA interference or virus-induced gene silencing to study the \textit{in vivo} role of PAT in aromatic amino acid biosynthesis (Pagnussat et al., 2005). Therefore, the only
way to truly characterize this gene was using a biochemical approach; however, this study has now given rise to new ideas that would require further investigation.

Although we identified a new potential effector of PAT, L-cysteine, further *in vivo* experiments are needed to reveal the role of cysteine in aromatic amino acid regulation. One such experiment would include analyzing flux of prephenate to the three aromatic amino acids or aromatic amino acid-derived products in the presence of cysteine. Because Arabidopsis inflorescence shoots have rapid rates of growth and lignification, cysteine could be fed to these tissues, and the transcriptional and metabolic responses to high cysteine could be measured over time. It is odd that cysteine is a positive effector of CM3, increasing the pool of downstream prephenate, while cysteine is a competitive inhibitor of PAT, which uses prephenate as a substrate; it seems that under high cysteine concentrations, the chloroplast pool of prephenate would increase, and there is not a known prephenate-using enzyme in the chloroplast besides PAT. Our understanding of this role of cysteine in its connection to prephenate is incomplete.

Part of the reason a biochemical and structural characterization of PAT was undertaken was because of the interesting findings by Dornfeld and colleagues that two potential active site mutations introduced a gain-of-function, where the new mutant protein functioned as a tyrosine aminotransferase (2014). To understand the importance of these two residues, three crystal structures were solved and compared, including wild-type, the T84V mutant, and the T84V/K169V double mutant; however, the data did not give us direct insight as to how this protein gained new function. We do not know how the individual active site residues contribute to the overall protein dynamics and kinetics. Computational molecular dynamics simulations of PAT through the ping-pong reaction may answer this question.
Additionally, none of the structures presented in our research had the substrate or product, prephenate or arogenate, respectively, bound in the active site. Despite our attempts, diffraction quality crystals would not form in the presence of these compounds. Currently neither prephenate nor arogenate have been co-crystalized in any of the three-dimensional protein structures found in the Research Collaboratory for Structural Bioinformatics Protein Data Bank. Additionally, we were not able to capture either the cyclized or non-cyclized cysteine-PLP adduct in any of the structures to confirm cyclization. Therefore, we cannot definitively say this occurs except by evidence from studies of similar aminotransferases (Liu et al., 2013). One way to potentially confirm whether the PLP-cysteine adduct is cyclic or non-cyclic may be to purify the adduct from the enzyme reaction and use mass-spectrometry and NMR to determine its chemical structure.

By comparing the amino acid sequence of Arabidopsis PAT to the five Arabidopsis AATs, we identified a single residue that was involved in the enzymes ability to recognize prephenate as a substrate. In future studies, the role of this residue in the AATs should be investigated to determine if prephenate activity could be introduced into these proteins by the reciprocal mutation (Lys→Glu). Also, these AATs should be analyzed for cysteine inhibition since they also use PLP as a cofactor and are highly similar to PAT; the same could be done for plant tyrosine aminotransferases.

While we identified a potential negative effector of PAT, the transcriptional and post-translational modifications that govern this enzyme also remain unknown. Because this enzyme was purified from *E. coli* and not from the plant, it is likely that there are modifications that we do not see. Also, no studies have looked at the transcript levels of PAT over time and with various biotic and abiotic stressors. Because both phenylalanine and tyrosine are required for
protein biosynthesis, and phenylalanine is a precursor to lignin, we would expect that this gene is turned on during times of intense growth and is turned down when protein breakdown contributes more to the tyrosine and phenylalanine pools. Because tyrosine and phenylalanine are precursors to defense metabolites, such as salicylic acid, cyanogenic glycosides, and isoquinoline alkaloids, PAT could also be expressed when plants are under attack by pathogens or herbivores. However, this is purely speculation, and the specific roles of PAT during growth and defense should be explored through gene expression analysis. In conclusion, this data answers the question of how the active site of an enzyme evolved both AAT and PAT activity and reveals new insights into how tyrosine and phenylalanine metabolic pathways are regulated by another amino acid, cysteine.

Chapter 6: Summary of the substrate fidelity of auxin-using GH3 proteins

GH3 proteins, or acyl acid-amido synthetases, are responsible for regulating hormone concentrations in plants by conjugating acyl acid hormones to amino acids as a means of activating or inactivating the molecule (Staswick et al., 2002; Staswick et al., 2005). Amino acid-hormone conjugates are generated using a two-step ping-pong mechanism that involves both an adenylation and transfer reaction (Staswick et al., 2002; Staswick et al., 2005; Terol et al., 2006; Chen et al., 2010; Westfall et al., 2010). In the first half-reaction, the acyl acid substrate is adenylated, and pyrophosphate is released (Chen et al., 2010; Westfall et al., 2010), followed by a transferase reaction where an amino acid displaces AMP to form the conjugate.

The largest class of GH3 proteins are the auxin-using GH3s, which includes AtGH3.2 and AtGH3.5 (Westfall and Zubieta et al., 2012). Using amino acid sequence alignments and the x-ray crystal structure of AtGH3.5, we noticed that the active site residues that would be in contact
with the acyl acid ligand were the same in both AtGH3.2 and AtGH3.5 (Westfall and Sherp et al., 2016). Despite having identical active sites and both using aspartate as an amino acid substrate, AtGH3.5 uses multiple acyl acid substrates (indole-3-acetic acid (IAA), phenylacetic acid (PAA), and benzoic acid (BA)) with almost equal catalytic efficiencies, while AtGH3.2 is fairly specific for IAA.

Using single-turnover kinetics with labeled ATP, the rate of the first half-reaction, $k_1$, was determined for GH3.2, GH3.5, and GH3.11, a jasmonate-conjugating GH3 protein, with IAA, BA, and PAA for the auxin-using enzymes and JA for GH3.11 using a TLC assay. When visualizing the formation of adenylated acyl acids on the TLC plates, we noticed that AMP was accumulating over time in the absence of amino acid, which suggested that the adenylated intermediate may be hydrolyzed if the amino acid is not present for the transferase reaction to occur. This rate of hydrolysis ($k_{\text{AMP}}$) was reported with all of the substrates for each of the GH3 proteins analyzed. Overall, the data suggested that IAA is the preferred substrate for both GH3.2 and GH3.5, and the PAA-AMP and BA-AMP intermediates are hydrolyzed at a faster rate than the IAA-AMP intermediate.

GH3 proteins have a large N-terminal domain and a smaller C-terminal domain that pivots between an open and closed position depending on whether AMP or ATP is bound in the active site (Westfall and Zubieta et al., 2012; Round et al., 2013). Because the amino acids found in contact with the acyl acid substrate in AtGH3.2 and AtGH3.5 are the same, we wondered how distal residues influenced substrate specificity and the rate of hydrolysis of the adenylated intermediate. To study this, chimeric proteins were generated by swapping the C-terminal domains of GH3.2 and GH3.5. Steady state kinetics revealed that GH3 N5/C2 (N-terminal domain of GH3.5 with the C-terminal domain of GH3.2) was inactive, while GH3 N2/C5 was
active with IAA, PAA, and BA as substrates. However, our studied failed to confirm that the GH3 N5/C2 protein was properly folded, which could have been the reason for the lack of activity. Single-turnover experiments with GH3 N2/C5 showed that the protein functioned similarly to GH3.2 and has a higher $k_1$ value with IAA as the acyl acid substrate.

Based on these data, the auxin-using GH3 proteins have different rates for the first half-reaction based on the acyl acid substrate and hydrolyze adenylated intermediates in the absence of an amino acid substrate. The ratio of $k_1$ to $k_{AMP}$ for each of the acyl acid substrates shows that a BA-AMP or PAA-AMP is more likely to be hydrolyzed than an IAA-AMP, suggesting that IAA is the true substrate of these promiscuous enzyme. We hypothesize that in the absence of the amino acid substrate, the active site remains in the open conformation when a noncognate substrate binds, allowing solvent to access the adenylated intermediate and cleave the bond. However, when a cognate substrate binds, the enzyme closes in anticipation of the transferase reaction. An alternative hypothesis is that the enzyme is able to cleave the noncognate adenylate intermediate. This single-turnover kinetic analysis of auxin-using GH3 proteins broadens our understanding of the reaction chemistry of an important hormone-modifying class of enzymes and sheds light on the promiscuity of the enzyme in substrate selection.

Chapter 6: Future directions

While we now have a deeper understanding of the kinetic basis of substrate selection in the auxin-using GH3 protein family, we still do not know the exact reaction sequence for intermediate hydrolysis and if hydrolysis is enzyme-mediated or solvent-mediated. It is unlikely that the enzyme would encounter this scenario in vivo since amino acids are presumed to be in the millimolar concentration range in plant cells. Also, only amino acids have been tested as the
second substrate for these enzymes, and it may be that they are also able to use other cellular metabolites as the second substrate. Nonetheless, it is unlikely in vivo that a GH3 proteins second substrate would be limiting in a cell.

The concentration of the acyl acid substrate, however, would be expected to fluctuate in a cell over time. Only about 1% of IAA is found in a free form, with about 90% in an amide linkage and about 10% in an ester linkage (Normanly et al., 1993; Tam et al., 2000). Therefore, the concentration of available IAA is much lower than IAA in a modified form; the concentrations of free versus modified BA and PAA are not known. Along those same lines, it is unclear what the physiological roles of PAA and BA conjugates are in plants. The conjugates were previously confirmed to be found in plants using mass spectrometry (Westfall and Sherp et al., 2016). It is only assumed that modifying PAA takes it out of the pool of PAA that is able to bind the TIR1 receptor and that modifying BA likely regulates the concentration of downstream metabolites, like salicylic acid, aromatic cytokinins, ubiquinone, folic acids, and other specialized metabolites (Widhalm and Dudareva, 2015).

An interesting thought about the evolution of this large family of hormone-modifying enzymes is that, like all proteins, we are catching them in the act of evolving. It does not seem unreasonable to think that as the homologs evolve, they will become more different from each other and one may become more promiscuous than the other, which is what we see for GH3.5. While it does prefer IAA in the single-turnover experiments, if there is less selective pressure for strict substrate selectivity, these proteins may have even further relaxed substrate preferences in the future as they continue to evolve.

Also, the hydrolases for PAA-Asp and BA-Asp have not yet been identified. It may be that the IAA-specific hydrolases are also able to release the hormones from these amino acid
conjugates to regulate physiological responses. A combination of genetic and biochemical experiments would be necessary to identify and confirm the presence of these hydrolases. Despite all of these unknowns, this study provided resolution for the question of substrate selection (and promiscuity) in a physiologically important class of hormone-modifying enzymes in plants.

Chapter 7: Summary of the chorismate-conjugating GH3 proteins AtGH3.7 and AtGH3.12

This study investigated the substrates and physiological roles of two acyl acid-amido synthetases that are specific to Brassicaceae plants—AtGH3.7 and AtGH3.12 (or PBS3). Prior to this investigation, GH3.12 was identified as a 4-hydroxybenzoate-glutamate synthetase using an assay that measured pyrophosphate release in the first half-reaction. While most GH3 proteins are acyl acid hormones, like jasmonate and indole-3-acetic acid, 4-hydroxybenzoate is neither a plant hormone nor a major plant metabolite. It seemed unlikely that this little-known compound was the substrate of this protein. A substrate for GH3.7 had remained to be identified, but it shares 79% of its amino acid sequence with GH3.12 and likely has similar function. Based on a published protein phylogeny of the GH3s that used only the acyl acid binding pocket residues, GH3.7 and GH3.12 are the only members in this clade (Westfall and Zubieta et al., 2012).

Using an enzyme assay that measured AMP release in the full reaction (versus measuring pyrophosphate release from the first half-reaction), the aromatic amino acid precursor chorismate was found to be the acyl acid substrate of these two proteins. While GH3.12 conjugates chorismate to glutamate, GH3.7 conjugates chorismate to cysteine. The substrate of GH3.12 was likely misidentified because of the assay conditions of the earlier publication (Okrent et al., 2009). When conducting the pyrophosphate release assay in the absence of the amino acid, we
too saw that there is higher activity with 4-hydroxybenzoate than with chorismate; however, when glutamate was included in the assay, chorismate had about 35 times more activity than 4-hydroxybenzoic acid. Steady-state kinetics, single turnover kinetics, and mass spectrometry further confirmed these results. Additionally, an x-ray crystal structure of AtGH3.12 was solved in complex with chorismate to identify key residues for binding and catalysis, and site-directed mutagenesis of those residues revealed interactions that are necessary for catalysis, which includes three tyrosines and two arginines for hydrogen bonding and charge-charge interactions, respectively.

Arabidopsis gh3.12 (or pbs3) was initially identified from a mutant screen for its increased susceptibility to both virulent and avirulent strains of the plant pathogen Pseudomonas syringae (Warren et al., 1999) and was found to have decreased accumulation of pathogen responsive gene transcripts and salicylate-glucosides (SAG) after infection (Jagadeeswaran et al., 2007; Nobuta et al., 2007). Because this phenotype could be rescued by exogenous SA application, gh3.12 was believed to be upstream of salicylic acid biosynthesis in plants. In vitro assays showed that both AtGH3.7 and AtGH3.12 were inhibited by SA, but chorismate-glutamate formation was inhibited by 7 µM SA, while chorismate-cysteine formation was inhibited by 2.4 mM SA, which may not be a physiologically relevant concentration of SA.

To test if AtGH3.7 had redundant functions as AtGH3.12 in planta, gh3.7 knockout lines were infected with Pseudomonas syringae pv. tomato DC3000. Interestingly, these lines showed wild-type responses to pathogen infection and were not more susceptible to disease unlike gh3.12 mutants. Because GH3.12 and GH3.7 are the only proteins in their clade and are specific to Brassicaceae, we wanted to address the combined role of these two proteins in plant pathogen response. To do this, an Arabidopsis gh3.7/gh3.12 double mutant was generated and infected
with *P. syringae*. Interestingly, this double mutant was more susceptible to disease than the *iscochorismate synthase (sid2-2)* mutant negative control, suggesting that the combination of these two proteins is critical for pathogen defense in Arabidopsis. Taken together, this data provides evidence that the GH3 protein family in Brassicaceae is able to conjugate amino acids to the aromatic amino acid precursor chorismate as a way of regulating multiple growth and defense pathways.

**Chapter 7: Future directions**

Despite our new understanding of the substrates of this Brassicaceae-specific GH3 protein family, many questions remain to be answered. We now know that GH3.12, or PBS3, is a chorismate-glutamate synthetase, and the physiological data suggests that this protein is upstream of salicylic acid biosynthesis. While chorismate is a precursor to salicylic acid through isochorismate, it is not clear how the conjugate participates in SA biosynthesis. Because SA is an incredibly potent inhibitor of GH3.12, this protein could act to regulate SA biosynthesis by conjugating chorismate, the SA precursor, to an amino acid when SA levels are low. When SA levels increase during pathogen attack, SA is needed to regulate pathogen defense genes, and GH3.12 would be inhibited.

There are two possible scenarios for the fates of the chorismate conjugates based on what we know about other amino acid conjugates. One is that the chorismate-amino acid conjugates are a storage or degradation form of the molecule that take chorismate out of the active pool, preventing the synthesis of downstream amino acids (tyrosine, phenylalanine, and tryptophan) and hormones (indole-3-acetic acid, salicylic acid, and phenylacetic acid) simultaneously. In this case, the chorismate conjugates would be regulating multiple hormones at once. The enzyme that
would free chorismate from the conjugate, a hydrolase, is not known. The second scenario is that the conjugates are an active molecule. It could be that chorismate-glutamate is a pathway intermediate in SA biosynthesis, potentially as a substrate for isochorismate synthetase, but there is no concrete evidence for this. It may be of interest to test either experimentally or computationally through molecular docking and molecular dynamics simulations if chorismate-amino acid conjugates are able to bind the known NPR (nonexresser of pathogenesis-related) SA receptors.

Interestingly, this research also begs the question of whether or not chorismate or the chorismate conjugates are plant hormones. While it seems more likely that it is not a ubiquitous plant hormone, since this clade of GH3 proteins is specific to Brassicaceae, it is curious that these plants find that modifying chorismate is a worthwhile physiological investment. While we do not know the function of these conjugates, we also do not know where these conjugates are stored in the plant cells or if they are transported long distances. A long-term goal of this project could be to take a forward genetics approach to identifying Arabidopsis mutants where a phenotype is restored when grown on chorismate or a chorismate-conjugate. The inherent unstable nature of chorismate would make such an experiment challenging, and because it is a screen, there is no promise that such a mutant would be identified.

Because GH3 proteins were initially identified as auxin-responsive genes, there is still a component of auxin/SA crosstalk that remains unknown. The direct regulation of these chorismate-conjugating GH3 proteins either transcriptionally or post-translationally are also not known. To understand the role of these proteins on a global scale, RNAseq experiments in combination with metabolic analysis would give a clearer picture of the broader role of these chorismate-conjugating GH3 proteins in plant metabolism.
Additionally, the function of AtGH3.7 in plants is still not understood. While it does aid GH3.12 in pathogen defense, preliminary localization data suggests that this protein is found in the roots of plants. It could be that GH3.7 is specific for defense against root pathogens, but this has not been tested. Also, many GH3 protein knockouts do not have obvious phenotypes, but the overexpression lines do. Arabidopsis lines that are overexpressing gh3.12 have been generated, and gh3.7 overexpression lines are at the T2 stage now. Once these plants are homozygous, they will be analyzed for obvious growth phenotypes that would shed light on the role of these proteins in plant development. Also, an Arabidopsis gh3.7 native promoter:YFP line has been generated and is also at the T2 stage now. While there are many experiments left to confirm the role of chorismate-amino acid conjugates in plants, this study was instrumental in shedding light on the possibility of modulating the concentrations of multiple plant hormones using a single upstream metabolite as a regulator.

Conclusions of the thesis

The goals of my dissertation research were to contribute to our understanding of how plants synthesize and regulate metabolites that are essential to plant physiology. For my research, I took a biochemical and structural approach to studying pathway enzymes to revisit key aromatic amino acid biosynthetic enzymes that had been largely ignored. It was assumed that microorganisms and plants shared many key features in common in tyrosine and phenylalanine biosynthesis, but it turns out that plants have taken a different approach in many cases.

One key theme that kept appearing in my research is the role of cysteine in modifying aromatic amino acid precursors and metabolism (Figure 1). Taken together, my data show that
Plants synthesize the three aromatic amino acids, Tyr, Phe, and Trp, in the chloroplast of all plants. PDH is a cytosolic, legume-specific pathway to tyrosine that is not found in all plants. A functional PDT remains to be characterized. GH3.12 and GH3.7 and cytosolic, Brassicaceae-specific proteins, and GH3.12 is thought to be upstream of salicylic acid in plants based on genetic work. The role of GH3.7 in plants is unknown. Abbreviations: Cysteine (Cys); isochorismate synthase (ICS); phosphonoyl pyruvate (PEP); erythrose-4-phosphate (E4P); chorismate mutase (CM); prephenate aminotransferase (PAT); arogenate dehydratase (ADT); arogenate dehydrogenase (ADH); prephenate dehydrogenase (PDH); prephenate dehydratase (PDT); phenylpyruvate aminotransferase (PPY-AT); 4-hydroxyphenylpyruvate (HPP-AT).

Figure 1. Overview of the regulatory network of aromatic amino acid metabolic pathways. High levels of cysteine in the plastid would increase the pool of chorismate, since cysteine is a positive effecter of chorismate mutase, and would decrease the conversion of prephenate to arogenate by PAT. Furthermore, chorismate was conjugated to cysteine in the cytosol by GH3.7, suggesting an added layer of aromatic/cysteine crosstalk in plants. Another consideration is that all of these enzymes are found in Brassicaceae species; while PAT was found to be inhibited by cysteine outside of Arabidopsis, the early plant CMs were not affected by cysteine, and CMs outside of Arabidopsis have not been investigated biochemically. This cysteine-regulation may be specific to the mustard family.
Another sulfur-containing class of metabolites in mustards are the glucosinolates, which are plant defense compounds that are either aromatic, indolic, or aliphatic. Because these sulfur-containing metabolites, which give mustard its bitter taste, the presence of these and other sulfur metabolites could be regulating the concentration of sulfur, including in the form of cysteine, in plants. It may also be that cysteine-mediated regulation of aromatic amino acids and their precursors is a result of redox states in the cell.

Based on the idea that GH3.12 may be upstream of SA biosynthesis, this may also be an example of a hormone regulation that evolved in Brassicaceae that is as yet unknown in other plant families. Chorismate-amino acid conjugates have not yet been identified in a plant metabolite analysis yet, and it may be that chorismate is regulated by other mechanisms outside of Brassicaceae.

Overall, what we do know is that we have many more gaps in knowledge left to fill before we can fully understand how plants synthesize and regulate important protein and metabolite precursors. More so, it may end up that, as is the case with other metabolites, aromatic amino acid pathways and their regulatory mechanisms in Arabidopsis may vary from other plants and may be more of the exception and less of the rule. While my goal was to advance our understanding of plant metabolism, future discoveries are needed before we can fully appreciate and understand the complexity of plant metabolism and its regulation in its entirety.
REFERENCES


Okrent RA, Brooks MD, Wildermuth MC (2009) Arabidopsis GH3.12 (PBS3) conjugates amino acids to 4-substituted benzoates and is inhibited by salicylate. J. Biol. Chem. 284, 9742-54


Staswick PE, Tiryaki I, Rowe ML (2002) Jasmonate response locus JAR1 and several related Arabidopsis genes encode enzymes of the firefly luciferase superfamily that show activity on jasmonic, salicylic, and indole-3-acetic acids in an assay for adenylation. *Plant Cell* 14, 1405-15


Appendix I

Identification of a noroxomaritidine reductase with
Amaryllidaceae alkaloid biosynthesis related activities

Matthew B. Kilgore\textsuperscript{1,2}, Cynthia K. Holland\textsuperscript{2}, Joseph M. Jez\textsuperscript{2}, Toni M. Kutchan\textsuperscript{2}

\textsuperscript{1}Donald Danforth Plant Science Center, St. Louis, MO 63132
\textsuperscript{2}Department of Biology, Washington University in St. Louis, St. Louis, MO 63130
Identification of a Noroxomaritidine Reductase with Amaryllidaceae Alkaloid Biosynthesis Related Activities

Matthew B. Kilgore, Cynthia K. Holland, Joseph M. Jez, and Toni M. Kutchai

From the Donald Danforth Plant Science Center, St. Louis, Missouri 63132 and the Department of Biology, Washington University in St. Louis, St. Louis, Missouri 63130

Amaryllidaceae alkaloids are a large group of plant natural products with over 300 documented structures and diverse biological activities. Several groups of Amaryllidaceae alkaloids including the hem anthamine- and crinine-type alkaloids show promise as anticancer agents. Two reduction reactions are required for the production of these compounds: the reduction of norcausograndine to norbelladine and the reduction of noroxomaritidine to normaritidine, with the enantiomer of noroxomaritidine dictating whether the derivatives will be the crinine-type or hemanthamine-type. It is also possible for the carbon-carbon double bond of noroxomaritidine to be reduced, forming the precursor for maritamine or elwesine depending on the enantiomer reduced to an oxomaritamine product. In this study, a short chain alcohol dehydrogenase/reductase that co-expresses with the previously discovered norbelladine 4′-O-methyltransferase from Narcissus sp. and Galanthus sp. was cloned and expressed in Escherichia coli. Biochemical analyses and x-ray crystallography indicates that this protein functions as a noroxomaritidine reductase that forms oxomaritamine from noroxomaritidine through a carbon-carbon double bond reduction. The enzyme also reduces norcausograndine to norbelladine with a 400-fold lower specific activity. These studies identify a missing step in the biosynthesis of this pharmacologically important class of plant natural products.

The family of bulbous plants known as the Amaryllidaceae includes ornamental plants such as members of the genera Narcissus (daffodils) and Galanthus (snowdrops). These plants produce a group of alkaloids with a range of potential pharmaceutical uses. Of the Amaryllidaceae alkaloids, galanthamine has been used clinically for treatment of Alzheimer disease symptoms (1, 2). In addition to galanthamine, the chemically diverse Amaryllidaceae alkaloids have many documented pharmacological activities (3). For example, several have cancer fighting potential including hemanthamine, crinine, and lycorine (2, 4, 5). Continued development of Amaryllidaceae alkaloids for clinical and pharmaceutical applications requires the establishment of a cost effective production system for these molecules; however, limited information about the genes and proteins involved in the biosynthesis of Amaryllidaceae alkaloids is available.

The biosynthesis of all Amaryllidaceae alkaloids, norbelladine, which is made from the condensation of tyramine and 3,4-dihydroxybenzaldehyde followed by imine reduction, is a pathway intermediate (6–8). 3,4-Dihydroxybenzaldehyde is hypothesized to originate from a branch of the phenylpropanoid pathway with demonstrated intermediates trans-cinnamic acid, 4-hydroxycinnamic acid, and either 3,4-dihydroxycinnamic acid or 4-hydroxybenzaldehyde (9). The conversion of 3,4-dihydroxycinnamic acid to 3,4-dihydroxybenzaldehyde could be catalyzed by an enzyme encoded by a parologue of vanillin synthase. Vanillin synthase converts ferulic acid to vanillin (10). This reaction is similar to the proposed conversion of 3,4-dihydroxycinnamic acid to 3,4-dihydroxybenzaldehyde via a hydratase/lyase-type mechanism. An analogous, non-oxidative pathway is proposed for benzaldehyde biosynthesis. Reactions similar to the β-oxidative and non-oxidative CoA-dependent pathways documented in the biosynthesis of benzaldehyde-type compounds including vanillin and 4-hydroxybenzaldehyde are also a possibility (11). The condensation of tyramine and 3,4-dihydroxybenzaldehyde to the Schiff-base norcausograndine is analogous to known amine-aldehyde condensing enzymes including norcoacanthine synthase and strictosidine synthase (12, 13).

Considering the simple chemistry required for norcausograndine formation, it is possible that the condensation reaction occurs either nonenzymatically or in the active site of the enzyme that reduces norcausograndine to norbelladine. Enzyme superfamilies capable of catalyzing reductions similar to the reduction of norcausograndine to norbelladine include the aldo-keto reductases and short-chain alcohol dehydrogenase/reductases (SDR)s (14, 15). An example of an imine reductase from the alcohol dehydrogenase branch of the SDR superfamily functions as a tetrahydropalmatine synthase in Catharanthus roseus (16). Modification of norbelladine then leads to multiple Amaryllidaceae alkaloids.

8 This work was supported in part by National Institutes of Health Grant 1R2GM092561 from the NIH and National Science Foundation Grant DBI-0521520 for acquisition of the QTRAP LC-MS/MS. A patent has been filed for the sequence of noroxomaritidine reductase. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
9 This article contains supplemental Table S1.
10 The atomic coordinates and structure factors (codes SF85, SF79, and SF75) have been deposited in the Protein Data Bank (http://wwwpdb.org/).
1 Supported by the National Science Foundation Graduate Research Fellowship Program Grant DGE-1143954.
2 To whom correspondence should be addressed: 975 N. Warson Rd., St. Louis, MO 63132. Tel.: 314-587-1473; Fax: 314-587-1573; E-mail: bkutchai@danforthcenter.org.
Identification of Noroxomaritidine Reductase

Norbelladine undergoes methylation and phenol-coupling reactions to produce the vast diversity of Amaryllidaceae alkaloids, including galanthamine, lycorine, hem anthamine, and crinine (Fig. 1) (1, 2). Recent studies have identified enzymes catalyzing steps in the core Amaryllidaceae alkaloid biosynthesis pathway, including norbelladine 4′-O-methyltransferase (N4OMT) and the cytochrome P450 CYP96T1 (44). N4OMT catalyzes the 4′-O-methylation of norbelladine to 4′-O-methylnorbelladine (17). CYP96T1 is a para-para′ phenol-phenol coupling cytochrome P450 that generates an enantiomeric mixture of noroxomaritidine (18), which can be further modified into hemanthamine- or crinine-type alkaloids depending on the stereochemistry of noroxomaritidine (19–21). CYP96T1 also converts 4′-O-methylnorbelladine to N-demethylmarxwedine, but not as efficiently as it forms noroxomaritidine (18). N-Demethylmarxwedine generated by CYP96T1 may contribute to the synthesis of galanthamine-type alkaloids.

The enzymes that reduce noroxomaritidine to nor maritidine (a precursor of the common hemanthamine-type alkaloids) or oxomaritamine (a precursor of less prolific alkaloids including maritamine and elwesine) remain to be identified.

In previous work, co-expression analysis of the Narcissus sp. aff. pseudonarcissus, Galanthus sp., and Galanthus elwesii transcriptomes successfully identified N4OMT and CYP96T1 in the core Amaryllidaceae alkaloid biosynthetic pathway (17, 18). For several missing steps in the hemanthamine, crinine, maritamine, and elwesine pathways, we hypothesized that members of either aldo-keto reductase or SDR enzyme families may catalyze norbelladine synthesis and/or reduction of the noroxomaritidine enantiomers.

Here we identify a N. sp. aff. pseudonarcissus SDR gene that co-expresses with the previously discovered N4OMT (17). Biochemical analysis of the recombinant protein reveals that this N. sp. aff. pseudonarcissus SDR reduces the carbon-carbon double bond of the Amaryllidaceae alkaloid noroxomaritidine to form one of the enantiomeric forms of oxomaritamine. The enzyme can also produce norbelladine from a solution containing 3,4-dihydroxybenzaldehyde and tyramine with a spe-
Identification of Noroxomartidine Reductase

FIGURE 2: Phylogenetic relationship of NR to other plant reductases and dehydrogenases. Amino acid sequences of all of the enzymes were taken from NCBI (IP08 3AE2, O07106.1, AT3E07.3, AKR00286.1, AAC48976.1, ABW24051.1, and AC579634.1). The phylogenetic tree was constructed using MEGA (40), and the amino acid sequences were aligned using MUSCLE. The LG model was used to construct a Maximum Likelihood tree using a bootstrap replicate of 500. The percent identity of the plant reductases and dehydrogenases used in the tree were calculated using the Needleman-Wunsch Global Alignment tool from NCBI with the gap costs set at an existence of 11 and an extension of 2.

cific activity above background, but still 400-fold lower than noroxomartidine conversion. Competition experiments and x-ray crystal structures provide insight on how the N. sp. aff. pseudonarcissus SDR binds tryamine and piperonal, an analogue of 3,4-dihydroxybenzaldehyde, and potentially noroxomartidine. Given the high specific activity for noroxomartidine reduction relative to norcamphor reduction, we designate this gene/protein as noroxomartidine reductase (NR).

Results

Identification and Cloning of NR—In previous work, the assembly of a de novo transcriptome for N. sp. aff. pseudonarcissus with 106,451 sequences and a workflow for the identification of Amaryllidaceae alkaloid biosynthetic pathway genes from the 9,505 contigs that co-express with the accumulation of galanthamine was described (17). This resource was used to identify candidate oxidoreductases in the pathway. Because N4OMT catalyzes a reaction in the middle of the pathway, we reasoned that candidate oxidoreductase genes would co-express with the N4OMT gene. When interrogating for N4OMT co-expressing oxidoreductase homologues, 36 transcripts were identified in the N. sp. aff. pseudonarcissus transcriptome.

Of these 36 transcripts, medp_9arc_20101112.58880 was found to co-express with N4OMT in the previously described ABySS and MIRA assemblies for N. sp. aff. pseudonarcissus, G. elwesi, and Galanthus sp. (17, 18). This transcript was also co-expressed with N4OMT in the previously described Trinity-based assembly for Galanthus sp. (18). This made medp_9arc_20101112.58880 distinctive because all the other 36 transcripts were found to co-express with N4OMT in 3 or less of these assemblies. Because the medp_9arc_20101112.58880 was incomplete in the 5’ and 3’ regions, a full-length homolog (medp_9arc_20101112.14238) was used to design primers for PCR amplification. A full-length cDNA was obtained with the GenBank accession number KT295569 that differed from the medp_9arc_20101112.58880 at four nucleic acids (A141G, T183C, C243G, and T300C) with one nonsynonymous change (D81E).

The identified clone from N. sp. aff. pseudonarcissus encodes a 271-amino acid protein (NR) with a predicted molecular mass of 29.0 kDa. Sequence comparison indicates that NR is a member of the SDR enzyme family (Fig. 2). Comparison with representative members of the SDR family shows that NR shares 52% sequence identity to tropine reductase II from Datura stramonium and less than 20% identity with Papaver somniferum salutaridine reductase, C. roseus alcohol dehydrogenase, Linum usitatissimum pinosiolin-lariciresinol reductase, Medicago sativa isoflavone reductase, C. roseus tetrahydroaloinone synthase, and Digitalis lanata progesterone 5β-reductase.

Noroxomartidine Reduction Catalyzed by NR—During the biosynthesis of several prominent alkaloids including hemizanthine in N. sp. aff. pseudonarcissus, the ketone group of noroxomartidine is reduced to a corresponding alcohol (Fig. 1). The biosynthesis of the Amaryllidaceae alkaloids maritamine or elwesine requires the reduction of the carbon-carbon double bond in noroxomartidine, potentially through oxomartidine (Fig. 1). Because some members of the SDR enzyme family catalyze similar reduction reactions, for example, the reductions of tropine or 5β-progesterone (22, 23), we examined the ability of NR to reduce noroxomartidine. For biochemical and structural studies, the NR protein from N. sp. aff. pseudonarcissus was expressed as an N-terminal His$_6$-tagged protein and purified by affinity and size exclusion chromatography.

Noroxomartidine was reduced by NR in the presence of NADPH as indicated by the presence of product +2 m/z from the noroxomartidine substrate (Fig. 3A). In assays lacking either noroxomartidine or NADPH, no product formation was observed. Similarly treated TALON resin purified protein extracts from Escherichia coli transformed with empty pet28a vector showed no activity. Incubations using NADH instead of NADPH showed no conversion of noroxomartidine. Due to limited substrates the enzymatic activities were only quantified by specific activity measurements. The specific activity of NR for reduction of noroxomartidine was 8600 ± 1250 pmol min$^{-1}$ mg of protein$^{-1}$ (Table 1).

The noroxomartidine substrate is an enantiomeric mixture. The enantiomers were therefore resolved by chiral chromatography, as previously described for CYP96T1 (18). Following a 2-h incubation, only one of the noroxomartidine enantiomers was notably consumed and assays lacking NADPH or the NR enzyme did not show this preferential substrate disappearance (Fig. 4A). More than half of the enantiomeric mixture was never consumed, indicating that the enzyme acts preferentially on one of the two enantiomers (Fig. 4B). Optically pure standards of known configuration are lacking, so the absolute configuration of
Identification of Noroxomartidine Reductase

FIGURE 3. Noroxomartidine reduction by NR. Enzymatic activity of NR for reduction of noroxomartidine (m/z 274.3) was monitored using LC-MS/MS with the same LC time program found in Ref. 18A. A, enzyme assays of NR with 100 mM sodium phosphate buffer, pH 7.0, 100 μM noroxomartidine, 1 mM NADPH, and 10 μg of pure protein in 100 μL at 30 °C for 2 h. Traces from top to bottom are: noroxomartidine standard; complete assay with NR, noroxomartidine, and NADPH; assay without NADPH; assay without noroxomartidine; and assay with TALON resin purified E. coli empty vector protein extract substituted for NR protein. B, EPI MS/MS of NR product. C, EPI MS/MS of noroxomartidine reduced with NaBH₄.

TABLE 1

| Substrate | Product monitored | Specific activity (m/z)  
|------------|-----------------|-------------------|
| Noroxomartidine | 9-O-Demethyloxomartamine | 219.1, 204.1, 190.1, 176.1, m/z 165.0 and 151.0.  
| 8.9-O-Dimethyloxomartidine | 9-O-Demethyloxomartamine | 219.1, 204.1, 190.1, 176.1, m/z 165.0 and 151.0.  
| 3,4-Dihydroxybenzaldehyde and tyramine | Norbelladine | 219.1, 204.1, 190.1, 176.1, m/z 165.0 and 151.0.  
| Isovanillin and tyramine | 3',4'-O-Methyl-norbelladine | 219.1, 204.1, 190.1, 176.1, m/z 165.0 and 151.0.  
| Vanillin and tyramine | 3',4'-O-Methyl-norbelladine | 219.1, 204.1, 190.1, 176.1, m/z 165.0 and 151.0.  
| Piperonal and tyramine | 4,4'-((1,3-Benzoxodiox-5-yl)-oxy)trimethylphenol | 219.1, 204.1, 190.1, 176.1, m/z 165.0 and 151.0.  

* Measurements are with 500 μg substrate and 1 mM NADPH.

A Product made but not quantified. Values shown are mean ± S.D. for n = 3.

the preferred substrate remains unknown. Comparison of the MS/MS fragmentation patterns of noroxomartidine reduced by NR (Fig. 3B) with noroxomartidine reduced by sodium borohydride (Fig. 3C) reveals a clear difference. Aldehyde and ketone double bonds are the typical substrates of sodium borohydride (24). Although carbon-carbon bonds can be reduced by sodium borohydride, the ketone will preferentially be reduced. For this reason, product generated by the sodium borohydride reduction is likely normartidamine. The 1–2 carbon-carbon double bond is, therefore, likely reduced by NR into oxomartamine.

To examine the ability of NR to accept substrates similar to noroxomartidine, the demethylated form 8.9-O-dimethyloxomartidine was tested as substrate. 8.9-O-Dimethyloxomartidine was reduced by NR yielding a product similar to oxomartamine in its mass spectrum, indicating that the product is also reduced at the carbon-carbon double bond and therefore is 9-O-demethyloxomartamine. Fragments that are identical for these two molecules include m/z 191.1, 94.1, 108.1, and 136.1. Fragments that represent the presence of an additional methyl group in oxomartamine compared with 9-O-demethyloxomartamine are m/z 246.1 and 232.1, m/z 219.1 and 205.1, m/z 204.1 and 190.1, m/z 190.1 and 176.1, and m/z 165.0 and 151.0. Identification of Norcaugogodine Reduction Catalyzed by NR—In addition to the reduction of noroxomartidine, NR was found to catalyze the reduction of the imine in norcaugogodine to form norbelladine, albeit with a substantially lower specific activity. The amine aldehyde condensation of tyramine and 3,4-dihydroxybenzaldehyde to form norcaugogodine can spontaneously occur in solution (Fig. 1). It is not known if a biosynthetic enzyme in this pathway acts on tyramine and 3,4-dihydroxybenzaldehyde as substrates or the pre-condensed substrate norcaugogodine to make norbelladine.

For the assays examining reduction to norbelladine, tyramine and 3,4-dihydroxybenzaldehyde were incubated with NR and NADPH and the resulting product examined by LC-MS/MS. A peak with the same retention time as authentic norbelladine was observed in the complete reaction mixture (Fig. 5A). In the absence of tyramine and 3,4-dihydroxybenzaldehyde, no norbelladine was observed. Assays lacking either NADPH or NR or assays replacing NR with E. coli proteins that co-purify
Identification of Noroxomaritidine Reductase

**A**

(10bS,4aR and 10bR,4aS)-Noroxomaritidine standard  

NR assay  
No NADPH  
No substrate  
No enzyme  
Vector control  

Time (min)  
2 4 6 8 10 12 14 16 18 20 22 24 26 28

**B**

(10bS,4aR)-Noroxomaritidine  
(10bR,4aS)-Noroxomaritidine  
NR?  
(10bR,4aR)-Oxomaritinamine  
(10bS,4aS)-Oxomaritinamine  

**FIGURE 4.** NR consumption of 20 μM (10bS,4aR or 10bR,4aS)-noroxomaritidine using the same assay conditions as for specific activity assays incubated 3 h. A, LC-MS using a Chem-Tech, Inc., Chiral-CB H 100 × 4.6-mm, 5-μm column and the same LC setup and time program as for chiral separations in Ref. 18. Samples were monitored at m/z 272.3 with CE (25) and DP (70) on the OTRAP 6500 for the following samples top to bottom: (10bS,4aR and 10bR,4aS)-noroxomaritidine mixed standard; complete assay with NR, (10bS,4aR and 10bR,4aS)-noroxomaritidine and NADPH; assay without NADPH; assay without (10bS,4aR and 10bR,4aS)-noroxomaritidine substrate; assay without NR enzyme; and assay with TALON resin purified E. coli empty vector protein extract substituted for NR protein. B, the two enantiomers of noroxomaritidine and corresponding expected products.

**FIGURE 5.** Norcaugosidine reduction to norbelladine by NR. Enzymatic activity of NR for reduction of norcaugosidine was monitored using the same assay conditions and LC-MS/MS setup as in specific activity assays. A, MRM (m/z 260.1/238.0) of NR assays. Traces from top to bottom are: norbelladine standard; complete assay with NR, tyramine, 3,4-dihydroxybenzaldehyde, and NADPH; assay without NADPH; assay without tyramine and 3,4-dihydroxybenzaldehyde; assay without enzyme; and assay with TALON resin purified E. coli empty vector protein. B, MRM (m/z 260.1/238.0) relative quantification of assays shown in panel A in triplicate. C, EPI MS/MS of norbelladine standard. D, EPI MS/MS of NR product.
Identification of Noroxomaritodine Reductase

**FIGURE 6.** A comparison of the tyramine (ICE 15), DP 60, m/z 138.1/121.0 and m/z 138.1/93.0, levels in daffodil organs (above ground leaf (agl), below ground leaf (bg), leaf scale (ls), bulb scale (bs), bulb core (bc), root (r), flower stalk above ground (fsg), flower stalk below ground (fsg), and flower (fl)). These samples were collected during the months of January (well developed roots with well developed flower primordium), March (all plants form this month lack flower primordia due to sampling limitations), April (giant all flowering), and August (dormant bulbs). The instrument used for the experiment was a Quattro Premier XLS with a Phenomenex Luna 5 μm C18 (2) 250 × 4.60-mm LC column using the following LC method: A = 0.1% formic acid and B = acetonitrile, start with 10% B for 2 min followed by a linear increase to 40% B for 9 min, linear increase to 90% B for 0.1 min, hold for 3 min, linear decrease to 10% B for 0.1 min, and hold 7 min. The following was also monitored by MRM across these conditions but were not detected 3,4-dihydroxybenzaldehyde (ICE 15), DP 60, m/z 139.0/111.0 and 139.0/93.0, norbelladine (ICE 15), DP 60, m/z 260.1/138.0 and 260.1/121.0, noroxomaritidine (ICE 25), DP 70, m/z 272.3/259.0 and 272.1/212.1), and oxomaritidine (ICE 20), DP 60, m/z 274.1/219.1 and 274.1/136.1). Alkaloid samples were prepared by utilizing liquid nitrogen, a mortar and pestle for grinding of frozen tissue, followed by 70% ethanol extraction, PTFE membrane filtration, extract drying, and extract resuspension in mobile phase.

During the TALON protein purification resulted in low levels of norbelladine probably resulting from a background reaction between tyramine and 3,4-dihydroxybenzaldehyde (Fig. 5D); however, these levels were below that observed in the complete assay mixture containing enzyme. Comparison of the MS/MS fragmentation patterns of authentic norbelladine (Fig. 5C) and the product of the NR reaction (Fig. 5D) confirm the identity of the enzymatic product. Given the low abundance of the product formed by NR, assays with varied pH and temperature were performed to maximize enzymatic yields relative to the chemical background reaction. Final assay conditions for norbelladine formation by NR were pH 6.0 at 35°C with a specific activity of 21.2 ± 1.2 pmol min⁻¹ mg of protein⁻¹ (Table 1).

Variation of the aldehyde paired with tyramine in the NR-catalyzed reaction was also examined (Table 1). In place of 3,4-dihydroxybenzaldehyde, vanillin, isovanillin, and piperonal were used for incubations with NR, tyramine, and NAADPH and led to formation of 3′-O-methylnorbelladine, 4′-O-methylnorbelladine, and 4-(2-(1,3-benzodioxol-5-ylmethyl)amino)ethylphenol, respectively (Table 1). The structures of the enzymatic products were confirmed by comparing LC-MS/MS fragmentation patterns to the products produced by chemical condensation and subsequent reduction of vanillin, isovanillin, or piperonal with tyramine in methanol containing sodium cyanoborohydride. The specificity of the NR-catalyzed reaction with isovanillin and tyramine was determined to be 4-fold lower (5.33 ± 0.21 pmol min⁻¹ mg of protein⁻¹) than the reaction with 3,4-dihydroxybenzaldehyde and tyramine (Table 1). Other rates were not of direct interest to the synthesis of Amaryllidaceae alkaloids and so were not determined. These results indicate that NR can reduce an imine in the conversion of norcaugosidine to norbelladine but the low specific activity could indicate this is a nonspecific reaction.

**Inhibition of Noroxomaritodine Reductase**—The reduction of noroxomaritidine provided the opportunity to test the occupancy of the active site of NR by 3,4-dihydroxybenzaldehyde and tyramine. The IC₅₀ value of tyramine is 5.230 ± 752 μM. Piperonal has a lower IC₅₀ value (425 ± 62 μM) than 3,4-dihydroxybenzaldehyde (4,960 ± 559 μM). This could result from either the increased stability of piperonal relative to 3,4-dihydroxybenzaldehyde or a tighter binding to the active site. Although these values are high, concentrations of tyramine can be up to 3 mM in N. sp. aff. pseudomarcussis (Fig. 6) and strictosidine synthase, another alkaloid biosynthesis enzyme, has been shown to have a Kₘ ~ 4 mM for tryptamine and secologanin (25). To determine whether these results were likely the result of allosteric interactions or direct competition, several of these compounds were crystallized with NR.

**X-ray Crystal Structure of NR**—To understand the molecular basis of noroxomaritidine reduction by NR and examine the potential binding of the individual components tyramine and 3,4-dihydroxybenzaldehyde for norbelladine reduction, the
Identification of Noroxomaritidine Reductase

TABLE 2
Summary of crystallographic data collection and refinement statistics

<table>
<thead>
<tr>
<th>Crystal</th>
<th>NR/NADP⁺</th>
<th>NR/NADP⁺-tyramine</th>
<th>NR/NADP⁺-piperonal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Space group</td>
<td>P₄₁₂₂</td>
<td>P₂₁₋₁₋₁</td>
<td>P₃₁₂</td>
</tr>
<tr>
<td>Cell dimensions (Å)</td>
<td>a = b = 60.38; c = 156.1</td>
<td>a = 61.88; b = 86.96; c = 186.1</td>
<td>a = b = 73.39; c = 167.9</td>
</tr>
<tr>
<td>Data collection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wavelength (Å)</td>
<td>0.98</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>Resolution range (Å) (highest shell resolution)</td>
<td>45.2−1.73 (1.79−1.73)</td>
<td>39.6−1.81 (1.88−1.81)</td>
<td>35.9−1.50 (1.55−1.50)</td>
</tr>
<tr>
<td>Reflections (total/mosaic)</td>
<td>356,653/27,021</td>
<td>943,975/90,846</td>
<td>862,122/84,666</td>
</tr>
<tr>
<td>Complementary (highest shell)</td>
<td>99.7% (98.6%)</td>
<td>99.2% (91.6%)</td>
<td>99.9% (98.7%)</td>
</tr>
<tr>
<td>Rmerge (highest shell)</td>
<td>23.9 (2.4)</td>
<td>25.7 (2.3)</td>
<td>17.6 (2.3)</td>
</tr>
<tr>
<td>Rmerge (highest shell)</td>
<td>10.6% (52.1%)</td>
<td>9.8% (57.4%)</td>
<td>7.1% (66.8%)</td>
</tr>
<tr>
<td>Model and refinement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rmerge (%)</td>
<td>18.7% (21.4%)</td>
<td>18.2% (21.3%)</td>
<td>14.9% (16.0%)</td>
</tr>
<tr>
<td>No. of protein atoms</td>
<td>1870</td>
<td>7,600</td>
<td>3971</td>
</tr>
<tr>
<td>No. of water molecules</td>
<td>146</td>
<td>443</td>
<td>644</td>
</tr>
<tr>
<td>No. of ligand atoms</td>
<td>48</td>
<td>174</td>
<td>107</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>1.05</td>
<td>1.13</td>
<td>1.18</td>
</tr>
<tr>
<td>Bond lengths (Å)</td>
<td>0.006</td>
<td>0.008</td>
<td>0.007</td>
</tr>
<tr>
<td>Bond angles (°)</td>
<td>37.9, 37.7, 41.8</td>
<td>41.2, 41.2, 42.4</td>
<td>20.2, 18.2, 33.1</td>
</tr>
<tr>
<td>Average B-factor (Å²) protein, solvent, ligands</td>
<td>98.8, 12.6, 0.6%</td>
<td>97.2, 2.5, 0.6%</td>
<td>97.2, 2.5, 0.6%</td>
</tr>
</tbody>
</table>

FIGURE 7. Three-dimensional structure of NR. A, ribbon diagram of the NR tetramer complexed with NADP⁺ and tyramine is shown. Each monomer is colored differently. NADP⁺ and tyramine are shown as space-filling models. B, NADP⁺ binding in NR. Residues interacting with NADP⁺ are shown as stick figures. Dotted lines represent hydrogen bond and charge-charge interactions.

The x-ray crystal structure of the enzyme was determined. Because of the limited amounts of noroxomaritidine available and instability of 3,4-dihydroxybenzaldehyde, crystallization efforts used tyramine and piperonal as a substrate mimic. No enzymatic activity was observed when these individual molecules were incubated with NR and NADPH. The x-ray crystal structures of N. sp. aff. pseudomaritimus NR complexed with NADP⁺, NADP⁺ and tyramine, and NADP⁺ and piperonal were determined at 1.73-, 1.81-, and 1.50-Å resolution, respectively (Table 2). The NR-NADP⁺ complex was originally intended to be apoenzyme; however, the cofactor co-purified with the protein. The overall structure of NR reveals a canonical SDR-fold (Fig. 7A). NR forms a tetrameric structure with each 271-amino acid monomer adopting an alternating α/β structure formed by seven parallel β-sheets sandwiched between two layers of three α-helices that shape the NADPH-binding domain (Fig. 7A).

NR uses NADPH as an electron donor for noroxomaritidine reduction. The crystal structure of NR shows that NADP⁺ binds the active site of NR through a combination of polar and nonpolar interactions (Fig. 7B). The adenine ring, which is in the anti-conformation, interacts with the side chain carboxylate of Val83 through its exocyclic N6 (3.0 Å) and with an oxygen of Asp272 through hydrogen bonding with N1 (2.8 Å). The 2'-phosphate of the adenine ribose forms charge-charge interactions with the side chain of Arg90 (2.7–2.8 Å) and a hydrogen bond with the backbone amide of Arg265 (2.8 Å). Additionally, the adenine ribose phosphate interacts with the hydroxyl group of Ser56 (2.8 Å). These interactions are key structural features that distinguish the NADPH-specific SDRs from the NADH-specific members of the family (26, 27). The nicotinamide ribose adopts a C3'-endo conformation with the C3 hydroxyl group hydrogen bonding to the side chain hydroxyl group of Thr57 (2.8 Å). The 2'-hydroxyl of the nicotinamide ribose interacts with the catalytic tyrosine (Le Tyr175, 2.6 Å) and the 3'-hydroxyl of the nicotinamide ribose interacts with Lys109 (3.0 Å), the second catalytic residue, and the carbonyl of Asn110 (2.7 Å). A water molecule forms hydrogen bonds that bridge two of the nicotinamide ribose phosphates and interacts with the side chain amide nitrogen of Asn110 (2.9 Å). Another hydrogen bond forms between an oxygen on the one nicotinamide ribose phosphate and Thr216 (2.8 Å). The syn conformation of the nicotinamide ring is stabilized by a hydrogen bond involving the carbonyl group to the backbone amide-nitrogen of Thr216.
(2.8 Å). The A-face of the nicotinamide ring packs against Ile<sup>209</sup>, Pro<sup>221</sup>, Gly<sup>226</sup>, and Ala<sup>327</sup> to orient the B-face toward the substrate binding pocket.

The x-ray crystal structures of NR in complex with NADP<sup>+</sup> and either tyramine (Fig. 8A) or piperonal (Fig. 8C) identify how these ligands fit within the active site. In the NR structure with tyramine (Fig. 8A), the amine group of the ligand is positioned roughly equidistant between the hydroxyl group of Tyr<sup>213</sup> (3.2 Å) and the C4-position of NADP<sup>+</sup> (3.2 Å). The phenolic portion of the ligand bonds in a pocket defined by Tyr<sup>218</sup> and Phe<sup>116</sup> on opposite sides with the tyramine hydroxyl group forming a hydrogen bond with the side chain of Glu<sup>256</sup> (3.1 Å). Notably, the ligand binding site of NR is much larger than tyramine (Fig. 8B) with additional space extending toward Ile<sup>277</sup> and Arg<sup>283</sup>. The NR-NADP<sup>+</sup>-piperonal complex shows a similar mode of binding for the ligand (Fig. 8C). The aldehyde group is bound between Tyr<sup>213</sup> and the C4 of NADP<sup>+</sup>. The benzodioxole ring is sandwiched between Tyr<sup>116</sup> and Phe<sup>116</sup> with a hydrogen bond interaction made between Ser<sup>143</sup> and the O3 of piperonal.

Experiments on tyramine, piperonal, and 3,4-dihydroxymandeldehyde leave the substrate of the norbelladine reduction by this NR undefined. It could be the individual components tyramine and the corresponding aldehyde or preformed norcraugosidine. The uncondensed components bind the NR active site, which could indicate a contribution of the enzyme to condensation through an increased local concentration of the two components but the location of binding is the same. This is not an ideal setup for a condensation on the enzyme because they would be in the same place. Also, none of the NR structures have the enzyme and NADP<sup>+</sup> in an orientation that would obviously facilitate the condensation reaction of these substrates. For these reasons, the individual components bind the enzyme active site and can inhibit the noroxomaritidine reduction. It remains unclear if the enzyme actually uses these substrates to make norcraugosidine or takes norcraugosidine from solution.

Comparison of the NR structures reveals conformational flexibility in the NR active site (Fig. 8D). In the "open" active site conformation, the Arg<sup>209</sup>-Lys<sup>220</sup> loop is extended out into solution away from the ligand binding site. In contrast, this region hinges at Arg<sup>209</sup> and Lys<sup>220</sup> to adopt a "closed" active site conformation in the NR-NADP<sup>+</sup>-tyramine complex. This shift in structure moves Thr<sup>213</sup> and Phe<sup>216</sup> 10.2 and 12.0 Å, respectively, into the ligand binding site. The Arg<sup>209</sup>-Lys<sup>220</sup> loop also adopts the closed active site conformation in the NR-NADP<sup>+</sup>-piperonal complex.

Because of limited availability of noroxomaritidine, computer-aided docking of this ligand into the x-ray structure of NR was performed (Fig. 9). The protein model contained the crystallographically determined NADP<sup>+</sup> and the Arg<sup>209</sup>-Lys<sup>220</sup> loop was in the open active site conformation. The closed active site conformation was too constrained to allow for fitting of the ligand in the site. Movement of the Arg<sup>209</sup>-Lys<sup>220</sup> loop provides sufficient space in the NR active site to accommodate noroxomaritidine. The ligand binds with its ketone group and carbon-carbon bond of the A-ring oriented toward the Tyr<sup>217</sup>-Lys<sup>219</sup> catalytic dyad and the C4 position of the cofactor nicotinamide ring to allow for reduction of the substrate.

**Discussion**

The Amaryllidaceae alkaloids are a large group of plant natural products with over 300 documented structures and diverse biological activities. Moreover, these molecules are of pharmaceutical interest either as potential anti-tumor agents or for the symptomatic treatment of Alzheimer disease (1, 2, 4, 28). Although the general chemical steps in the core Amaryllidaceae alkaloid biosynthetic pathway have been established (Fig. 1),

---

AUGUST 5, 2016 • VOLUME 291 • NUMBER 32

ASBMB

JOURNAL OF BIOLOGICAL CHEMISTRY

16747

257
the genes and enzymes of the pathway remain largely unstudied. Recent transcriptome analysis for co-expression of compounds and biosynthetic genes involved in Amaryllidaceae alkaloid biosynthesis have provided molecular insights on this pathway, including the identification and characterization of NoOMT, which catalyzes the methylation of norbelladine, and CYP96T1, which can form noroxomaritidine (17, 18). A similar strategy now identifies a member of the SDR enzyme family from N. sp. *aff. pseudonarcissus* as a reductase (*i.e.* NR) that preferentially uses noroxomaritidine as a substrate. 400-Fold lower norcraugosine reductase activity, which was above background, was also detected.

NR belongs to the SDR superfamily (Fig. 2) (14, 29). Although originally identified as oxidoreductases of ketones and aldehydes, including multiple enzymes involved in the biosynthesis of different types of alkaloids (22, 30, 31). The biochemical repertoire of the SDR superfamily has expanded to include carbon-carbon double bond reductases (23) and imine reductases (16). The co-expression of NR with *NAOMT* suggested a possible role in the biosynthesis of Amaryllidaceae alkaloids.

Given that SDRs, such as *Sp* progesterone reductase (23), can catalyze the reduction of carbon-carbon double bonds, NR was tested using noroxomaritidine as a substrate (Fig. 3 and Table 1). Biochemical analysis of NR indicates that the enzyme functions to reduce noroxomaritidine using NADPH. Although the enantiomeric forms of noroxomaritidine have not been isolated, assays suggest a preference of NR for one of these forms. Depending on the enantiomer being reduced this activity could contribute to the biosynthesis of either martianamine or elwesine (32, 33). Both of these compounds share a reduced carbon-carbon double bond, but are derived from the (10β,4αS)-noroxomaritidine and (10β,4αR)-noroxomaritidine skeletons, respectively. *N*. sp. *aff. pseudonarcissus* has not been previously documented to have these compounds and proper standards to determine their presence or absence are lacking. So, the product profiles in *N*. sp. *aff. pseudonarcissus* are unknown.

The x-ray crystal structures of NR in complex with NADP*+* and either tyramine or piperonal (Figs. 6 and 7) clearly define the common structural features, including the catalytic dyad (*i.e.* Tyr*175* and Lys*179*) and nucleotide cofactor binding site of the SDR superfamily (14). Moreover, the various structures of NR reveal the conformational flexibility of the Arg*200*-Lys*209* loop in NR. Although the molecular determinants of noroxomaritidine binding in NR remain to be examined in detail, computer-aided docking of the substrate into the open conformation active site (Fig. 9) suggests a possible chemical mechanism for noroxomaritidine reduction.

Extensive studies on multiple SDR enzymes support a common reaction mechanism in which the active site tyrosine serves as a general acid to protonate the keto-group of a bound substrate and the lysine in the catalytic dyad lower the pKₐ of the catalytic tyrosine to promote proton donation (14). The arrangement of Tyr*175* and Lys*179* in the NR active site (Fig. 8, A and B) suggests a similar chemical mechanism for noroxomaritidine reduction (Fig. 10A). In the reaction catalyzed by NR, binding of noroxomaritidine positions the substrate ketone in proximity to Tyr*175* and NADPH. Electrostatic interaction with Lys*179* reduces the pKₐ of Tyr*175* to polarize the substrate carbonyl group for protonation and hydride transfer. The resulting enolate-like structure can then tautomerize to yield the carbonyl product (*i.e.* oxomaritamine) with a reduced carbon-carbon double bond. The overall reaction is comparable with those proposed for various steroid carbon-carbon double bond reductases (23, 34, 35). The reduction of noroxomaritidine catalyzed by NR also requires positioning of the carbon-carbon double bond in proximity to the C₄ of NADPH to allow for hydride transfer. It is possible that the flexibility of the NR active site loop can aid in the proper orientation of the substrate for catalysis. A similar role for the active site loop as an adaptive feature for allowing the binding of different substrates has been proposed for the SDRs that catalyze reductive terpene cyclization in iridoid alkaloid biosynthesis (36).

Reduction of an imine is required in the biosynthesis of norbelladine, a central metabolite in the Amaryllidaceae alkaloid pathway (Fig. 1). Because NR was the oxidoreductase that was co-expressed most consistently with *NAOMT*, we postulated that NR may also reduce norcraugosine to norbelladine. Biochemical assays show that incubation of NR with tyramine, 3,4-dihydroxybenzaldehyde, and NADPH leads to norbelladine synthesis at a rate above background, but also with a specific activity substantially lower than that for noroxomaritidine reduction (Fig. 5 and Table 1). Comparable chemistry was also observed when other aldehydes (*i.e.* vanillin, isovanillin, and piperonal) were used in place of 3,4-dihydroxybenzaldehyde. This substrate flexibility and the low specific activity for this reaction may indicate this is a nonspecific enzymatic reduction.

For NR to catalyze norbelladine synthesis, nucleophilic attack of the tyramine amine group on the aldehyde via a condensation reaction leads to formation of norcraugosine (Fig. 10B). Whether NR serves to enhance the rate of this reaction relative to the non-enzymatic rate is unclear. The three-dimen-
sional structures of NR and the competition assays demonstrate that tyramine and aldehyde substrates bind at the active site; however, it is not obvious how both molecules would bind for the reaction leading to norcraugusone. Nonetheless, our data indicates that the reduction of the imine increases in the presence of NR and NADPH. For this reaction, binding of norcraugusone in the NR active site may position the imine group in proximity to Tyr179 and NADPH. As described above polarization of Tyr179 via interaction with Lys179 would allow the hydroxyl group to serve as a general acid with hydride transfer from NADPH, which results in norbelladine synthesis (Fig. 16B).

The specific activity differences for noroxomaritidine reduction (8600 ± 1250 pmol min⁻¹ mg⁻¹ of protein⁻¹) compared with norcraugusone reduction (21.2 ± 1.2 pmol min⁻¹ mg⁻¹ of protein⁻¹) (Table 1) suggest that former reaction is preferred and that the latter is a nonspecific reaction. It remains to be determined if the norbelladine formation in planta results from the low side activity of NR or an as yet unidentified enzyme. It should be noted that derivatives of the carbon-carbon double bond reduction of noroxomaritidine are rarely reported in planta. It is possible that NR is normally outcompeted by an unidentified reductase that reduces the ketone group of noroxomaritidine, which is a necessary step for the production of the hemithamine and crinine alkaloids (20). Ultimately, the identification of NR, along with previous work on N4OMT and CYP96T1 (17, 18), provide biochemical insight on Amaryllidaceae alkaloid biosynthesis and will be useful in future co-expression analysis to search for other genes related to this pathway and to determine the connections between Amaryllidaceae alkaloid biosynthetic gene expression and the accumulation of various alkaloids in different Amaryllidaceae plants.

**Experimental Procedures**

**Plant Tissue and Chemicals**—Plant tissues and chemicals used in this study were previously described (17, 18). In addition, piperonil was purchased from Sigma; 8,9-O-dimethyleoxomaritidine and noroxomaritidine were obtained from natural product collection; dimethyl sulfoxide was from New England BioLabs; NADP⁺ was from MP Biomedicals; anhydrous citric acid (ACS reagent grade 99.5%) and sodium borohydride (99%) were from Acros Organics; and sodium cyanoborohydride (reagent grade 95%) was from Aldrich.

**Candidate Gene Identification**—Candidate genes were selected as in the discovery of CYP96T1 (18), but instead of a list of cytochromes P450, a list of oxidoreductases (supplemental Table S1) were used in the BLASTP query with an e-value cutoff of 1 × 10⁻⁴ against ESTScan predicted peptides for each transcriptome. A contig for an SDR was found to co-express with N4OMT in 4 of the 5 available Amaryllidaceae transcriptomes, whereas other BLASTP results showed co-expression with N4OMT in 3 or less of the transcriptomes (18).

**Cloning and Recombinant Enzyme Purification**—N. sp. aff. pseudornarcissus bulb cDNA was prepared as previously described (18). The outer PCR contained the following components: 25 ng N. sp. aff. pseudornarcissus bulb cDNA, 1X Phusion HF reaction buffer, 1 unit of New England Biolabs Phusion High-Fidelity DNA polymerase, 3% dimethyl sulfoxide, 0.4 μM dNTPs, 0.4 μM NR forward outer (5'-dGGAAAGGCCTCADGAGGAGATT-3'), and 0.4 μM NR reverse outer (5'-dAGATAGCACCCTGAGGAT-3') primers. The parameters for the PCR were 98 °C 10 s for 1 cycle; 98 °C 10 s, 50 °C 30 s, 72 °C 60 s for 35 cycles; 72 °C 5 min 1 cycle; 4 °C until removed. The inner PCR mixture was the same except 0.2 μM NR forward
Identification of Noroxomaritidine Reductase

<table>
<thead>
<tr>
<th>Substrate</th>
<th>Product</th>
<th>Parameters (parent ion m/z)(CE)(DP)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noroxomaritidine</td>
<td>Oxomaritamine</td>
<td>(274.13)(350)(70)</td>
</tr>
<tr>
<td>6-OH-Dihydroxybenzoic acid</td>
<td>Norbelladine</td>
<td>(260.13)(150)(70)</td>
</tr>
<tr>
<td>3,4-Dihydroxybenzaldehyde</td>
<td>Vanillin</td>
<td>(254.28)(160)</td>
</tr>
<tr>
<td>and tyramine</td>
<td>3'-O-Methylorobelladine</td>
<td>(264.28)(150)</td>
</tr>
<tr>
<td>L-Valine and tyramine</td>
<td>4'-O-Methylorobelladine</td>
<td>(274.13)(200)</td>
</tr>
<tr>
<td>Piperonal and tyramine</td>
<td>4'-O-(1,3-Benzodioxol-5-yilmethyllaminooethyl)phenol</td>
<td>(274.21)(200)</td>
</tr>
</tbody>
</table>

The pET28a-NR construct was transformed into E. coli Rosetta II (DE3) cells (EMD Millipore). Cells were cultured in Terrific broth up to $A_{600\text{nm}} = 0.6 - 0.8$. Induction of protein expression was performed with a final concentration of 1 mM isopropyl 1-thio-β-D-galactospyranoside in 16°C. The enzyme used in enzyme assays was purified with a TALON-cobalt column as previously (17). For protein crystallization, crystals were pelleted by centrifugation and resuspended in lysis buffer (50 mM Tris, pH 8.0, 500 mM NaCl, 20 mM imidazole, 10% glycerol, and 1% Tween 20). Following sonication, cell debris were removed by centrifugation, and the resulting lysate was passed over either a Ni$^{2+}$-nitrilotricarboxylic acid (Qiagen) or a TALON-cobalt column equilibrated in lysis buffer. The column was then washed with 50 mM Tris, pH 8.0, 500 mM NaCl, 20 mM imidazole, and 10% glycerol. Bound His-tagged protein was eluted with 50 mM Tris, pH 8.0, 500 mM NaCl, 250 mM imidazole and 10% glycerol. For protein crystallization, the His-tag was removed by overnight dialysis at 4°C using thrombin (1:2000 total protein). Dialyzed protein was loaded onto an mixed benzamidine-Sepharose/ Ni$^{2+}$-nitrilotricarboxylic acid column. The flow-through was loaded onto a Superdex-75 26/10 Hi Load size exclusion column (GE Healthcare) equilibrated with 25 mM HEPES, pH 7.5, and 100 mM NaCl using an Akta Explorer FPLC system. Concentration of all protein purifications was determined by the Bradford method (Protein Assay, Bio-Rad) with bovine serum albumin as a standard.

Enzyme Assays—Assays were performed with 100 mM citrate buffer, pH 6.0, at 35°C in 40-μl reactions except where noted in figure legends. Compounds screened for enzymatic activity were monitored with the parameters listed in Table 3. Assay-specific conditions are described in appropriate figures and tables. Assays were extracted with ethyl acetate at pH 9.5 as previously described (17). After drying, extracts were re-suspended in mobile phase matching the solvent composition at the beginning of the HPLC program. Extracts, except where noted otherwise, were run on a Li-20AC XR prominence liquid autosampler coupled to a 20AD XR prominence liquid chromatography instrument with a Phenomenex Luna 5-μm C8(2) 250 x 4.60-mm column and a QTRAP 4000 as used for CYP96T1 (18). The LC time programs used for assays were as follows: a isocratic time program with 20% acetonitrile and 0.1% formic acid was used for pH optimum, temperature optimum, IC$_{50}$ and specific activity assays but initial screening for substrates for enzymatic activity used the same time program as for substrate screening of N4OMT (17). The MRM parameters used, with the exception of specific activity measurements, to monitor norbelladine were m/z 260.0/138.0 and 260.0/121.0 with CE (15) and DP (50) and oxomaritamine was monitored with the following MRM m/z 274.3/136.1 and 274.3/219.1 with CE (35) and DP (70).

Specific activity measurements required quantification of oxomaritamine and a new LC-MS/MS setup. Oxomaritamine was quantified by incubating noroxomaritidine and NR overnight and equating the quantity of noroxomaritidine consumed to the quantity of oxomaritamine produced using the same HPLC program used to analyze enzyme assays during N4OMT characterization at A$^{280\text{nm}}$ (17). This oxomaritamine standard was used in specific activity experiments on the LC-MS/MS to determine product quantity. For specific activity measurements, a QTRAP 6500 was used for MRM analysis with an isotropic flow of 0.1% acetonitrile and 0.1% formic acid in H$_2$O. The MRM parameters for specific activity measurements were norbelladine m/z 260.1/138.0 and 260.1/121.0 CE (10) and DP (60), 4'-O-methylorobelladine m/z 274.1/137.0 and 274.1/122.0 CE (20) and DP (60), and oxomaritamine m/z 274.1/136.1 and 274.1/219.1 CE (30) and DP (70).

Norbelladine Analog Synthesis—4-(1,3-Benzodioxol-5-yilmethylamino)ethylphenol, 3'-O-methylorobelladine, and 4'-O-methylorobelladine were prepared by mixing ~100 mM sodium cyanoborohydride and ~10 mM tyramine with ~10 mM piperonal, vanillin, or isovanillin, respectively, in 2.5 mL of anhydrous methanol. After incubation overnight at room temperature, the reactions were evaporated to 0.5 mL with subsequent addition of 2 mL of sodium carbonate, pH 9.5, and extracted twice with 2 mL of ethyl acetate. After drying, all extracts were dissolved in 1 mL of water, diluted 1:20, and purified by fraction collection with the same HPLC program used to analyze enzyme assays during N4OMT characterization (17).

Protein Crystallization—Purified NR was concentrated to 8 mg mL$^{-1}$ and crystallized using the hanging-drop vapor-diffusion method with a 2-μl drop (1:1 concentrated protein and crystallization condition). Diffraction quality crystals of the NRNADP$^+$ complex were obtained at 4°C with 20% PEG-8000 and 100 mM HEPES buffer, pH 7.5. Crystals of the NRNADP$^+$-tyramine complex formed at 4°C in 2 mM ammonium sulfate, 100 mM CAPS buffer, pH 10.5, 200 mM lithium sulfate, 1.5 mM NADP$^+$, and 6 mM tyramine. For the NRNADP$^+$-piperonal complex, 35% 2-methyl-2,4-pentanediol, 100 mM sodium acetate, pH 4.5, and 3 mM piperonal was used for crystallization. Individual crystals were flash-frozen in liquid nitrogen with the mother liquor containing 25%...
Identification of Noroxomaritidine Reductase

glycerol as a cryoprotectant. Diffraction data (100 K) was collected at the Argonne National Laboratory Advanced Photon Source 19-ID beamline. The data were indexed, scaled, and integrated with HKL3000 (37). Molecular replacement implemented in Phaser (38) used D. stramonium tropine reductase-II (PDB code 2AE2) (39) as a search model to determine the structure of the NR-NADP⁺-tyramine complex. Iterative rounds of manual model building and refinement, which included translation-libration-screen models, used COOT (40) and PHENIX (41). The crystal structures of NR-NADP⁺ and NR-NADPH-piperonal were determined by molecular replacement using Phaser and the NR-tyramine structure with ligands removed as the search model and with building and refinement as described above. Data collection and refinement data are summarized in Table 2. The final model of the NR-NADP⁺ complex included residues Ser26 to Gly271, NADP⁺, and 136 waters. The final model of the NR-NADPH-tyramine complex included residues Ser26 to Gly251 for chain A, residues Met19 to Gly271 for chain B, residues Leu14 to Asn270 for chain C, and residues Leu16 to Gly251 for chain D. NADP⁺ in chains A-C, tyramine in chains B and D, and 443 waters. The final model of the NR-NADPH-piperonal complex included Met19 to Gly237 and NADP⁺ in chains A and B, piperonal in chain A, and 644 waters. Coordinates and structure factors for NR complexed with NADPH (PDB 5EU1), NADPH and tyrosine (PDB 5FF9), and NADPH and piperonal (PDB 5FF7) have been deposited in the RCSB Protein Data Bank.

Computational Docking of Substrates—Molecular docking of noroxomaritidine into the three-dimensional structure of the NR-NADP⁺ complex was performed using AutoDock Vina (version 1.1.2) (42). The ligand was generated using ChemDraw 3D and energy minimized. The ligand was manually placed into the active site by using the position of NADP⁺ in the active site as a guide with docking using a grid box of 30 × 30 × 30 Å and the level of exhaustiveness set to 8.

Author Contributions—As a part of his Ph.D. thesis M. B. K. conceived the study, performed candidate selection, cloning, and enzymatic experiments, and wrote the majority of the manuscript. C. K. H. performed crystallography and docking experiments and contributed to the writing of the manuscript. J. M. J. contributed to conception of crystallography and docking experiments, revisions of the manuscript, and data analysis. T. M. K. contributed to conception of experiments and revisions of the manuscript.

Acknowledgements—Portions of this research were carried out at the Argonne National Laboratory Structural Biology Center of the Argonne National Laboratory, a national user facility operated by the University of Chicago for the Department of Energy Office of Biological and Environmental Research under Grant DE-AC02-06CH11357.

References

August 5, 2016 • Volume 291 • Number 32

JOURNAL OF BIOLOGICAL CHEMISTRY

16751
Identification of Noroxomartidine Reductase


Appendix II

Dissonance Strikes a Chord in Stilbene Synthesizers

Cynthia K. Holland\textsuperscript{1}, Barbara Cascella\textsuperscript{1}, Joseph M. Jez\textsuperscript{1}

\textsuperscript{1}Department of Biology, Washington University in St. Louis, St. Louis, MO 63130
Dissonance Strikes a Chord in Stilbene Synthesizers

Cynthia K. Holland,1 Barbara Cascella,1 and Joseph M. Jez1*2

1Department of Biology, Washington University in St. Louis, One Brookings Drive, CB1137, St. Louis, MO 63130, USA
Correspondence: jez2@wusl.edu
http://dx.doi.org/10.1016/j.chembiol.2016.11.016

In this issue of Cell Chemical Biology, Mori et al. (2016) combine X-ray crystallography and biochemistry to discover a new mechanism for stilbene synthesis in bacteria. The dialkyl-condensing enzyme StID catalyzes formation of cyclohexanediones using a non-canonical β-ketosynthase active site. Aromatization by StIC completes production of the stilbene product.

In the delicate balance between mutualism and parasitism, lea makes a soil microbe and an aromatic natural product, the stilbene, that shrinks pocket that was traditionally thought of as plant defense compounds involved in deterring insects, nematodes, and herbivores (Chong et al., 2009); however, chemical ecologists recognized that Photobacterium spp. bacteria synthesize stilbenes as part of their mutualistic relationships with host nematodes that parasitize insects, particularly moths (Richardson et al., 1988). Once Heterorhabditis nematode parasites reach the hemolymph of a suitable insect host, the nematode regurgitates the Photobacterium mutualism from its gut to kill the insect (Eleftherianou et al., 2007). In these insect cadavers, stilbene levels soar to 3 mg/g of insect tissue, suggesting that the microbe synthesizes this aromatic compound both to suppress the insect immune response and to embroil other microbes from joining the feast (Klu and Webster, 2003).

To date, certain plants and Photobacterium bacteria are the only known stilbene synthesizers. In plants, type III polyketide synthases (PKSs) produce stilbenes (Tropp et al., 1995). Plant stilbene synthesis begins with the loading of p-coumaroyl-Coenzyme A (CoA) onto an active site cysteine, followed by three sequential condensation reactions with activated acetyl units derived from decarboxylation of malonyl-CoA; cyclization of the tetraakide intermediate generates the final stilbene product (Figure 1, left) (Austin et al., 2004). In contrast, early work proposed that Photobacterium used cinnamoyl-acetyl carrier protein (ACP) and isovaleryl-ACP as substrates for a diacyl condensation reaction catalyzed by StID, followed by aromatization by StIC to produce a stilbene product (Figure 1, right) (Joyce et al., 2008). While both plants and Photobacterium use p-coumarate as the starting molecule for stilbene synthesis, the different substrates of the bacterial and plant condensing enzymes hinted that these organisms synthesize stilbenes using different chemistries.

In this issue of Cell Chemical Biology, Takahiro Mori and colleagues (Mori et al., 2016) investigate StID and StIC, the last two enzymes in the stilbene biosynthetic pathway of P. luminescens, using a combination of functional assays and X-ray crystallography to confirm for the first time that these two enzymes catalyze condensation and aromatization reactions, albeit in notably different ways than their plant homologs. The authors found that StID does not condense cinna- moyl-CoA with malonyl-CoA, as the plant stilbene synthases do, but rather prefers longer chain variants of each. The reaction product of StID contains a carboxylated cyclohexanecione ring, and the presence of the carboxylic acid on the ring is essential for recognition by StIC. Interestingly, these enzymes are capable of producing a variety of unnatural stil- bene products, as they exhibit tolerance of a range of chain lengths in the β-ke- tous synthase superfamilies, as expected. Surprisingly, the authors found that mutation of His302 in StID to an alanine resulted in enhanced enzymatic activity by expanding the active-site cavity, which suggested that StID uses alternative reactive residues for the condensation reaction. This is the first enzyme identified in the β-ketosyn- thase superfamilies that contains a non-reactive histidine within the conserved active-site triad.

StID differs from the plant stilbene synthases because it catalyzes a
non-decarboxylative Claisen condensation reaction of the β-ketoacyl and an unsaturated acyl substrate. Instead, the cyclic product of SSID is decarboxylated and aromatized by SIC to afford the final stilbene product. Analysis of the SSID active site revealed a lone candidate residue that could compensate for the displaced His302—a glutamate located near the conserved triad. Mutation of Glu154 to glutamine abolished the dialkyl condensation activity of SSID. This catalytic base appears to work in concert with a neighboring water molecule to condense the two substrates and promote cleavage of the cyclic product from the active site cysteine. Thus, SSID’s mode of action is akin to that of ketosynthase enzymes that perform head-to-head condensations of alkyl thioester substrates during production of 2,5-dialkylferulic acids in Chitinophaga pinensis and other bacteria (Fuchs et al., 2013).

Figure 1. Evolution of Stilbene Biosynthesis: Common Scaffold with Different Reaction Mechanisms in Plants and Bacteria

The SSID dialkyl condensing enzyme from Photohablobas fumariformis adopts the canonical thiolester bond in β-ketosynthases (ribbon diagram). The plant stilbene synthases use a conserved active site to generate stilbene from p-coumaroyl-CoA and three malonyl-CoA (left). The SSID enzyme uses a different active site architecture to condense two dialkyl thioester substrates into a cyclohexeneone scaffold, which SIC aromatizes into the final stilbene product (right).

Overall, SSID catalyzes the condensation of isovaleryl-β-ketoacyl and cinna- moyl (β-κetoacyl) thioesters to generate a carboxy-cyclohexeneone moiety, which the aromatase SIC converts to a stilbene product. The SSID reaction sequence begins with loading of a β-κetoacyl starter to Cys126, which then rotates toward the short acyl-binding pocket and allows for entry of the second acyl substrate to the active site. From there, Glu154 abstracts a proton from the β-κetoacyl starter to spark the Claisen condensation, resulting in C-C bond formation. An intramolecular Michael addition yields a cyclic product that is cleaved from Cys126 by a Glu154-activated nucleophilic water molecule. Release of the carboxy-cyclohexeneone product from the active site follows reorganization of the base-catalyst Glu154 by a ketoenol tautomerization. SIC uses this product to generate the decarboxylated, aromatic stilbene.

These findings shed light on the previously uncharacterized mechanism of bacterial stilbene production. Nature converged two variations on stilbene biosynthesis with bacterial synthetase, causing dissonance by breaking from the

REFERENCES

Appendix III

Arabidopsis: The original plant chassis organism

Cynthia K. Holland¹, Josesph M. Jez¹

¹Department of Biology, Washington University in St. Louis, St. Louis, MO 63130
Arabidopsis: the original plant chassis organism

Cynthia K. Holland¹ · Joseph M. Jez¹

Received: 14 February 2018 / Accepted: 11 April 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Arabidopsis thaliana (thale cress) has a past, current, and future role in the era of synthetic biology. Arabidopsis is one of the most well-studied plants with a wealth of genomics, genetics, and biochemical resources available for the metabolic engineer and synthetic biologist. Here we discuss the tools and resources that enable the identification of target genes and pathways in Arabidopsis and heterologous expression in this model plant. While there are numerous examples of engineering Arabidopsis for decreased lignin, increased seed oil, increased vitamins, and environmental remediation, this plant has provided biochemical tools for introducing Arabidopsis genes, pathways, and/or regulatory elements into other plants and microorganisms. Arabidopsis is not a vegetative or oilseed crop, but it is as an excellent model chassis for proof-of-concept metabolic engineering and synthetic biology experiments in plants.

Keywords Arabidopsis thaliana · Chassis organism · Genome resources · Metabolic engineering · Plant biochemistry · Synthetic biology

Introduction
Though small in size, Arabidopsis thaliana (thale cress), the model plant organism, has influenced the field of plant biology in a mighty way. Some may argue that it is a useless weed with no agronomic significance (which is true); however, Arabidopsis has many benefits that make it a suitable organism for genetic and biochemical studies and as a synthetic biology chassis. Friedrich Laibach first proposed using Arabidopsis as a genetic model in the 1940s (Laibach 1943; Meyerowitz 2001). The same features that made this plant attractive as a model organism for genetics—its short 6–8-week lifespan, easy maintenance, sequenced genome, small space requirements, and genome of five chromosomes—also make it useful for biochemists, metabolic engineers, and synthetic biologists (Laibach 1907; Leonelli 2007).

With the birth of molecular biology and genetic engineering, the tools to modulate expression of genes or to introduce genes from other organisms provided the first opportunities to explore new applications. While synthetic biology focuses on designing novel synthetic biological pathways and/or organisms or redesigning existing natural systems, metabolic engineering concentrates on optimizing genetic and regulatory processes to increase production of a particular molecule. Metabolic engineering and synthetic biology aim to either induce or alter transcriptional, signaling, and/or metabolic pathways for the production of desired traits and products (Purnick and Weiss 2009). While Arabidopsis may not be practical or economically feasible as a large-scale production platform, the benefits of using a well-understood and resource-rich model plant as a starting point for proof-of-concept metabolic engineering and synthetic biology experiments offers certain advantages. Indeed, many of the findings from the Arabidopsis literature have influenced work in other agronomic, nutritional, and economically important plants (Provart et al. 2015); however, key disadvantages to large-scale Arabidopsis engineering are its small biomass, difficulties with converting plant tissues into suspension cell cultures, and difficulties in engineering tissues outside of the leaf.

Here we review the current state of plant metabolic engineering and synthetic biology involving A. thaliana and to showcase Arabidopsis as a chassis for developing plants as green factories for energy, the environment, and human health.
The Arabidopsis toolkit

Many resources exist for the synthetic biologist considering Arabidopsis as a chassis organism (Table 1). For the novice plant biologist (and perhaps expert engineer), this plant is an excellent system to learn. Arabidopsis has excelled as a model organism due to the ability to do both forward and reverse genetics easily and quickly. There are Agrobacterium tumefaciens transfer-DNA (T-DNA)-induced insertion mutants in almost every gene (O’Malley et al. 2015), and Arabidopsis is a genetically tractable organism that can be transformed using disarmed A. tumefaciens transformed with a vector containing a gene of interest by a straightforward floral-dip method (Clough and Bent 1998). The Arabidopsis genome was the first plant genome to be sequenced (The Arabidopsis Genome Initiative 2000), and since that time, many more genomics and genetics tools have become available, including the use of genome editing approaches for specific modifications (Wood et al. 2011; Li et al. 2013; Mao et al. 2013).

<table>
<thead>
<tr>
<th>Resource</th>
<th>Uses</th>
<th>Website</th>
</tr>
</thead>
<tbody>
<tr>
<td>TAIR (The Arabidopsis Information Resource)</td>
<td>Genome information&lt;br&gt;DNA and seed stocks&lt;br&gt;Publications</td>
<td>Arabidopsis.org</td>
</tr>
<tr>
<td>AraPort (Arabidopsis Information Portal)</td>
<td>Gene structure and expression&lt;br&gt;Protein function&lt;br&gt;Interaction networks</td>
<td>AraPort.org</td>
</tr>
<tr>
<td>Virtual Plant</td>
<td>Visualization, integration, and analysis of genomic data&lt;br&gt;Exploring gene networks&lt;br&gt;Identifying molecular interactions</td>
<td>Virtualplant.bio.nyu.edu</td>
</tr>
<tr>
<td>Bio-Analytic Resource (BAR) for Plant Biology</td>
<td>Gene expression and protein tools&lt;br&gt;Molecular markers and mapping tools&lt;br&gt;Functional genomics&lt;br&gt;Interactive data visualization</td>
<td>Bart.toronto.ca</td>
</tr>
<tr>
<td>ePlant</td>
<td>Visualization of connections in genomics and molecular research&lt;br&gt;Expression Angler to find genes based on expression patterns&lt;br&gt;Find genes based on mutant phenotype&lt;br&gt;Visualizes natural variation, environmental effects, phenotypes, and responses, subcellular localization, protein networks, DNA sequences, protein sequences, and 3D structures</td>
<td>Bart.toronto.ca/eplant</td>
</tr>
<tr>
<td>eFP Browser</td>
<td>Electron fluorescent photographs of a gene of interest’s expression patterns&lt;br&gt;Subcellular localization&lt;br&gt;Seed coat localization&lt;br&gt;Spatio-temporal root stress expression data&lt;br&gt;mRNA levels on polyribosomes in seedlings under hypoxic stress</td>
<td>Bart.toronto.ca/eFP/cgi-bin/eFPWeb.cgi</td>
</tr>
<tr>
<td>ArGen Express</td>
<td>Microarray gene expression data for light, pathogen, development, abiotic stress, and ecotypes</td>
<td>jsp.weigelworld.org/ArGenExpress/resources</td>
</tr>
<tr>
<td>Genevestigator</td>
<td>Microarray and RNAseq gene expression data across tissues, diseases, nutrients, chemicals, genotypes</td>
<td>Genevestigator.org</td>
</tr>
<tr>
<td>1001 Genomes</td>
<td>Genome sequences for over 1000 Arabidopsis strain genomes, including visualization and analysis tools&lt;br&gt;Seeds are available for several hundred strains</td>
<td>1001genomes.org</td>
</tr>
<tr>
<td>1001 Methylome</td>
<td>Epigenome and transcriptome data for over 1000 Arabidopsis strains</td>
<td><a href="http://signal.salk.edu/1001.php">http://signal.salk.edu/1001.php</a></td>
</tr>
</tbody>
</table>
As Arabidopsis research progressed, the need for organized collections of seeds and other resources became necessary. In the United States, the Arabidopsis Biological Resource Center (ABRC) at The Ohio State University “collects, preserves, reproduces, and distributes diverse seed and other stocks of Arabidopsis thaliana and related species” (https://abrc.osu.edu). ABRC provides seed stocks and is a repository of clones, cell lines, protein chips, resources, cloning vectors, host strains, and education kits. In the United Kingdom, the Nottingham Arabidopsis Stock Centre (http://arabidopsis.info/) provides seed and molecular services. An abundance of inbred lines allows for genome-wide association (GWA) studies to identify the genomic basis of a phenotypic trait or characteristic (Atwell et al. 2010). Recently, 1135 natural Arabidopsis accessions were sequenced to enable analysis of genetic and epigenetic variation in Arabidopsis (http://1001genomes.org) (1001 Genomes Consortium 2016). Similarly, the DNA methylomes from ~1000 Arabidopsis inbred lines unveiled the epigenetic diversity and its relationship to phenotypic diversity (Ecker 2016; Kawakatsu et al. 2016).

Online databases ensure that Arabidopsis data and materials are easily accessible. TAIR (The Arabidopsis Information Resource; http://www.arabidopsis.org) is a genetic and molecular biology data resource with genome information, DNA and seed stocks, publications, and links to other Arabidopsis resources. The Arabidopsis Information Portal (http://www.araport.org) contains open-access resources for gene structure and expression, protein function, and interaction networks and offers a community space to share information.

Easy visualization and summary tools help users access the constantly growing information on Arabidopsis. Virtual Plant supports visualization, integration, and analysis tools for systems biology research for identifying molecular interactions within a set of genes and exploring gene networks (http://virtualplant.bio.nyu.edu; Katari et al. 2010). The Bio-Analytic Resource (BAR) for Plant Biology (http://bar.utoronto.ca) provides gene expression and protein tools, molecular markers and mapping tools, and other genomic tools and widgets for functional genomics and interactive data visualization. As part of BAR, ePlant helps researchers visualize connections in genomics and molecular research, including molecular structures and protein-protein interactions, to aid in hypothesis generation. The Arabidopsis electronic fluorescent pictograph (eFP) browser allows scientists to visualize the expression patterns of a gene of interest using fluorescent pictographs. Similarly, Expression Angler 2016 lets users search for genes with similar expression patterns and shows an eFP image of the expression data for each gene and includes features to predict cis-elements in the promoters of co-expressed genes.

Advances in bioinformatics, modeling, multigene engineering, sequencing technologies, and genome editing allow scientists to more accurately predict the results of metabolic modifications (Farre et al. 2015). Starting with “-omes” (genomics, transcriptomics, metabolomics, proteomics, etc.), genes, regulatory networks, and feedback regulation can be deciphered for knowledge-based metabolic engineering. Expression data sets from published experiments have been collected and compiled electronically in user-friendly formats including AtGenExpress and Genevestigator. These search engines for gene expression datasets provide a starting point for analyzing experiments to find a gene of interest, identifying either conditions that affect expression or genes that are expressed during a specific condition, and comparing gene expression among a range of conditions.

For designing metabolic engineering strategies, computational pipelines for genome-scale, subcellular compartmentalized network modeling in Arabidopsis tissues and cell cultures are available (Mintz-Oron et al. 2012). For expressing an entire metabolic pathway in a plant, binary plasmids supporting expression of up to seven genes have been developed and can be used with multiple rounds of transformation and/or genetic crosses for expression of additional genes (Shockey et al. 2015). These can be combined with sequence elements to coordinate overexpression of some genes while simultaneously silencing others, making it possible to introduce and regulate the expression of an entire heterologous pathway in Arabidopsis. Furthermore, metabolic pathway assembly has been advanced by parallel DNA assembly technologies and the exchange and reuse of DNA in multigene constructions using MoClo molecular toolkits (Golden Gate Modular Cloning; Engler et al. 2014; Owen et al. 2017).

For the metabolic engineer and synthetic biologist, microorganisms are generally simpler, faster, and easier chassis supported with a comparable depth of available resources, but for explorations in engineering a multicellular, developmentally complex, and photosynthetic organism, Arabidopsis is often the starting point.

Examples of engineering metabolism in Arabidopsis

There are many examples of engineering Arabidopsis for producing nutritionally, environmentally, and pharmaceutically relevant compounds that build on the community resources for this chassis organism (Fig. 1). Here we highlight a few examples of how researchers have targeted Arabidopsis and its metabolism for the production of a variety of biomolecules and purposes.
Targeting metabolite production

Terpenes are a class of plant-specialized metabolites that have a host of biological activities, industrial value, and use as pharmaceuticals. Isoprenoids function as membrane components, pigments, electron transport systems, phytohormones, antioxidants, fragrances, flavors, nutritional supplements, and pharmaceuticals (Lange et al. 2015). Analysis of each individual Arabidopsis gene involved in the cytosolic, peroxisomal, and plastidial pathways for isoprenoid, mevalonate, and methylerythritol phosphate synthesis evaluated transcriptional control at each step of this process and to infer flux bottlenecks (Lange et al. 2015). Such approaches using Arabidopsis provide insights into what genes and/or metabolic steps to target for engineering of high-value metabolites derived from isoprene.

Lignin, a durable aromatic structural polymer component of plant cell walls, is another attractive engineering target. This polymer is difficult to breakdown due to carbon–carbon and ether linkages between the aromatic moieties, which pose a barrier for the breakdown of plant cell wall biomass as a carbon source for bioenergy applications. Targeted alterations in the chemical composition of lignin and/or the amount of lignin can ease processing of this valuable material (Bonawitz and Chapple 2013; Vanholme et al. 2013). Arabidopsis knockout lines targeting caffenyl shikimate esterase in lignin biosynthesis resulted in more efficient breakdown of cell wall polysaccharides into simple sugars, or saccharification, relative to wild type, but with a 40% reduction in plant size (Vanholme et al. 2013). Because saccharification enzymes can be inhibited by lignin, depressing the biosynthesis of lignin precursors provides a 25% improvement in saccharification yield without impacting plant growth and biomass yield (Zhang et al. 2012). Because many of the core lignin biosynthetic enzymes are shared among Arabidopsis and other plants, findings can be translated to bioenergy crops, including poplars, eucalyptus, switchgrass, alfalfa, and maize [for specific examples, see Li et al. (2014)].

Similarly, efforts to target production of the lignin-related product sinapyl alcohol 4-O-glucoside (syringin), which also has health-promoting pharmacological activity, were successful (Chu et al. 2014). By combining expression of a chimeric UDP-dependent glucosyltransferase (UGT72E3/2),
ferulate 5-hydroxylase, and a lignin-specific transcriptional activator MYB58 to divert carbon flux towards sinapyl alcohol, a 56-fold increase in production was achieved (Chu et al. 2014). The use of a transcription factor in this study also highlights efforts to manipulate expression networks for metabolic engineering. For example, microRNAs like miR1A56 and miR828 regulate gene transcripts in the anthocyanin floral pigment pathway in Arabidopsis and suggests that regulation of microRNAs could aid tissue-specific engineering of biosynthetic pathways, especially if they target genes encoding transcription factors (Gou et al. 2011; Yuan and Grotewold 2015).

Specialty oils, such as long-chain polyunsaturated fatty acids (LC-PUFAs) found in fishes that acquire them from algae through their diet, are of interest for nutritional and industrial applications (Napier et al. 2014). Several engineering approaches have optimized production of the omega-3 fatty acids docosahexaenoic acid (DHA) and eicosapentaenoic acid (EPA) in Arabidopsis to levels comparable with fish oil (Petric et al. 2012; Sayanova et al. 2012; Ruiz-Lopez et al. 2013). LC-PUFAs have been produced in Arabidopsis seeds, where the biosynthetic precursors are found in high amounts, by tissue-specific expression of fatty acid chain elongation and desaturation enzymes from nematode (Caenorhabditis elegans) and zebrafish (Danio rerio) (Robert et al. 2005). To improve seed oil content, changing expression of one gene is often insufficient. As another example of engineering flux through metabolism using transcriptional regulators instead of single genes, seed-specific overexpression of WRINKLED1, a transcriptional regulator of glycogen and fatty acid biosynthesis, and diacylglycerol acyltransferase coupled with suppression of a tracylglycerol lipase was needed to increase seed oil percentage and seed mass (van Erp et al. 2014); however, ectopic expression of the Brachypodium distachyon WRINKLED1 led to cell death in the leaves associated with high free fatty acid content (Yang et al. 2015).

Although Arabidopsis is not ideal for industrial-scale oil production, Arabidopsis research can be translated into oilseed crops (Napier et al. 2014). To increase hydroxy-fatty acids in Arabidopsis, substituting its native phosphatidylcholine diacylglycerol cholinephosphotransferase with a castor bean homolog enhanced flux in a mutant background (Hu et al. 2012). Oils can also be produced in the leaves of Arabidopsis by expressing diacylglycerol acyltransferases and/or olcosin genes from tung tree (Vernicia fordii) in wild-type plants and mutants with disrupted fatty acid breakdown (Yurechenko et al. 2017). Accumulating oils in the vegetative biomass of plants could enhance yields beyond those of oilseed crops (Vanherecke et al. 2014). For example, transcriptome and biochemical data identified a metabolic bottleneck in Nicotiana tabacum (tobacco) that could be overcome by overexpressing the Arabidopsis LEC2 transcription factor in a lipase mutant background for a 30% increase of triacylglycerol in tobacco leaves (Vanhercke et al. 2014).

Plant metabolic engineering has the potential to enhance food nutritional content, which can serve as a preventative to chronic deficiency diseases, obesity, metabolic disorders, cardiovascular disease, and cancer (Martin and Li 2017). Vitamin B1, or thiamin diphosphate, is obtained through diet in humans and serves as a cofactor in glucose metabolism, the Krebs cycle, and branched-chain amino acid biosynthesis (Goyer 2010). While crops such as rice, wheat, and maize are poor sources of thiamin (Fitpatrick et al. 2012), studies in Arabidopsis show that overexpressing key biosynthetic enzymes significantly increases thiamin pools in leaves and seeds (Dong et al. 2015). The nutritional and dietary effects of these changes in vitamin content in a food crop remain to be evaluated, but initial studies in Arabidopsis are promising.

Similar efforts to improve the nutritional value of seeds have focused on transporters that move primary metabolites using Arabidopsis as a test platform (Weber and Brautigam 2013). Arabidopsis and other Brassicaceae contain glucosinolates, which are a diverse family of sulfur-containing metabolites. Some glucosinolates can also be anti-nutrients in animal feed. By mutating two glucosinolate transporters, the import of glucosinolates to Arabidopsis seeds was prevented, which suggested that altered tissue-specific expression of transporters could eliminate anti-nutritional compounds from valuable products (Nour-Eldin et al. 2012). Recent work translated these insights from Arabidopsis to the oilseed crops Brassica juncea (Indian mustard) and B. rapa (canola) (Nour-Eldin et al. 2017). This work and the examples above highlight the growing pullet of tools and approaches developed using Arabidopsis as a chassis organism.

**Targeting environmental applications**

The use of plants for environmental remediation dates back to the Romans, but genetic engineering approaches broaden the ability of plants to clean-up a variety of contaminants (Jez et al. 2016). Three examples highlight how Arabidopsis contributes to moving such approaches forward into real-world applications.

Many explosives, such as 2,4,6-trinitrotoluene (TNT), are residual contaminants at sites around the world, but are not easily degraded by microbes and plants. Plants can degrade TNT, but not at the level needed for viable remediation (Johnston et al. 2015). Recent work identified the mechanism behind this limitation as the generation of a nitro radical during TNT degradation (Johnston et al. 2015). This radical can react with oxygen to form toxic superoxide. By knocking out the gene for the protein that creates the nitro radical (i.e., monodehydroascorbate reductase 6),
Arabidopsis plants displayed enhanced tolerance for TNT, opening the door for using this approach in plants better suited for field remediation.

Other approaches for engineering plants for environmental applications focus on improving existing detoxification systems. Plants synthesize phytochelatins, which are heavy-metal-binding peptides of polymerized glutathione (Cahoon et al. 2015). Using protein engineering, a variant of phytochelatin synthase was obtained that enhanced cadmium tolerance and accumulation in Arabidopsis and B. juncea, which in addition to being an oilseed crop has been used for phytoremediation efforts (Cahoon et al. 2015). Unexpectedly, the evolved enzyme was catalytically inferior to the starting version, yet conferred an improved phenotype. Additional work showed that selection of a less active variant prevented the depletion of redox-active metabolites to maintain growth during cadmium exposure.

In addition to remediation technologies, the use of plants as biosensors for monitoring field contamination is under development. One example is a fluorescent zinc biosensor introduced into Arabidopsis and poplar trees to serve as sentinel plants to monitor contamination-prone areas (Adams et al. 2012). A modified zinc transporter with flanking fluorescent proteins signals when the transporter was operational. Although many years away and requiring additional work, the use of plants for remediation and monitoring of contamination could provide new tools for maintaining and improving the environment.

Plants producing bacterial genes

In addition to the chemical arsenal of plants, microorganisms contribute to the diverse toolbox of the synthetic biologist and metabolic engineer seeking to move microbial pathways into a plant. Expression of bacterial genes in the Arabidopsis can help produce specialty compounds, deregulate metabolic pathways, and detoxify environmental contaminants. Butanetrol (1,2,4-butanol), a precursor to butanetrol trinitrate used as a propellant and energetic plasticizer for the defense and energy industries is one metabolic engineering target (Abdel-Ghani et al. 2013). Using different promoters and codon-optimized synthetic genes, introduction of the four bacterial biosynthetic genes into Arabidopsis resulted in plant-based production of butanetrol (Abdel-Ghani et al. 2013).

At the interface of primary and secondary metabolism, 3-deoxy-o-arabino-heptulosonate 7-phosphate synthase regulates the shikimate pathway that gives rise to the three aromatic amino acids and their downstream metabolites. Expressing a mutant bacterial AraG gene, which encodes a feedback-insensitive variant, in transgenic Arabidopsis deregulated this bottleneck and results in increased levels of compounds ranging from aromatic amino acids to secondary metabolites such as glucosinolates and indole-3-acetic acid (the predominant auxin) (Tzin et al. 2012).

Exploiting the ability to move multiple genes into Arabidopsis yielded a new microbe-derived xenobiotic degradation pathway in a plant. This example of engineering Arabidopsis for a phytoremediation goal involved the transfer of genes encoding two subunits of naphthalene dioxygenase from Mycobacterium vanbaalenii and genes for flavoprotein reductase and ferredoxin from Pseudomonas putida to build an pathway for the detoxification of harmful polycyclic aromatic hydrocarbons (Peng et al. 2014). As with microbes, the mixing and matching of genes, proteins, and regulatory elements and the ability to ‘knock-in’ and/or ‘knock-out’ functions can introduce new pathways into plants.

Conclusions and future directions

Arabidopsis had been the model of the plant biology research community for several decades and has become a chassis for synthetic biologists. The explosion of online tools and resources to predict gene functions and to find the biosynthetic pathways responsible for this plant’s chemical diversity has only accelerated. Yet, there is more to be done. The genome of Arabidopsis was released nearly two decades, but thousands of genes remain to have their biological functions identified (Provart et al. 2015).

For all of the positive qualities of this remarkable plant, Arabidopsis is not a vegetative crop. Thus, one of the major challenges with Arabidopsis research is the translation of its outcomes to oilseed plants and/or agricultural crops; however, studies in Arabidopsis can be useful to metabolic engineering of those plants for quality trait improvement (Ewing et al. 1999–2000; Mintz-Oron et al. 2012).

Another potential issue with translating Arabidopsis metabolic engineering and synthetic biology efforts into higher plants is that the metabolic landscape in plants varies. Arabidopsis seeds are not the same as soybean (Glycine max) seeds, and legumes form symbioses with nitrogen-fixing bacteria and mycorrhizal fungi (like most other plants), while Arabidopsis does not—theyir biologies are quite different. The pool of precursors for an introduced metabolic pathway may also differ in plant species, and the regulation of pathways will likely vary. These concerns all need to be considered when translating Arabidopsis findings into other plants.

Several private biotechnology companies have tried, without great success, to use Arabidopsis as a synthetic biology platform. Area Biodetection tried to develop RedDetect, a plant-based biosensor to detect nitrogen dioxide gas released from buried landmines in the soil, using visual color changes from green to red using a combination of gene mutations and the introduction of a MYB transcription factor.
and a chalcone synthase gene (EU Joint Research Centre 2006). The company was acquired by Plant Route in 2011, and to our knowledge, an explosive-detecting Arabidopsis plant is not available (http://plant-route.dk/?page_id=10). Challenges with this technology may stem from the small ground coverage limits of Arabidopsis, which would also impose difficulties on other phytoremediation efforts using this plant. Similarly, the first crowdsourced-funded synthetic biology project, Glowing Plant from BioCurious, was supposed to deliver a glow-in-the-dark Arabidopsis to provide lighting without electricity (Zhang 2017). The Glowing Plant project faced difficulties in introducing all six necessary genes into the plant at once, so at best, the plants glowed dimly. Although unsuccessful, the project did spark public interest in plant-based synthetic biology.

While we understand that Arabidopsis itself will never be a production chassis, we do advocate its use for proof-of-concept work. The genetic, genomic, and biochemical tools, online resources, and large research community make this a desirable organism as a test platform. There are challenges, but the ease of transformation and short lifespan make work in Arabidopsis fast and relatively easy for the synthetic biologist-turned-plant biologist. By unlocking the potential to improve plants and to use plants as chassis for more nutritious food with fewer inputs, biofuel, or biopharmaceuticals, scientists have the opportunity to change the way we feed, power, build, clothe, and care for humans of the future.

Acknowledgements The authors acknowledge support from the National Science Foundation (MCB-1614539 to JM) and DGE-1143954 to CKH.

Author contribution statement Both authors researched and wrote this review.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References


274


274
Appendix IV

Structural Biology of Jasmonic Acid Metabolism and Responses in Plants

Cynthia K. Holland¹, Joseph M. Jez¹

¹Department of Biology, Washington University in St. Louis, St. Louis, MO 63130

Jasomates (JA) constitute a diverse family of modified oxylipins that are involved in hormonal defense responses, growth regulation, and many other aspects of plant physiology. To respond to external cues, plants synthesize jasmonyl-isoleucine (JA-Ile), the active form of the hormone, using the enzyme JAR1. JA-Ile then targets the jasmonate F-box receptor, the COI1-ASK1 complex, and recruits JASMONATE-ZIM DOMAIN PROTEIN (JAZ), resulting in the degradation of this repressor to regulate jasmonate gene responses. Alternatively, when JA-Ile levels decline, JAZ binds an array of co-repressors and transcription factors to down-regulate JA responses. Three-dimensional structures of enzymes, proteins, and peptides involved in JA signalling shed light on the molecular architecture of the binding interactions between proteins and the specific chemistry of the enzymes. Additionally, structures of the JA biosynthesis and modifying enzymes provide insight into the strict stereo-specificity of the formation of the precursor to active JA, 3R, 7S-jasmonate, as well as decorated JAs. Here we provide an up-to-date look into the structural investigations in JA signaling and metabolism and suggest future areas of study that remain to be understood on the molecular level.
Chapter 5

Structural Biology of Jasmonic Acid Metabolism and Responses in Plants

5.1 Introduction to Jasmonates

To coordinate growth and development during times of stress and predation, plants evolved the capability of synthesizing and responding to an array of complex modified oxylipins, or oxygenated fatty acid derivates, known as jasmonates (Kombrink 2012). Methyl jasmonate (MeJA) was first isolated from the oil of jasmine (*Jasminum grandiflorum*), although its effect on plant physiology remained elusive for nearly two decades (Demole et al. 1962; Ueda and Kato 1980; Dathe 1981). Since this discovery, jasmonates have been recognized as critical phytohormones involved in plant defense and fertility.

Chemically, the jasmonates encompass jasmonic acid (JA) and its derivatives, including MeJA, 3R, 7S-jasmonyl-L-isoleucine (JA-Ile), cis-jasmone, and jasmonyl-1-aminocyclopropane-1-carboxylic acid (JA-ACC) (Figure 1a) (Browse 2009). Jasmonate biosynthesis proceeds through the lipooxygenase-catalyzed peroxidation of linoleic acid to form 13-hydroperoxide, which is modified and cyclized to form 12-oxo-phytodienoic acid (OPDA). OPDA is reduced and then undergoes three rounds of β-oxidation to form JA. Further biochemical modifications of JA, such as glycosylation, decarboxylation, reduction of either the C₆ carbonyl or C₉,₁₀ double bond, or hydroxylation of either C₁₁ or C₁₂, can convert the core JA scaffold into over 30 different jasmonate compounds with diverse physiological roles.
Investigations of the role of jasmonates in plant physiology reveal their role in stamen and trichome development, vegetative growth, cell cycle regulation, senescence, regulation of anthocyanin pigment biosynthesis, and responses to various biotic and abiotic stresses (Creelman and Mullet 1997; Wasternack 2007; Howe and Jander 2008; Browse 2009; Avanci et al. 2010; Pauwels and Goossens 2011). Jasmonates are also key players in induced resistance of plants (Kunkel and Brooks 2002), including Rhizobacteria-mediated induced systemic resistance (Pieterse et al. 1998).

As one of the major plant defense compounds, jasmonates induce defense mechanisms to ward off insects, herbivores, and a broad spectrum of fungal and bacterial pathogens (Howe and Jander 2008; Browse 2009). Wounding of the plant tissue leads to elevated levels of jasmonates at the site of attack, which in turn induces a systemic increase of JAs through the regulation of JA-responsive genes that allow for production of diverse compounds as the attack continues (Reymond et al. 2004; Glauser et al. 2008). These compounds include glucosinolates, camalexins, alkaloids, artemisinin, and volatile organic compounds, such as terpenes, which provide a chemical arsenal to combat predators (Farmer et al. 1992; Bolter 1993; Paré and Tumlinson 1999; Engelberth et al. 2004; De Geyter et al. 2012).

When stressed or otherwise stimulated, various jasmonates accumulate and are converted to JA-Ile, the only known bioactive jasmonate hormone (Figure 1b) (Wasternack and Hause 2007; Fonseca et al. 2009a; Fonseca et al. 2009b). JA-Ile forms a co-receptor complex with the F-box protein CORONATINE INSENSITIVE 1(COI1) and a JASMONATE-ZIM DOMAIN PROTEIN (JAZ). This complex allows for ubiquitinylation of the JAZ protein, which results in its degradation by the 26S proteasome (Fonseca et al., 2009a; Thines et al., 2007; Chini et al., 2007). JAZ proteins are nuclear repressors of transcription factors and repress JA-responsive
Figure 1. Overview of jasmonate molecules and jasmonate-linked responses. a) Chemical structures of representative jasmonates. The * denotes a chiral center that can be either in the $R$ or $S$ configuration. The carbon-carbon double-bond drawn as a squiggle can be either cis or trans. b) Overview of jasmonate signaling and gene regulation pathways. Stimulus that leads to production of JA-Ile by the enzyme JAR1 results in perception of the bioactive hormone by the COI1-ASK1 receptor. Ubiquitinylation of JAZ proteins by the complex activates JA-responsive gene expression. Under low JA conditions, JAZ proteins interact with different transcriptional factors, such as TOPELESS and MYC3, to repress expression.

genes through recruiting the co-repressor TOPELESS (TPL) by the adaptor protein NOVEL INTERACTOR OF JAZ (NINJA) and also competing with MEDIATOR25 (MED25) for interaction with members of the MYC transcription factor family (Pauwels et al. 2010; Zhang et al. 2015). Degradation of JAZ proteins thereby alters transcriptional patterns in response to JA-mediated signals.

Structural efforts to characterize the JA biosynthetic enzymes, starting from chloroplast membrane lipids, have allowed for a better understanding of the generation of $3R$, $7S$-jasmonate needed to produce JA-Ile to elicit a physiological response. Although the jasmonate biosynthesis and JA-related signaling pathways are well-studied, atomic-level studies aimed at deciphering the mechanisms and molecular interactions responsible for large-scale physiological responses
are just beginning. Recent structural investigations reveal the biochemistry and protein-protein interaction details of enzymes and proteins involved in bioactive jasmonate biosynthesis, metabolism, perception, and transcriptional responses.

5.2 Production of JA-Ile, the Bioactive Jasmonate Phytohormone

Plant responses to developmental cues and the environment are primarily modulated by the activity and concentration of phytohormones (Westfall et al. 2010). Fluctuation in concentrations of plant growth regulators and modifications of the chemical structure of a hormone affects receptor interactions to alter developmental responses. Although jasmonates can be decorated with methyl groups, sugars, amino acids, and other chemical moieties (Staswick and Tiryaki 2004), the functions of all of the various modified jasmonates remain unclear. Presumably, plants have developed mechanisms, such as conjugation with amino acids, to regulate hormone levels. For example, conjugation of isoleucine to JA by the acyl acid amido synthetase JAR1 gives rise to JA-Ile, which is the active form of the hormone (Figure 2a) (Staswick et al. 2002).

In plants, the GH3 (Gretchen Hagen 3) acyl acid amido synthetases contribute to maintaining hormone levels in plants by conjugating a hormone, including auxins, jasmonates, and benzoates, to an amino acid as either a storage or activation mechanism (Hagen and Guilfoyle 1985; Westfall et al. 2013). The biological function of this class of enzymes was first established when the jasmonate resistant 1 (jar1-1) mutant was identified in Arabidopsis thaliana and shown to exhibit decreased sensitivity to exogenous MeJA and reduced male fertility (Staswick et al. 1992, 2002). The A. thaliana JAR1 protein (also known as AtGH3.11) functions as a JA-amido synthetase and conjugates Ile to JA (Staswich and Tiryaki 2004). AtGH3.11/JAR1 functions as a
Figure 2. Conversion of JA to the bioactive jasmonate JA-Ile. a) Overall reaction catalyzed by the JAR1 acyl acid amido synthetase. Each half-reaction is shown. b) Three-dimensional structure of JAR1 from Arabidopsis thaliana. The ribbon diagram of JAR1 in complex with JA-Ile (space-filling model; purple) is shown. The N- and C-terminal domains of JAR1 are colored gold and blue, respectively. The 'hinge-loop' that alters in conformation, along with the C-terminal domain, is indicated. c) Active site of JAR1. This view shows a close-up of the JAR1 active site with JA-Ile bound. Key residues are noted. The ATP binding site, based on other GH3 protein structures, is also indicated.

A

B

C

monomer to catalyze the adenylation of JA using ATP in the first half reaction of a ping-pong kinetic mechanism with subsequent AMP release and JA-amino acid conjugate formation in the second half reaction (Chen et al. 2010). JA-Ile can then regulate JA-responsive genes through binding the F-box receptor protein COI1 (Feys et al. 1994; Xie et al. 1998).

The x-ray crystal structure of AtGH3.11/JAR1 was determined in complex with JA-Ile to reveal its catalytic mechanism (Westfall et al. 2012). The overall structure of AtGH3.11/JAR1 consists of a large N-terminal domain containing α-helices that flank a β-barrel and two β-sheets; the C-terminal domain is related to the adenylating firefly luciferase (ANL) enzyme superfamily (Figure 2b) (Gulick 2009). The active site is located at the intersection of the two
domains with a hinge loop between Leu\textsuperscript{427}-Arg\textsuperscript{439} providing the conformational flexibility necessary for pivoting the C-terminal domain at each stage of catalysis.

The structure of AtGH3.11/JAR1 in complex with JA-Ile provides information on how the enzyme binds the oxylipin molecule (Figure 2c). At the interface of the two domains, JA-Ile is situated in a large apolar binding site. Multiple residues, including Thr166, Val222, Phe223, and Ile304 encompass the cyclopentane ring of JA-Ile with a water-mediated hydrogen bond formed with His328. The pentenyl tail of JA-Ile binds in a pocket formed by Leu117, Thr121, Phe125, Thr166, Val169, and Trp336. The isoleucine moiety is oriented with its carboxylate toward residues of the nucleotide binding P-loop. Although no nucleotide was bound in the structure, the highly conserved adenosine binding pocket architecture can be inferred based on the structure of another \textit{A. thaliana} GH3 protein (i.e., AtGH3.12/PBS3) (Westfall et al. 2012). The nucleotide-binding site consists of a canonical phosphate-binding loop motif, a \(\beta\)-turn-\(\beta\)-structure motif, and a ribose interaction motif. When the C-terminal domain is oriented in the “open” active site conformation, the active site allows for binding of ATP, Mg\textsuperscript{2+}, and JA for the adenylation reaction to occur. Upon adenylation, pyrophosphate is released, and rotation of the C-terminal domain is triggered, trapping the adenylated intermediate for the second half-reaction. This rotation also creates a tunnel that allows access of isoleucine into the active site for nucleophilic attack on the intermediate. This releases AMP and results in formation of the conjugated product JA-Ile.

\textbf{5.3 Perception of JA-Ile by the Hormone Receptor F-box Protein COI1}

The hormone receptor for JA-Ile is the F-box protein COI1, which functions in recruiting substrates to the Skp1-Cullin1-F-box protein (SCF) ubiquitin E3 ligase complex (Feys et al.}
In the presence of JA-Ile, JAZ transcriptional repressors bind SCF\textsuperscript{COI1} (Chini et al. 2007; Yan et al. 2007). The active jasmonate hormone mediates protein-protein interaction, which leads to the poly-ubiquitylation and degradation of JAZ proteins by the 26S proteasome. This relieves repression of the MYC2 transcription factor, or other transcription factors, and allows for transcriptional changes in jasmonate-responsive genes (Lorenzo et al. 2004; Chini, et al. 2007).

**Figure 3. Perception of JA-Ile by the Jasmonate Receptor COI1.** a) Overall structure of the COI1-ASK1 jasmonate receptor complex. The overall structure is shown as a ribbon diagram with JA-Ile (green) and the JAZ1 peptide (purple) shown as space-filling models. b) JA-Ile as the molecular glue between COI1 and JAZ proteins. Top view of the COI1-ASK1 complex shown as a molecular surface model with JA-Ile (green) and the JAZ1 peptide (purple) represented as a space-filling model and a ribbon, respectively. c) Interactions in the JA-Ile binding site. A close-up view of the JA-Ile (green) binding site and polar interactions (dotted lines) made within the active site pocket with both COI1 (white) and JAZ1 (purple) is shown.

Structural studies of the *A. thaliana* COI1-ASK1 complex bound to either JA-Ile or the jasmonate mimic coronatine and a peptide of a bipartite JAZ1 degron reveals how the hormone acts as a molecular glue in this macromolecular assembly (**Figure 3a**) (Sheard, Tan, and Mao et al. 2010). The COI1-JAZ co-receptor forms through interactions between the C-terminal Jas motif of the JAZ1 protein (Glu\textsuperscript{200}-Val\textsuperscript{220}). The overall structure of COI1 resembles the F-box
complex of the auxin receptor, TIR1 (Tan et al. 2007). COI1 consists of a tri-helical N-terminal F-box motif that binds ASK1 and a horseshoe-shaped C-terminal solenoid of 18 tandem leucine-rich repeats (LRRs). The top of the LRR domain binds the hormone and the C-terminal Jas motif of JAZ1 through three intra-repeat loops (Loop 2, 12, and 14), and a fourth loop (Loop C) partially covers Loop 2. Large, polar residues provide multiple contacts for formation of the functional complex (Figure 3b). Any mutations of residues in the site disrupt COI1-JAZ1 interactions. The keto group of the ligand, either JA-Ile or coronatine, points upward and interacts with COI1 residues at the entrance of the pocket and is solvent exposed in the absence of the JAZ degron. Three arginine residues (Arg85, Arg348, and Arg409) bind the amide and carboxyl groups of the ligand forming the floor of the pocket. Additional density in the structure led to the conclusion that inositol pentakisphosphate (InsP$_5$) binds below the hormone binding pocket. Although the specific binding interactions and the mechanism remain to be explored, InsP$_5$ was identified as a critical component of the co-receptor for sensitive jasmonate perception. Therefore, the jasmonate receptor complex is a three-molecule complex, requiring COI1, the JAZ degron, and InsP$_5$ for hormone binding.

The complete structure of a JAZ protein remains to be determined. JAZ proteins appear to be intrinsically disordered in their native form, which allows for repressor plasticity in its binding interactions (Pazos et al. 2013; Pietrosemoli et al. 2013). *A. thaliana* contains 12 JAZ proteins that are thought to be redundant in signaling pathways as repressors (Chini et al. 2007; Yan et al. 2007; Browse 2009). In the COI1-ASK1 structure, the degron peptide of JAZ1 forms a bipartite structure. Five N-terminal amino acids interact with COI1 and completely cover the opening where JA-Ile binds, which traps the ligand in the pocket. These co-receptor interactions explain the stereospecific preferences for (3R, 7S)-JA-Ile versus (3R, 7R)-JA-Ile, as the
stereochemistry may place the aliphatic chain unfavorably close to nearby JAZ1 and COI1 residues. The α-helix of the C-terminus of the JAZ1 degron binds the surface of the COI1 LRR domain, thus strengthening the interactions that form the co-receptor complex. There is considerable variability on the amino acid sequence level of the N-terminal clamp region of the JAZ degron among JAZ proteins. Further investigation is needed to tease apart these interactions between COI1 and the variable JAZ proteins and the responses of these co-receptors.

5.4 Transcriptional Regulation of Jasmonate-Responsive Genes

In the absence of JA-Ile, JAZ repressors bind either helix-loop-helix transcription factors or adaptor proteins to recruit co-repressors needed to downregulate jasmonate-targeted gene expression (Chini et al. 2007; Thines et al. 2007; Pauwels et al. 2010; Ke et al. 2015). The general co-repressor TOPLESS (TPL) can be recruited by an array of repressors, transcription factors, and transcriptional adaptor proteins (Causier et al. 2012).

TPL regulates gene expression by interacting with the ethylene response factor-associated amphiphillic repression (EAR) motifs of certain transcriptional repressors (Causier et al. 2012). The N-terminal domain of TPL, or the TOPLESS domain (TPD), interacts with EAR motifs of NINJA proteins, and also JAZ repressors themselves, when JA-Ile levels are low. The N-terminal domain contains a lissencephaly homologous (LisH) dimerization motif and a C-terminal to LisH (CTLH) motif, followed by a proline- and glutamine-rich linker that connects the N-terminal domain to the two WD40 domains, a 40 amino acid motif that terminates with a tryptophan-aspartate dipeptide, of the C-terminus (Figure 4a). X-ray crystal structures of the TPD from A. thaliana and rice (Oryza sativa) revealed a dimer of dimers, with each monomeric unit containing two dimerization interfaces with a CTLH motif sandwiched in between (Figure
Each TPD monomer contains nine α-helices and two small 3_{10} helices. Co-crystallization of a TPD with peptides of the NINJA EAR motif showed that four NINJA peptides could bind one TPD tetramer in a shallow groove composed of the CTLH motif (Figure 4c). Homologous WD40 domains in animals and yeast co-repressors, notably Groucho and Tup1, have been found to interact with repressors, but repressors that interact with plant TPL WD40 domains remain to be identified, which would give further insight to the overall roles of TPL in plant growth and development (Goldstein et al. 1999; Jimenez et al. 1997; Muhr et al. 2001; Paroush et al. 1994; Zhang et al. 2001; Komachi et al. 1994).

**Figure 4. Transcriptional Control of Jasmonate Responses.** a) Structural insight on TOPLESS (TPD)-NINJA interaction. A ribbon diagram of the tetrameric N-terminal domain of TOPLESS (TPD) co-complexed with NINJA peptides (green space-filling molecule) is shown. Each monomer of the TPD is colored differently. b) Interaction between the LisH (blue) and CTLH (rose) domains of the TPD (gold; domains colored separately). c) NINJA binding site. The NINJA peptide (green) binds TPD near the CTLH domain (rose). d) Structure of the MYC3 transcription factor. The ribbon diagram shows the overall structure of MYC3 and is colored by secondary structure (α-helices, rose; β-strand, white). e) JAZ protein interaction promotes transcriptional activation. Binding of the Jas motif of a JAZ protein (purple ribbon) between Jas-interacting domain (JID) and the transcription activation domain (TAD) of MYC3 alters interaction with the MED25 subunit of the Mediator complex.
While the Jas domain of JAZ proteins is an integral part of the COI1 F-box receptor complex, this domain also directly regulates the transcription factor MYC3 (Zhang et al. 2015). *A. thaliana* MYC proteins contain a Jas-interacting domain and a transcription activation domain. In the three-dimensional structure, MYC3 forms a helix-sheet-helix sandwich fold of five anti-parallel β-sheets surrounded by eight α-helices (*Figure 4d*). This structure was the first example of a non-complexed well-resolved MYC transcription activation domain, as they are generally unstructured in their unbound state. Additionally, MYC3 has been co-crystallized with a Jas peptide (i.e., Ser218-Met239) of JAZ9 (*Figure 4e*) (Zhang et al. 2015). The Jas peptide was nestled between the Jas-interacting domain (JID) and the transcription activation domain (TAD). While the first five amino acids of the Jas domain peptide are necessary for COI1-JAZ binding interactions, they do not form critical interactions with MYC3. The Jas peptide was a continuous helix, in contrast to its bipartite loop and short C-terminal helix in the COI1 co-receptor complex structure. This suggests that JAZ proteins employ a switch mechanism to form distinct confirmations in the MYC-JAZ transcriptional repression resting state versus the COI1-JAZ activated complex. Such a mechanism implies that MYC and COI1 potentially compete for binding the Jas domain of JAZ proteins. Elucidation of how MYC and Jas domains interact provided the first molecular insight on how transcriptional repression is achieved in a plant hormone response pathway.

To promote transcription of jasmonate responsive genes, RNA polymerase II is recruited to promoters by the action of the MED25 subunit of the Mediator complex (Kazan and Manners 2013). MYC3 directly binds MED25, although JAZ repressors can outcompete MED25 for MYC3. While MED25 is a critical component of gene regulation in response to jasmonates, biophysical protein-protein interaction and structural data is absent and would provide insight
into the flexibility of MYC3 binding both MED25 and JAZ repressors (Chen et al. 2012; Cevik et al. 2012; Kidd et al. 2009). For further information on the MYC family of transcriptional activators, basic helix-loop-helix repressors, and other JAZ targets readers are referred to a recent review (Chini et al. 2016).

5.5 Biosynthesis and Chemical Modifications of Jasmonates

5.5.1 Jasmonic Acid Biosynthesis

To gain an understanding of JA-signaling regulation, structural efforts have also focused on the biosynthesis of jasmonates and their stereo-selectivity in generating 3R, 7S-jasmonate, the precursor to JA-Ile. Jasmonates, while they are a diverse class of compounds, all stem from the same polyunsaturated fatty acid (PUFA) intermediates, linoleic acid (18:2) and hexadecatrienoid acid (16:3), from the chloroplast membrane (Schaller and Stintzi 2009). These PUFAs can be oxidized to form fatty acid regio- and stereo-specific hydroperoxides using lipoxygenases within the chloroplast. The oxygenated fatty acid hydroperoxide products are dehydrated to form unstable allylic epoxides, or allene oxides, which are cyclized by allene oxide cyclases to form optically pure (9S,13S)-oxophytodienoic acid (OPDA). OPDA is then transported out of the chloroplast by an unidentified transporter.

X-ray crystal structures of the CYP74 allene oxide synthase (AOS) from A. thaliana and guayule (Parthenium argentatum) (Lee et al. 2008; Li et al. 2008) reveal that plant CYP74s share a common fold with other P450 enzymes with a heme-binding loop that is eight amino acids longer than similar P450s in other organisms (Fig. 5a). The allene oxide cyclase 2 (AOC2) crystal structure gave insight to the reaction chemistry, specificity, and evolution that gives rise to the first cyclic, optically pure bioactive molecule in the JA biosynthetic pathway, OPDA.
(Hoffman et al. 2006). The AOC2 trimer contains a lipocalin fold comprised of an eight-stranded antiparallel β-barrel followed by a partially helical C-terminus (Fig. 5b). To identify the active site, the enzyme was co-crystallized with the inhibitor vernolic acid, which was bound in the hydrophobic barrel cavity. Modeling of the 12,13(S)-epoxy-9(Z),11,15(Z)-octadecatrienoic acid (12,13-EOT) substrate and the OPDA product into the active site cavity led to the prediction of the mechanism of cyclization, which includes the opening of the epoxide to form an anionic intermediate followed by a trans to cis conformational change between C10 and C11 to allow for the pericyclic ring closure with absolute stereo-selectivity.

To complete jasmonate biosynthesis, the peroxisomal ATP-binding cassette (ABC) transporter COMATOSE (CTS) (Footit et al. 2002; Theodoulou et al. 2005) transports OPDA and other β-oxidation substrates into the peroxisome. Once inside, OPDA is reduced to 3-oxo-2-(2’(Z)-pentenyl)-cyclopentane-1-octanoic acid (OPC-8:0). The crystal structure of tomato 12-OPDA reductase 3 has revealed the active site architecture responsible for the substrate stereo-selectivity and its regulation through self-inhibited dimerization (Breithaupt et al. 2006). A Co-A ester is added by ATP-dependent acyl-activating enzymes to prepare OPC-8:0 for β-oxidation. OPC-8:CoA ligase 1 (OPCL1), is the only enzyme in the A. thaliana acyl-activating enzyme family that is JA-specific (Koo et al. 2006). Two or three rounds of β-oxidation result in the shortened JA-CoA, and a thioesterase cleaves the CoA moiety to release free JA.

Overall, the crystallographic investigations of enzymes in jasmonate biosynthesis have revealed the mechanisms underlying the variety of oxylipins that can be generated while maintaining strict specificity; however, the roles of JA biosynthesis pathway intermediates in signaling remain to be understood.
5.5.2 Methyl-Jasmonate

Plants use volatile emissions to attract pollinators and seed dispersers and repel unwanted herbivores (Seo et al. 2001). MeJA is one such volatile involved in communicating defense responses within the organism and amongst plants (Farmer and Ryan 1990; Seo et al. 2001; Karban et al. 2000; Kessler et al. 2006; Baldwin et al. 2006). Plant exposure to MeJA triggers JA-mediated defense responses, and exogenous MeJA application can lead to herbivore resistance in many plant species (McConnel et al. 1997; Baldwin 1998; Li et al. 2002; Wu et al. 2008).

To generate MeJA, the enzyme JA carboxyl methyltransferase (JMT) uses S-adenosyl-L-methionine (SAM) to catalyze the methylation of JA (Seo et al. 2001). Although a JMT structure remains to be determined, the x-ray crystal structure of a related enzyme, the *Clarkia breweri* salicylic acid carboxyl methyltransferase (SAMT) (Zubieta et al. 2003), was solved. Using this structure as a guide, the active site architecture can be inferred by sequence alignment, which predicts a highly conserved SAM/SAH binding site. *Clarkia* SAMT and *A. thaliana* JMT vary by six residues in the active site (Tyr147 to Ser, Met150 to His, Ile225 Gln, Met308 to Ile, Phe347 to Tyr, and Asn349 to Ile), and mutagenesis of these residues was able to introduce methylation of JA into the SAMT but not a complete switch in specificity. A structure of a plant JMT would allow for a more thorough investigation of key residues involved in JMT specificity.

As noted above, the exogenous application of MeJA elicits a physiological reponse, which results from metabolism of MeJA to JA (Wu et al. 2008). Although a MeJA-esterase catalyzes the hydrolysis of MeJA (Stuhlfelder et al. 2002, 2004), the three-dimensional structure of this enzyme remains to be determined and would contribute to our understanding of the molecular basis of plant JA perception.
5.5.3 Homeostasis of JA-Ile: Hydroxylation, Carboxylation, and Turnover

Mechanisms governing the homeostasis of JA-Ile are essential to our understanding of plant defenses (Koo et al. 2011). Cytochrome P450 oxygenases (CYPs) constitute a vast family of heme-containing enzymes that catalyze both monooxygenations and hydroxylations (Schuler and Werck-Rechhart 2003). CYPs can also regulate hormonal pathways by modifying the chemical structures of phytohormones (Mizutani and Ohta 2010). The JA metabolite 12-hydroxy-JA-Ile forms through hydroxylation of JA-Ile by both CYP94B3 and CYP94B1 (Koo et al. 2011, 2014). CYP94B3 is specific for JA-Ile and is unable to hydroxylate JA to form 12-hydroxy JA or 12-hydroxy-JA-Ile to form 12-carboxy-JA-Ile. In contrast, CYP94B1 can convert JA-Ile to 12-carboxy-JA-Ile, although to a lesser extent (Koo et al. 2011, 2014; Kitaoka et al. 2011; Heitz et al. 2012). Together, these two enzymes account for more than 95% of 12-hydroxylation of JA-Ile and are localized to the endoplasmic reticulum (Koo et al. 2014). The carboxy-derivative of JA is preferentially formed from 12-hydroxy-JA-Ile by a third CYP, CYP94C1 (Heitz et al. 2012). Because these CYP enzymes inactivate JA-Ile, structural studies are necessary to provide insight to the evolution of CYP substrate specificity and to understand the mechanisms underlying JA modifications as a means of decreasing the pool of physiologically active JA.

While JA-Ile is the active form of the jasmonate hormone, the amino acid must be removed for it to be deactivated (Widemann et al. 2013). Two amidohydrolases, IAR3 (indole-3-acetic acid (IAA)-Ala resistant 3) and ILL6 (IAA-Leu resistant1-like 6), have been characterized in the catalysis of JA-Ile turnover. Not only does IAR3 cleave JA-Ile, but it also hydrolyzes the 12-hydroxy-JA-Ile conjugate. Although there is an x-ray crystallographic structure for the IAA-
amino acid hydrolyzing ILL2, the structure of a JA-Ile specific amidohydrolase remains to be determined (Bitto et al. 2009).

5.6 Summary

In the past decade, the structural studies have deepened our understanding of the complexity regulating jasmonate biosynthesis and signaling. Active jasmonate is synthesized by the conjugation of the amino acid Ile to JA by JAR1, which is perceived by the COI1-ASK1 complex. In the presence of JA-Ile, the JAZ repressor binds the JA-Ile-bound COI1-ASK1 complex, which leads to the degradation of JAZ by the 26S proteasome. When JA-Ile levels decline, JAZ recruits the co-repressor TOPLESS through interactions with the adaptor protein NINJA to bind MYC repressors. While structural work on this signaling pathway has revealed the general scheme, efforts to understand the JAZ binding interactors and the mechanisms governing selectivity in this large family of repressors are necessary to gain a full appreciation of the complexity of JA-responsive gene regulation. The adapter protein NINJA has only recently been discovered, and the C-terminal domain should be investigated for its role in WD40 domain interactions in signaling and for, potentially, the discovery of cross talk between the JA signaling pathway and other hormone pathways. Jasmonates encompasses a broad family of compounds, and the physiological role of these decorated jasmonates continue to escape our understanding. Future investigations are needed to elucidate the physiological and potential signaling role of the diversity of jasmonate compounds, as well as the role of the modifying enzymes in regulating the JA pool, and thus regulating JA responses. Overall, structural, biochemical, and biophysical studies have brought to light the intricacies of JA signaling, but a great deal remains to be discovered about the untapped potential of other JAs in influencing physiological responses.
Figure 5. Overview of JA biosynthesis and structures of key enzymes. Biosynthesis of JA begins in the chloroplast with the conversion of α-linolenic acid to (9S, 13S)-OPDA. X-ray crystal structures of AOS and AOC, which catalyze the second and final steps in the chloroplast pathway, provide chemical insight on these biosynthetic conversions. OPDA is transported from the chloroplast to the peroxisome, where the enzyme OPR3 forms OPC-8. Subsequent β-oxidation of OPC-8 leads to formation of JA.

5.7 References


Feys BJF, Benedetti CE, Penfold CN, Turner JG (1994) Arabidopsis mutants selected for resistance to the phytotoxin coronatine are male sterile, insensitive to methyl jasmonate, and resistant to a bacterial pathogen. *Plant Cell* 6: 751-759


Curriculum Vitae

Cynthia K. Holland  
Department of Biology, Washington University in St. Louis  
One Brookings Drive, Campus Box 1137, St. Louis, MO 63130  
Cynthia.holland@wustl.edu | 501-351-7228

EDUCATION

2018 Washington University in St. Louis, St. Louis, MO  
Ph.D., Biology and Biomedical Sciences with a Teaching Citation, Degree to be conferred in August  
Thesis: Regulation of aromatic amino acid metabolism in plants  
Research Advisor: Prof. Joseph Jez

2014 Henderson State University, Arkadelphia, AR  
B.S., summa cum laude, Biology with Honors and Chemistry minor  
Research Advisor: Prof. Martin Campbell

PROFESSIONAL EXPERIENCE

Jul-Nov 2017 Mason Communications, St. Louis, MO  
Freelance Science Writer  
Assisted with a technical brochure and web content for a biopharmaceutical testing company

2012-2014 Henderson State University, Arkadelphia, AR  
Undergraduate Research Assistant  
Project: A phytochemical investigation of the secondary metabolites in deciduous holly (Ilex decidua)

SU 2013 Donald Danforth Plant Science Center, St. Louis, MO  
NSF REU Intern in the Laboratory of Toni Kutchan  
Project: Bioengineering the oilseed crop Camelina sativa for production of anti-cancer compounds

SU 2012 North Carolina State University, Raleigh, NC  
NSF REU Intern in the Laboratory of DeYu Xie  
Project: Improvement of Camelina sativa seeds for biofuel production

PUBLICATIONS

Springer-Verlag, NY (in press)

300


**INVITED TALKS**


**PRESENTATIONS**


10. Holland CK, Jez JM (September 2017) “Regulation of aromatic amino acid biosynthesis in plants.” *Oral Presentation*, Bioforum, Washington University in St. Louis Department of Biology, St. Louis, MO.


7. **Holland CK** and Jez JM (October 2016) “Structural basis for the evolution of tyrosine biosynthesis pathways in legumes.” *Oral Presentation*, Washington University in St. Louis Plant and Microbial Biosciences Program Retreat, Eureka, MO.


4. **Holland CK** and Jez JM (February 2016) “Structural insights into an aspartate aminotransferase gain-of-function mutant.” *Oral Presentation*, Plant and Microbial Biosciences Lunch Seminar, Washington University in St. Louis, Department of Biology, St. Louis, MO.


1. **Holland CK** and Campbell MC (October 2013) “Determination of bioactivity in *Ilex decidua.*” *Oral presentation*, Arkansas IDeA Networks of Biomedical Research Excellence, University of Arkansas, Fayetteville, AR.

**Fellowships, Awards, and Professional Activities**

**2017**  
Jul 2017 Discussion Leader, Gordon Research Seminar: Plant Metabolic Engineering  
Feb 2017 Third Place in the Sciences, Washington University Graduate Student Research Symposium  

**2016**  
Apr 2016 ASBMB Experimental Biology Meeting Travel Award  

**2015-**  
2015- William H. Danforth Plant Sciences Fellow  
2014-2017 National Science Foundation Graduate Research Fellow  
2013 American Society of Pharmacognosy Undergraduate Research Award  
2013 Arkansas Academy of Science Undergraduate Award  
2012 Arkansas Department of Higher Education SURF Award
**TEACHING EXPERIENCE**

**Washington University in St. Louis, St. Louis, MO**

**Student Mentoring**
- **SP 2018** Bram Osterhout, undergraduate student, Biochemistry major
- **SP 2018** Emily Walters, high school student, Fort Zumwalt High School Independent Research Project
- **FL 2017-SU 2018** Daniel Berkovich, undergraduate student, Molecular Biology major
- **SU 2017** Alejandro De Santiago-Perez, HHMI EXceptional Research Opportunities Program, Biochemistry major at UC-Riverside
- **SU 2017** Poorva Sheth, high school student
- **SP 2017-SP 2018** Madeleine Kohn, undergraduate student, Biology major
- **SU 2016-SP 2017** Regina Liu, undergraduate student, Chemistry major with English Honors
- **SU 2016 & SU 2017** Daniel Berkovich, high school student, Students and Teachers as Research Scientists Program
- **FL 2015-SP 2017** Kourtney Kroll, undergraduate student, Biomedical Engineering major
- **SU 2015** Keishla Sanchez, high school student, Washington University Young Scientists Summer Focus Program

**Graduate Teaching Assistant**
- **FL 2017** BIO3010: Biotechnology Project for HHMI Biotech Explorers Pathway
- **SP 2016** BIO4522: Laboratory in Protein Analysis, Proteomics, and Protein Structure

**Discussion Leader**
- **Aug 2017** Washington University School of Medicine Graduate Student Teaching Assistant Orientation
- **FL 2016** BIO5723: Graduate Seminar in Plant and Microbial Bioscience

**Henderson State University, Arkadelphia, AR**

**Undergraduate Teaching Assistant**
- **SP 2013 & SP 2014** Introduction to Biology Laboratory

**Undergraduate Laboratory Assistant**
- **FL 2012-SP 2014** University Chemistry I and II Laboratories, General Organic and Biochemistry Laboratory

**SCIENTIFIC OUTREACH**

- **Feb 2018** Invited Speaker, “Presenting your science to the non-scientist,” WUSTL Graduate Student Senate Pre-Symposium Workshop on Poster Preparation
- **Apr 2015 & Mar 2017** Guest Panelist on NSF GRFP, Journal Club for Undergraduates in Biomedical Engineering and Sciences, Washington University in St. Louis
- **Apr 2014** “Chemistry in the Garden” Community Class Lecturer, Henderson State University
- **Oct 2013** Family Science Night Organizer, Henderson State University
MEMBERSHIPS

Community Member Level, Washington University Center for the Integration of Research, Teaching, and Learning (CIRTL)
American Society of Biochemistry and Molecular Biology
American Society of Plant Biology
Young Scientists Program, Washington University School of Medicine