Winter 12-15-2015

The Identification of Alkaloid Pathway Genes from Non-Model Plant Species in the Amaryllidaceae

Matthew B. Kilgore
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Plant Biology

Dissertation Examination Committee:

Toni Kutchan, Chair
Elizabeth Haswell
Jeffrey Henderson
Joseph Jez
Barbara Kunkel
Todd Mockler

The Identification of Alkaloid Pathway Genes from Non-Model Plant Species in the Amaryllidaceae

by
Matthew Benjamin Kilgore

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

December 2015
St. Louis, Missouri

Table of Contents

List of Figures

Appendix A Figures

Appendix B Figures

Appendix C Figures

Appendix D Figures

List of Tables

Acknowledgments

First, I would like to thank the Donald Danforth Plant Science center for my graduate stipend, the Division of Biology and Biomedical sciences (Washington University in St. Louis) for fanatical support, the National Institutes of Health award number 1RC2GM092561 (NIGMS) for the support of the work, and the National Science Foundation Grant No. DBI-0521250 for acquisition of the QTRAP LC-MS/MS. Thanks to Laurence Davin (Washington State University) for collection and identification of *Galanthus elwesii*. I thank Judy Coyle for assistance in figure preparation, Robin S. Kramer (National Center for Genome Resources) for contributing the transcriptome assemblies, Bradley S. Evans (Donald Danforth Plant Science Center) for assistance in mass spectral interpretation, John E. Cronan (Department of Microbiology, University of Illinois) for the *E. coli* culture with the *pfs* construct, and J. Steen Hoyer for critiquing this manuscript. Thank you to Toni M. Kutchan for allowing me to join her lab, her advice over the years, and continual feedback on my experiments and manuscripts. Thank you Joseph Jez for serving as the committee chair during the thesis proposal and subsequent committee updates. Thank you to Barbara Kunkel, Elizabeth Haswell, Jeffrey Henderson, Joseph Jez, and Todd Mockler for attending my committee meetings and providing useful feedback. I would like to thank the Kutchan lab members/visiting scientists Megan Augustin, Jörg Augustin, Linna Han, Rachel Wellinghoff, Ashutosh Shukla, Xiaohong Feng, Yasuhiro Higashi, Lydia Walter, Michael Müller, Dan Ruzicka, Luis Figueroa, Victoria Gowen, Emily Xu, Alan Lu, and Christina Rouse for their advice, assistance, and support over the years.

Matthew Benjamin Kilgore

Washington University in St. Louis

December 2015

Dedicated to my friends and family.

ABSTRACT OF THE DISSERTATION

The Identification of Alkaloid Pathway Genes from Non-Model Plant Species in the Amaryllidaceae

for Arts & Sciences Graduate Students

by

Matthew B. Kilgore

Doctor of Philosophy in Biology and Biomedical Sciences

Plant Biology

Washington University in St. Louis, 2015

Dr. Toni M. Kutchan, Chair

Secondary metabolites are often restricted in their distribution to different groups of organisms. For this reason, attempts to study these often useful and interesting products of metabolism require an ability to work in a diversity of non-model species. Methods for gene discovery with low investment and high efficiency are needed to effectively identify the biosynthetic genes in these diverse pathways. During this work, a workflow for efficiently identifying biosynthetic genes was developed and applied to Amaryllidaceae alkaloid biosynthesis. Genes discovered during this work include a norbelladine 4'-*O*-methyltransferase, a cytochrome P450 capable of phenol-phenol coupling 4'-*O*-methylnorbelladine to noroxomaritidine, and a short-chain dehydrogenase/reductase capable of forming norbelladine from tyramine and 3,4-dihydroxybenzaldehyde. These enzymatic discoveries support the future application of this workflow to other biosynthetic pathways and organisms.

Chapter 1: Introduction: The Amaryllidaceae Alkaloids: Biosynthesis and Methods for Enzyme Discovery

This chapter is in press as a review in Phytochemistry Reviews as:

The Amaryllidaceae alkaloids: biosynthesis and methods for enzyme discovery

Matthew B. Kilgore[1] and Toni M. Kutchan[1]

[1]Donald Danforth Plant Science Center, St. Louis, MO, USA

Amaryllidaceae alkaloids are an example of the vast diversity of secondary metabolites with great therapeutic promise. The identification of novel compounds in this group with over 300 known structures continues to be an area of active study. The recent assembly of a transcriptome for *Lycoris aurea* highlights the potential for discovery of Amaryllidaceae alkaloid biosynthetic genes with new technologies. Recent technical advances of interest to studies in non-model species include those in enzymology, next generation sequencing, nuclear magnetic resonance spectroscopy (NMR), and mass spectrometry (MS). In this thesis, improvements in sequencing and co-expression analysis introduced in this chapter are utilized to discover three Amaryllidaceae alkaloid biosynthetic genes. This is done using the hypothesis that genes involved in the biosynthesis of particular Amaryllidaceae alkaloids are co-regulated and this leads to correlations between biosynthetic gene expression and corresponding end product accumulation. At the end of this introduction, a summary of the individual chapters in the thesis is presented.

1.1 Amaryllidaceae Alkaloids

1.1.1 Amaryllidaceae Description with a *Narcissus* spp. Emphasis

The Amaryllidaceae are a family of bulbous plants with leaves and flowers growing from the top of a bulb and roots growing from the base of the bulb (Figure 1.1A, B, and C). During this work, experiments were conducted on members of the *Galanthus* and *Narcissus* genera of this family with the majority of the enzymology on *Narcissus* sp. *aff. pseudonarcissus* proteins. Most *Narcissus* spp. bloom in the spring followed by leaf, inflorescence, and root senescence. This leads to a summer of primordia development inside the bulb. During the winter, the roots grow to full size while leaf and inflorescence tissues grow slowly. The leaf and inflorescence tissues grow quickly in the spring (Rees, 1969). It can take 3-8 years for a *Narcissus* sp. plant to reach flowering age from seed. Starch storage is mediated by bulb scales and old leaf bases that make up the layers of the bulb. Over the years, bulb depth in the soil is maintained by unbranched contractile roots (Rees, 1969; Gordon, 2002). The presence of mycorrhizal interactions with *Narcissus* sp. roots have been observed (Chilvers and Daft, 1981). Wild *Narcissus* spp. reproduction is through the generation of bulblets or seed. Axillary buds inside the bulb develop into bulblets, with approximately one initiated per year (Rees, 1969). The ability to efficiently reproduce by seed is limited for many commercial cultivars because they are the result of crosses between species, which renders them semi-sterile. These cultivars are propagated by bulb mediated bulblet generation or tissue culture methods for bulblet generation from bulb fragments (Chen and Ziv, 2005).

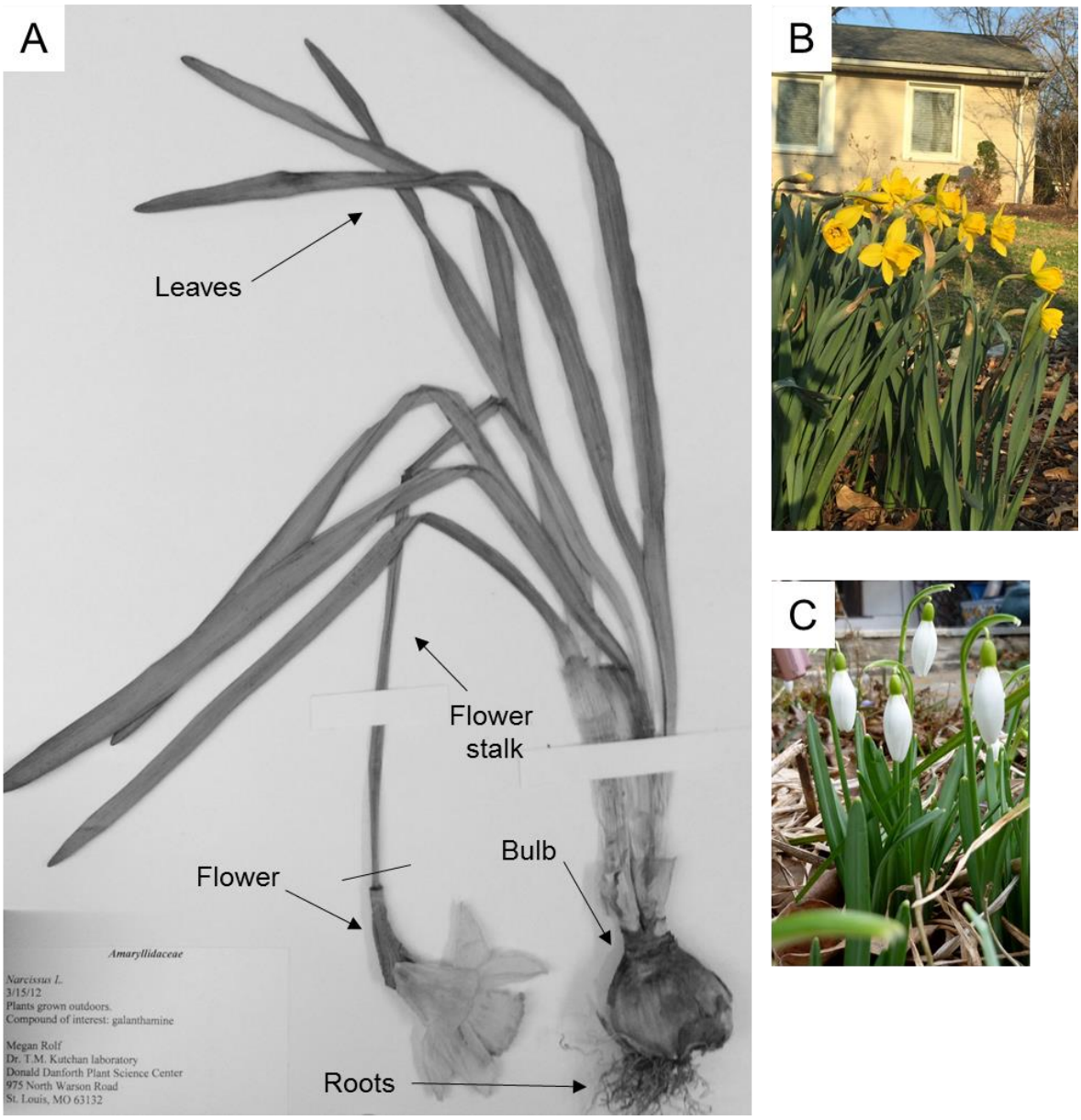

Figure 1.1 Photos of *Narcissus* sp. *aff. pseudonarcissus* and *Galanthus* sp.(A) An herbarium sample of *Narcissus* sp. *aff. pseudonarcissus* collected and pressed by Megan Augustin with major organs labeled. (B) A clump of *Narcissus* sp. *aff. pseudonarcissus* plants in flowering in late-march. (C) *Galanthus* sp. flowering in early-march.

1.1.2 Amaryllidaceae Alkaloid Introduction

The Amaryllidaceae have a long history as medicines in the cultures throughout their natural range spanning the Mediterranean, Central America, South America, Europe, Asia, and Africa (Plaitakis and Duvoisin, 1983; Louw et al., 2002; Cabezas et al., 2003; Howes and Houghton, 2003). The

Amaryllidaceae alkaloids are responsible for the medicinal properties of the Amaryllidaceae and are largely restricted to the family Amaryllidaceae, specifically the subfamily Amaryllidoideae (Chase et al., 2009). A noteworthy exception is the collection of alkaloids that have been found in the genus *Hosta* that is in the order Asparagales along with Amaryllidaceae (Chase et al., 2009; Li et al., 2012). New Amaryllidaceae alkaloid structures and the biosynthesis of these alkaloids have recently been reviewed (Kornienko and Evidente, 2008; Bastida et al., 2011; Jin, 2013; Takos and Rook, 2013). Galanthamine is a prime example of the Amaryllidaceae alkaloids. It is one of the three primary drugs used for the treatment of Alzheimer's disease. The Amaryllidaceae alkaloids and their derivatives have been a source of novel acetylcholine esterase inhibitors, anti-cancer compounds, anti-viral compounds, and antibacterial compounds with potential clinical applications. The last 10 years of progress in the identification of these bioactive compounds has recently been reviewed (He et al., 2015).

The *in planta* roles of these alkaloids are not as thoroughly examined as their potential medicinal roles. Some potential *in planta* roles of these compounds can be postulated based on their observed biological activities *in vivo* and *in vitro* on both mammals and other plant species. Several Amaryllidaceae alkaloids have been shown to have anti-cancer properties including the following examples haemanthamine, lycorine, narciclasine, lycoricidine, pancratistatin, β-crinane distichamine, and narciprimine (Liu et al., 2004; Havelek et al., 2014; He et al., 2015). These compounds have been highlighted as potential anticancer compounds because of their preferential cytotoxic action on cancer cell lines over normal cells, but there is usually back-ground toxicity for normal cells at higher doses. So, these cytotoxic compounds could be poisonous or illness inducing to an herbivore if consumed in adequate quantities; for this reason, it is possible these compounds act as herbivore deterrents. Galanthamine, sanguinine, ungiminorine, and 1-*O*-

acetyllycorine are acetylcholine esterase inhibitors (Irwin and Smith, 1960; He et al., 2015). Several insecticides including organophosphates and carbamates also act as acetylcholine esterase inhibitors and cause uncontrolled nerve firing that can lead to death (Casida, 1964). Hence, activity that makes these alkaloids potential Alzheimer's treatments my serve *in planta* as an herbivore deterrent acting on the nervous system. One of the side effects of galanthamine ingestion in humans is an upset stomach and diarrhea supporting its role as an herbivore deterrent (Hughes et al., 2004). Species of Amaryllidaceae frequently have a terrible taste and cause vomiting when ingested. As a result, poisoning from these species is rare. Oral administration of the Amaryllidaceae species *Boophone disticha* ethanol extracts to rats in doses of 240 mg/kg or higher results in death with symptoms reminiscent of acetylcholinesterase poisoning (tremors, convulsions, paralysis, and labored breathing) demonstrating the poisonous potential of Amaryllidaceae species (Gadaga et al., 2011). Some Amaryllidaceae alkaloids could have a similar role *in planta* as proposed for the clinic. Several Amaryllidaceae alkaloids are documented to have antiviral (lycorine, pancratistatin, hippeastrine, and haemanthamine), antibacterial (lycorine, ungeremine), and antiprotozoal (lycorine) properties which are as pertinent to plants as to humans when exposed to these classes of parasites (He et al., 2015). In addition to dealing with herbivores and parasites, Amaryllidaceae may use some of their compounds to modulate the growth of neighboring plants through allelopathy or perhaps modulate their own growth. The Amaryllidaceae alkaloid narciclasine is a prime example of a potential allelopathy (inhibitor of surrounding plant growth) agent with well documented inhibition of Arabidopsis seed germination and seedling growth. The effect is partly due to inhibition of polar auxin transport in roots and perhaps other tissues (Na et al., 2011).

There are still new alkaloids being discovered in the Amaryllidaceae alkaloids and even novel carbon skeletons with great potential to contribute to the list of known biologically active

compounds. A diversity of carbon skeletons are known for this group of alkaloids including hostasinine, belladine, galanthamine, crinine, lycorine, galanthindole, homolycorine, galasine, montanine, cripowellin, cherylline, buflavine, plicamine, tazettine, graciline, augustamine, pancratistatin, and gracilamine (Jin, 2009). Many of these compounds are of great potential pharmacological significance and their production through biological means is an area of great interest.

1.1.3 Prevalent Biosynthetic Gene Superfamilies

To further the biological production of these compounds, an understanding of their biosynthetic genes is requisite and the majority of the reactions are reaction types that are typically catalyzed by a collection of characterized enzyme families. The knowledge of these families can help inform efforts through homology searches to identify candidate genes. When studying secondary metabolism, in particular Amaryllidaceae alkaloid biosynthesis, several reaction types appear frequently including methylation, reduction, oxidation, condensation, hydroxylation, phenol-phenol' coupling, and oxide bridge formation. Examples of reductions found in the Amaryllidaceae include reduction of ketones, aldehydes, carbon-carbon double bonds, and imines. Two reductase superfamilies noted for their tendency to reduce aldehydes, ketones, carbon-carbon double bonds, and imines include aldo-keto reductases (AKRs) and short-chain dehydrogenase/reductases (SDRs) (Jörnvall et al., 1995; Penning, 2015). The SDR superfamily consists of three families including short-chain dehydrogenase/reductases, medium-chain dehydrogenase/reductases (MDRs) also known as alcohol dehydrogenases (ADH), and long-chain dehydrogenases/reductases (LDRs) (Kavanagh et al., 2008; Penning, 2015). The common feature of the SDR superfamily is a "Rossmann-fold" which is involved in the binding of dinucleotide cofactors including NADPH or NADH (Kavanagh et al., 2008). In Chapter 4, an SDR is identified

with the ability to reduce norcraugsodine, an early intermediate in Amaryllidaceae alkaloid biosynthesis. Oxidation reactions creating these various double bonds could be catalyzed by AKRs and SDRs as well because of the potential of these enzyme families to drive oxidations (Porté et al., 2013). 2-Oxoglutarate dependent dioxygenases and cytochrome P450 enzymes are well known for their ability to hydroxylate substrates thus making them good candidate gene families for the various hydroxylases in the biosynthesis of the Amaryllidaceae alkaloids (Lester et al., 1997; Nelson and Werck-Reichhart, 2011). The formation of an oxide bridge from a methoxy and hydroxyl group is probably catalyzed by a cytochrome P450 because CYP81Q1, CYP719A1, CYP719A13, and CYP719A14 are enzymes shown to catalyze this type of reaction (Ikezawa et al., 2003; Ono et al., 2006; Díaz Chávez et al., 2011). Phenol-phenol' coupling reactions are also likely catalyzed by cytochrome P450 enzymes. The cytochromes P450 CYP81Q1, CYP719A1, CYP719A13, and CYP719A14 have been shown to catalyze phenol-phenol coupling reactions and are proposed to act by a diradical mechanism (Ikezawa et al., 2003; Ono et al., 2006; Díaz Chávez et al., 2011). A phenol-phenol coupling cytochrome P450 forms a radical hydroxyl group on each of the phenol groups to be coupled. Resonance structures of these phenol radicals allow the single electron to be *ortho* or *para* to the initially radicalized hydroxyl. As a result, the electrons from the two phenol groups from a C-C bond with the possible phenol-phenol coupling possibilities of *ortho-para*, *para-para*, or *para-ortho*. In the case of the Amaryllidaceae alkaloids, a subsequent spontaneous nitrogen-ring closure (haemanthamine, crinine, and lycorine type) or oxygen-ring closure (galanthamine type) prevents the generation of a double phenol end product (Eichhorn et al., 1998). In other systems, however, with no further reactions or new quaternary carbons, both phenol groups are regenerated through rearrangement of hydrogens. In Chapter 3, CYP96T1 is shown to perform the *para-para'* phenol-phenol coupling reaction that makes the haemanthamine

and crinine carbon skeletons. Other enzyme groups noted for their ability to perform phenol-phenol' coupling reactions are laccases and peroxidases (Schlauer et al., 1998; Constantin et al., 2012). In the Amaryllidaceae two forms of methylation are common: *O*-methylation and *N*-methylation. *O*-Methyltransferases are divided into class I and class II methyltransferases (Ibdah et al., 2003). It is shown in Chapter 2 that the class I *O*-methyltransferase, N4OMT is responsible for the methylation of norbelladine to 4'-*O*-methylnorbelladine in Amaryllidaceae alkaloid biosynthesis (Kilgore et al., 2014). Other *O*-methylation reactions in the biosynthesis of these compounds could be catalyzed by homologues to *N4OMT* or homologues to other known *O*-methyltransferases including reticuline 7-*O*-methyltransferase, (*R,S*)-norcoclaurine 6-*O*-methyltransferase, columbamine *O*-methyltransferase, chavicol *O*-methyltransferase, and eugenol *O*-methyltransferase (Gang et al., 2002; Morishige et al., 2002; Ounaroon et al., 2003). Examples of *N*-methyltransferases that could share homology with *N*-methyltransferases involved in several Amaryllidaceae alkaloid biosynthetic pathways include coclaurine *N*-methyltransferase and caffeine synthase (Kato et al., 2000; Choi et al., 2002). Homologues of *O*-methyltransferases would be of potential interest when looking for an *N*-methyltransferase as well because of the close homology that exists between the *O*- and *N*- methyltransferases (Raman and Rathinasabapathi, 2003).

1.1.4 Core Biosynthetic Pathway

Intermediate Discovery

The core biosynthetic pathway of the Amaryllidaceae alkaloids consists of the reactions required to produce 3,4-dihydroxybenzaldehyde and tyramine, the condensation and reduction of these precursors to norbelladine, and the subsequent methylation of norbelladine to 4'-*O*-methylnorbelladine (Figure 1.2). Phenylalanine and tyrosine were shown to be precursors for

haemanthamine by incorporation of [3-^{14}C]phenylalanine and [3-^{14}C]tyramine into haemanthamine in *Nerine bowdenii* (Wildman et al., 1962b). Degradation experiments of haemanthamine generated from radiolabeled tyramine were used to demonstrate the placement of the labeled carbons on positions C11 and C12 in experiments with [2-^{14}C]tyrosine in *Sprekelia formosissima* and [1-^{14}C]tyrosine in *Narcissus* 'Twink' daffodil (Battersby et al., 1961a; Wildman et al., 1962a). [3-^{14}C]Tyramine has also been documented to incorporate into haemanthamine, haemanthidine, and 6-hydroxycrinamine in *Haemanthus natalensis* bulbs (Jeffs, 1962). Lycorine and norpluviine have been shown to incorporate [2-^{14}C]tyramine and [1-^{14}C]tyramine in *Narcissus* "Twink" (Battersby and Binks, 1960; Battersby et al., 1961b). [^{14}C]Phenylalanine and [^{3}H]3,4-dihydroxybenzaldehyde were both shown to be precursors to the aromatic half of haemanthamine and lycorine (Suhadolnik et al., 1962, 1963b). The pathway from phenylalanine to the intermediate 3,4-dihydroxybenzaldehyde was determined by feeding to *Narcissus pseudonarcissus* [3-^{14}C]*trans*-cinnamic acid, [3-^{14}C]4-hydroxycinnamic acid, [7-^{14}C]benzaldehyde, [7-^{14}C]4-hydroxybenzaldehyde, [^{3}H]3,4-dihydroxybenzaldehyde and [^{3}H]threo-DL-phenylserine and monitoring production of haemanthamine. The precursors [3-^{14}C]*trans*-cinnamic acid, [3-^{14}C]4-hydroxycinnamic acid, [^{3}H]3,4-dihydroxybenzaldehyde and [7-^{14}C]4-hydroxybenzaldehyde showed incorporation into haemanthamine. This led to the conclusion that the pathway for conversion of phenylalanine to 3,4-dihydroxybenzaldehyde is in the following sequence: phenylalanine, *trans*-cinnamic acid, 4-hydroxycinnamic acid, 3,4-dihydroxycinnamic acid or 4-hydroxybenzaldehyde, and 3,4-dihydroxybenzaldehyde (Suhadolnik et al., 1963a). 3,4-Dihydroxybenzaldehyde has been documented in *Hydnophytum formicarum* and other plants outside the Amaryllidaceae (Prachayasittikul et al., 2008). It is possible that the 3,4-dihydroxybenzaldehyde pathway is more phylogenetically spread than the latter steps or

convergent evolution of product formation has occurred. Carbon fourteen labeled norbelladine is incorporated into the alkaloids lycorine, crinamine, belladine, haemanthamine, and norpluviine (Battersby et al., 1961b; Battersby et al., 1961a; Wildman et al., 1962c). 4'-*O*-Methylnorbelladine is a precursor of all the primary alkaloid skeletons including crinine (crinine), haemanthamine (vittatine, 11-hydroxyvittatine), galanthamine (galanthamine, *N*-demethylgalanthamine, and *N*-demethylnarwedine), and lycorine (lycorine, norpluviine, and galanthine) (Kirby and Tiwari, 1966; Bruce and Kirby, 1968; Fuganti and Mazza, 1972b, a; Fuganti, 1973; Eichhorn et al., 1998). 4'-*O*-Methylnorbelladine has long been considered the direct substrate for creation of the *para-para'* and *ortho-para'* carbon skeletons. 4'-*O*-Methylnorbelladine has recently been established as the direct precursor of the *para-ortho'* skeleton as well (Eichhorn et al., 1998). This universal requirement in all phenol-phenol coupling branches for 4'-*O*-methylnorbelladine makes it the last common intermediate before a three way split in the Amaryllidaceae biosynthetic pathway. The three common divisions at 4'-*O*-methylnorbelladine are the *para-para'* coupling that leads to the crinine and vittatine enantiomeric series, the *ortho-para'* phenol coupling that is elaborated into the classic alkaloid lycorine, and the *para-ortho'* coupling that is elaborated into the most widely used Amaryllidaceae alkaloid galanthamine (Figure 1.3). Most other Amaryllidaceae alkaloid carbon skeletons are thought to be derivatives of these four skeletons. Examples include the pancratistatin and tazettine carbon skeletons derived from the haemanthamine skeleton and the homolycorine skeleton derived from the lycorine skeleton (Figure 1.4 and 1.5). The belladine-type alkaloids are thought to originate by the simple methylation of norbelladine, though the order of methylations is not determined. The cherylline skeleton is thought to originate from hydroxylation at the 11-position of the norbelladine skeleton and subsequent cyclization with the dioxygenated phenol group (Chan, 1973).

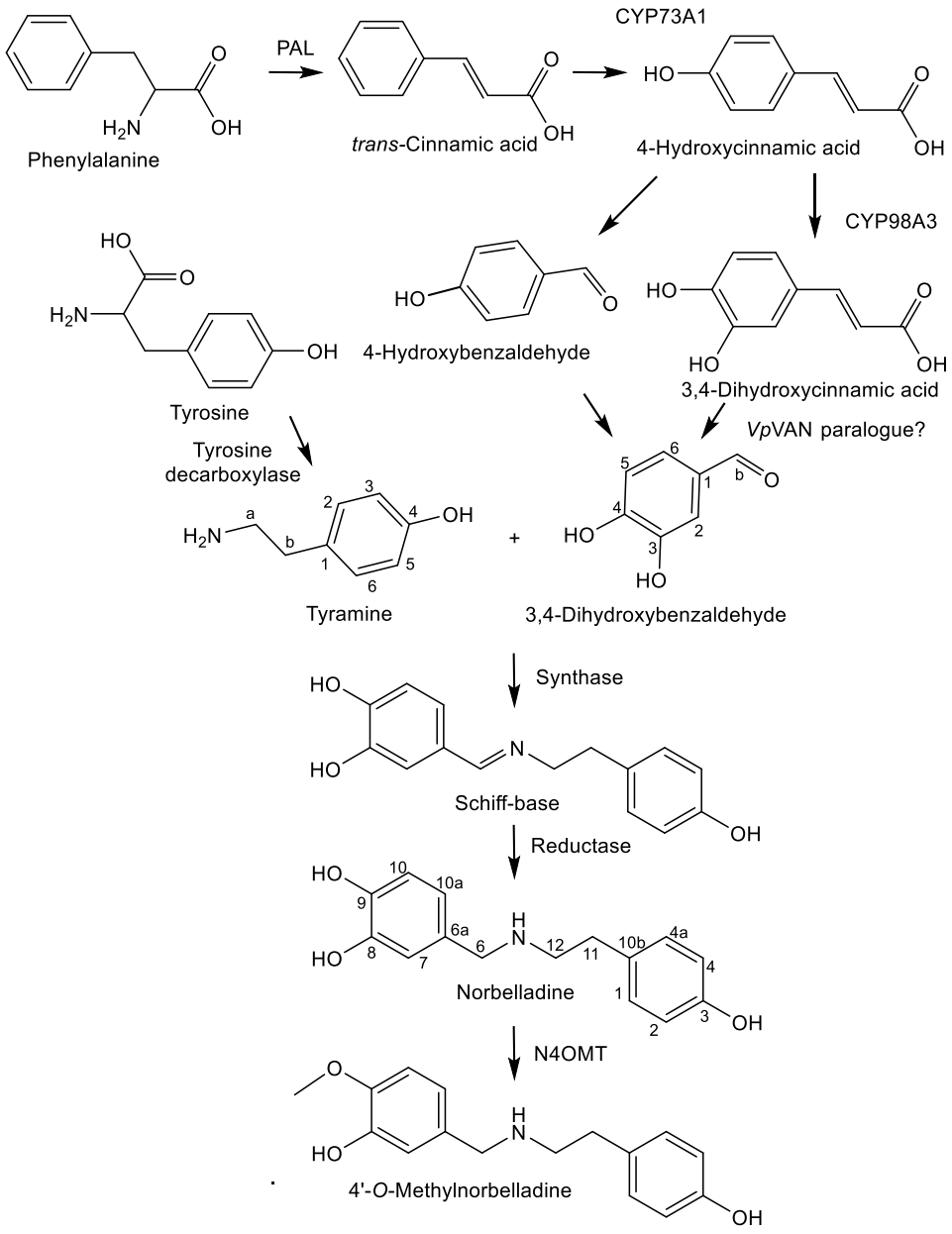

Figure 1.2. Core biosynthetic pathway for Amaryllidaceae alkaloids. Phenylalanine is converted to *trans*-cinnamic acid by phenylalanine ammonia lyase (PAL) and then to 4-hydroxycinnamic acid by CYP73A1. 4-hydroxycinnamic acid is potentially converted to 3,4-dihydroxycinnamic acid by CYP98A3 or to 4-hydroxybenzaldehyde and then to 3,4-dihydroxybenzaldehyde potentially by a *Vp*VAN paralogue. Tyrosine is converted to tyramine by tyrosine decarboxylase. 3,4-dihyroxybenzaldehyde and tyramine are condensed to form the Schiff-base, norcraugsodine, and reduced by an unknown reductase into norbelladine. Norbelladine is methylated by norbelladine 4'-*O*-methyltransferase (N4OMT) into 4'-*O*-methylnorbelladine.

Enzymology

The biosynthesis of 3,4-dihydroxybenzaldehyde from phenylalanine likely involves the early phenylpropanoid biosynthetic pathway all the way to caffeic acid (3,4-dihydroxycinnamic acid). Assuming the involvement of the phenylpropanoid pathway, 3,4-dihyroxycinnamic acid is a more likely intermediate in the biosynthesis than 4-hydroxybenzaldehyde. This is in agreement with the relatively low incorporation of 4-hydroxybenzaldehyde in radiolabeling experiments (Suhadolnik et al., 1963a). The deamination of phenylalanine to *trans*-cinnamic acid is done by phenylalanine ammonia lyase (PAL) (Tanaka et al., 1989). *Lr*PAL1 and *Lr*PAL2 have been cloned from the Amaryllidaceae plant *Lycoris radiata* demonstrating the presence of this enzyme in the Amaryllidaceae (Jiang et al., 2011; Jiang et al., 2013). The hydroxylation of *trans*-cinnamic acid to 4-hydroxycinnamic acid is done by cinnamate 4-hydroxylase (CYP73A1) (Fahrendorf and Dixon, 1993; Teutsch et al., 1993). CYP98A3 has been documented to hydroxylate the 3-position of 4-hydroxycinnamic acid (Franke et al., 2002). However, CYP98A3 prefers the shikimic acid or quinic acid esters over the free 4-hydroxycinnamic acid (Schoch et al., 2001; Franke et al., 2002). For this reason, it is possible a detour is required through shikimic acid, quinic acid, or acyl-CoA esters to get hydroxylated 3,4-dihydroxycinnamic acid. The conversion 3,4-dihydroxycinnamic acid to 3,4-dihydroxybenzaldehyde appears very similar to the conversion of ferulic acid to vanillin by vanillin synthase (*Vp*VAN), a hydratase/lyase (Gallage et al., 2014). The only difference is that the 3-hydroxyl is methylated in vanillin biosynthesis. Because of the substrate and reaction similarity, it is possible this reaction is catalyzed by an enzyme related to *Vp*VAN. Interestingly, there has been debate regarding the *Vp*VAN's preference for ferulic acid or 4-hydroxycinnamic acid (Havkin-Frenkel et al., 2003). If a similar enzyme in 3,4-dihydroxybenzaldehyde biosynthesis shares the ability to perform this reaction on substrates that

have or have not been hydroxylated at the 3-position, it would explain some of the ambiguity observed in earlier radiolabeling experiments. The conversion of tyrosine to tyramine is likely done by a homologue to the enzyme responsible for this reaction in other systems, tyrosine decarboxylase (Lehmann and Pollmann, 2009). This homologue in *Narcissus* sp. *aff. pseudonarcissus*, KT378599, has been cloned and confirmed to have tyrosine decarboxylase activity (Appendix A).

Prior to this thesis, the enzymology behind the condensation and subsequent reduction of 3,4-dihydroxybenzaldheyde and tyramine was unknown. The formation of the predicted Schiff-base intermediate, norcraugsodine, from tyramine and 3,4-dihydroxybenzaldehyde was possibly a spontaneous reaction occurring in solution, an enzymatically catalyzed condensation or both. The following reduction of the imine double bond to make norbelladine was potentially done by a reductase belonging to the AKR or SDR superfamilies. An SDR in the ADH family, tetrahydroalstonine synthase, from *Catharanthus roseus* can reduce the imine bond on strictosidine to form tetrahydroalstonine (Stavrinides et al., 2015). An enzyme, norcraugsodine reductase, capable of reducing norcraugsodine and potentially condensing 3,4-dihydroxybenzaldehyde and tyramine is discovered and described in Chapter 4. Several more NADPH-dependent imine reductases have been characterized in bacteria (Wetzl et al., 2015). After this reduction, norbelladine is shown in Chapter 2 to be methylated by the class I methyltransferase N4OMT in *Narcissus* sp. *aff. pseudonarcissus* (Kilgore et al., 2014). The three common phenol-phenol' coupling reactions that follow require the same biochemistry to operate and are likely done by cytochrome P450 enzymes, laccases, or peroxidases (Schlauer et al., 1998; Ikezawa et al., 2003; Ono et al., 2006; Díaz Chávez et al., 2011; Constantin et al., 2012).

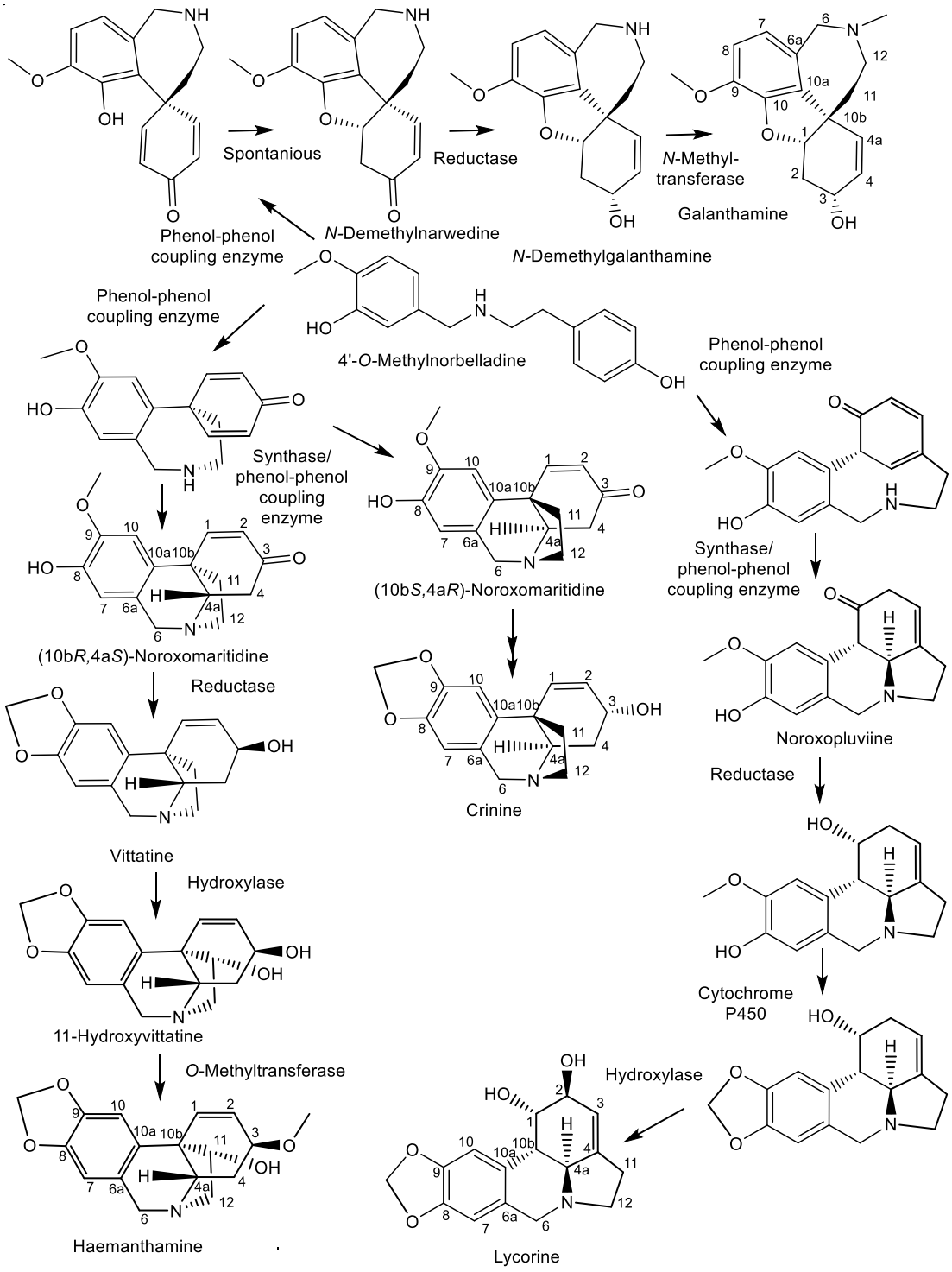

Figure 1.3. Primary Amaryllidaceae skeleton biosynthetic pathways with elaboration into the alkaloids haemanthamine, crinine, lycorine, and galanthamine.

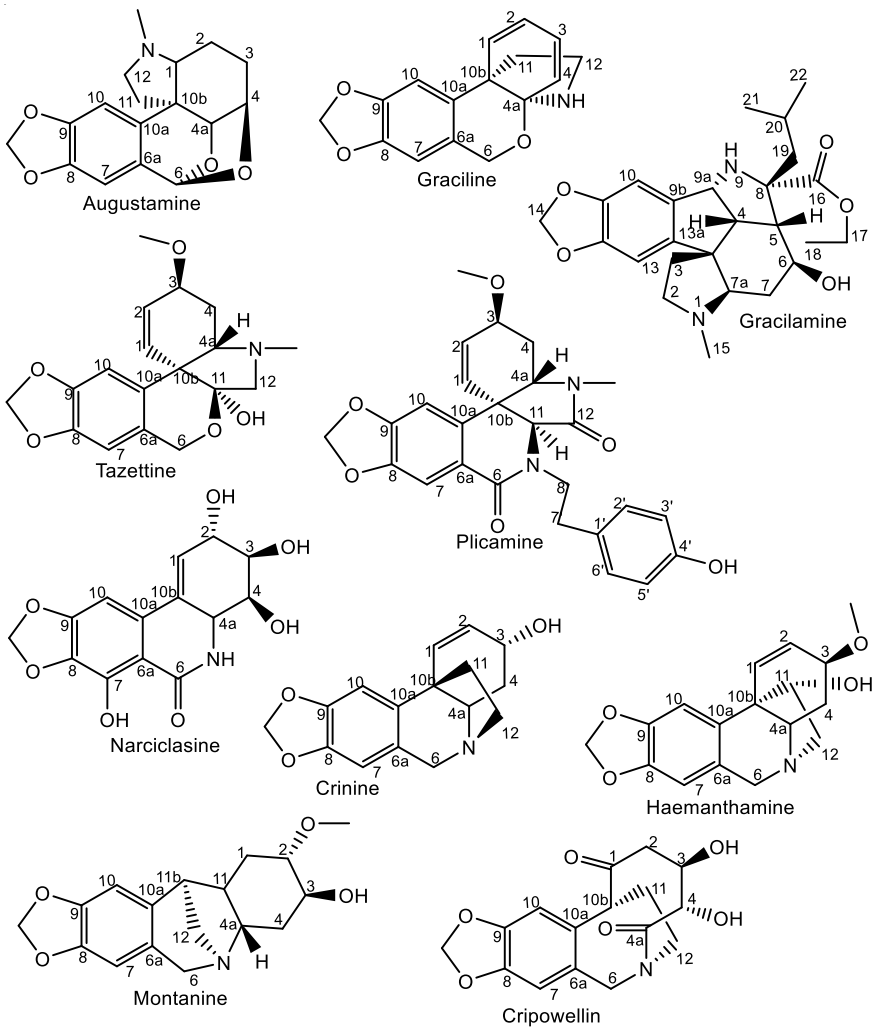

Figure 1.4. *Para-para'* phenol-phenol' coupling-derived skeletons found in the Amaryllidaceae.

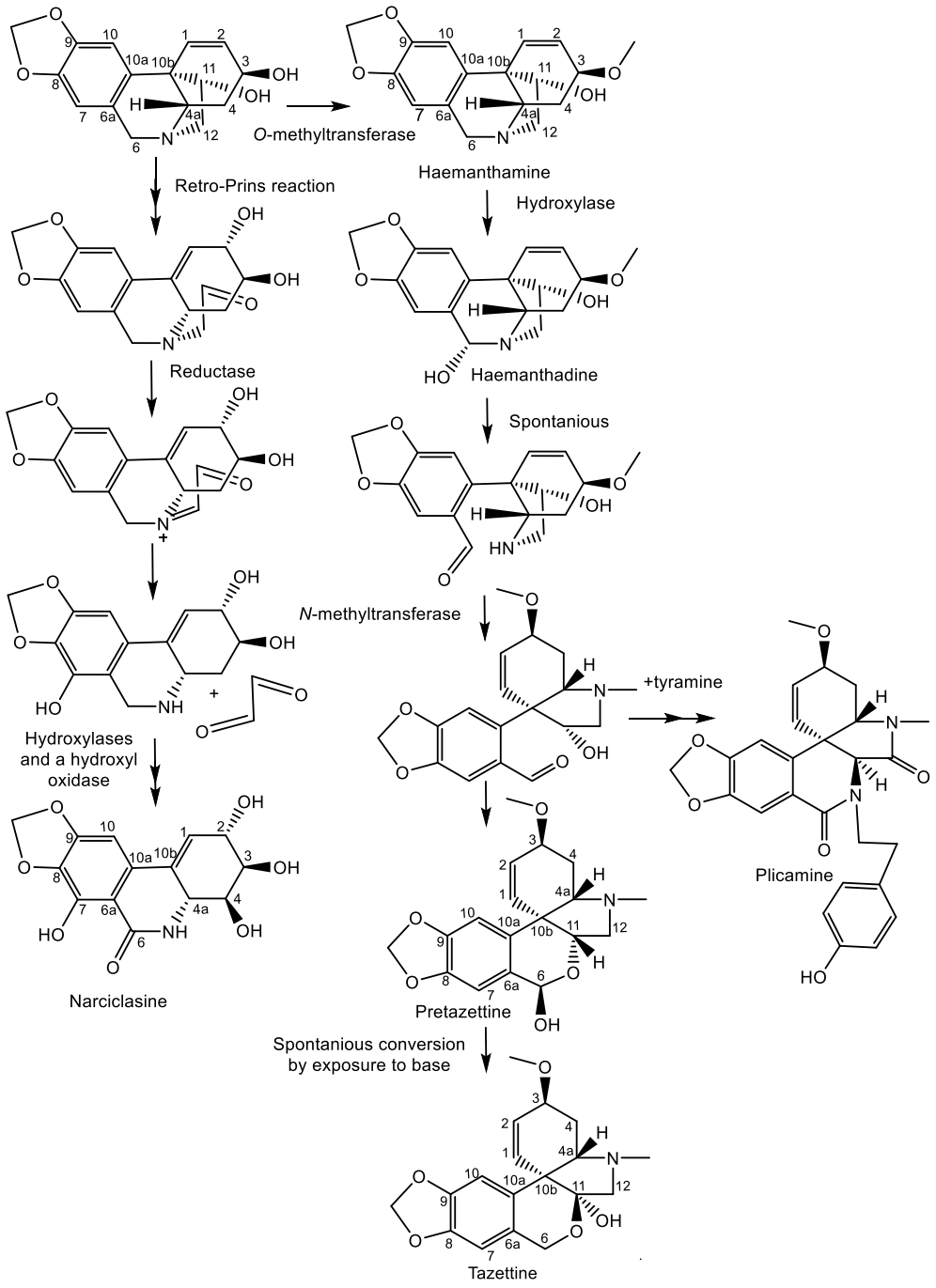

Figure 1.5. Carbon skeleton rearrangements in narciclasine, tazettine, and plicamine biosynthesis.

1.1.5 Galanthamine-Type Alkaloid Biosynthesis

Galanthamine is the representative alkaloid of the galanthamine type carbon skeleton. It is an Amaryllidaceae alkaloid used to treat the symptoms of Alzheimer's disease (Wilcock et al., 2000). Because of its importance, several studies have been done with radiolabeled isotopes to determine

its biosynthetic pathway (Barton et al., 1961; Barton and Kirby, 1962; Eichhorn et al., 1998). Initial studies by Barton et al. with 4'-*O*-methyl-[*N*-methyl-^{14}C]norbelladine, [*N*-methyl-^{14}C]norbelladine, and [4'-*O*-methyl-^{14}C]norbelladine as precursors of galanthamine showed no incorporation of [4'-*O*-methyl-^{14}C]norbelladine and a significant incorporation of 4'-*O*-methyl-[*N*-methyl-^{14}C]norbelladine and [*N*-methyl-^{14}C]norbelladine. This lead to the conclusion that *N*-methylation occurs at the norbelladine stage and subsequent *O*-methylation yields 4'-*O*-methyl-*N*-methylnorbelladine. This 4'-*O*-methyl-*N*-methylnorbelladine substrate was a precursor for phenol-phenol' coupling to narwedine that lead to galanthamine through a subsequent reduction (Barton and Kirby, 1962). This view was revised by subsequent work in *Leucojum aestivum* where it was shown that 4'-*O*-methylnorbelladine could be used as a substrate for phenol-phenol' coupling (Figure 1.3). The incorporation of 4'-*O*-methyl-*N*-methylnorbelladine was found to be one third the rate of incorporation of 4'-*O*-methylnorbelladine with an *N*-demethylation then occurring prior to conversion to galanthamine. *N*-Demethylgalanthamine was shown to be converted into galanthamine as well (Eichhorn et al., 1998). These two studies are an example of the limitations of conclusions drawn from negative results in radiolabeling experiments, especially considering the low incorporation rates observed in the initial study by Barton and Kirby (Barton and Kirby, 1962). The subsequent experiments by Eichhorn et al. revised the pathway establishing 4'-*O*-methylnorbelladine as the phenol-phenol' coupling substrate to make *N*-demethylnarwedine. *N*-Demethylnarwedine is reduced to *N*-demethylgalanthamine followed by *N*-methylation to make galanthamine (Eichhorn et al., 1998).

1.1.6 Haemanthamine, Pancratistatin, Gracilamine, Cripowellin, Tazettine, Plicamine, Hostasinine, Graciline, Augustamine, and Montanine-Type Alkaloid Biosynthesis

Haemanthamine and crinine are Amaryllidaceae alkaloids with well determined biosynthetic pathways and are the representative alkaloids of their respective skeleton types (Figure 1.3). The phenol-phenol' coupling reaction for the biosynthesis these skeletons proceeds with a *para-para'* coupling. A transient achiral intermediate is generated. Following the phenol coupling, the nitrogen attacks one of two carbon double bonds to generate each of the enantiomeric carbon skeletons. Given the absence of the crinine skeleton in many genera that contain the haemanthamine skeleton, an enzyme must be guiding this ring closure. In haemanthamine biosynthesis, (10b*R*,4a*S*)-noroxomaritidine results from phenol coupling which is then reduced to normaritidine. Normaritidine is oxidized to vittatine by the formation of an oxide bridge. The same enzymatic reactions are required for the conversion of the (10b*S*,4a*R*)-noroxomaritidine to crinine with some ambiguity surrounding the order of reduction and oxide bridge formation. If reduction is first, macowine is the intermediate, but if oxide bridge formation is first, oxocrinine is the intermediate (Fuganti and Mazza, 1972b). To make haemanthamine from vittatine requires hydroxylation and methylation. Hydroxylation to 11-hydroxyvittatine is considered the next step followed by methylation to haemanthamine as indicated by the incorporation of 11-hydroxyvittatine into haemanthidine which is a hydroxylated derivative of haemanthamine (Fales and Wildman, 1964; Fuganti, 1973). Since 11-hydroxyvittatine is not a catechol, this methylation is likely catalyzed by a homologue to a class II methyltransferase (Joshi and Chiang, 1998). A 2-oxoglutarate dioxygenase is found to hydroxylate vittatine into one of the diastereomeric forms of 11-hydroxyvittatine in Appendix B.

Radiolabeling studies with [2,4-$^{3}H_2$-4'-*O*-methyl-^{14}C]norbelladine and [1,4a-$^{3}H_2$-4'-*O*-methyl-^{14}C]norbelladine as precursors have been used to demonstrate the biosynthesis of narciclasine, a common pancratistatin-type alkaloid, through the *para-para'* phenol-coupling route (Figure 1.5). In corroboration, feeding of [^{3}H]norpluviine with no significant incorporation demonstrated the *ortho-para'* route to be an unlikely contributor to the biosynthesis of narciclasine (Fuganti et al., 1971). Narciclasine has been shown to be labeled from the racemic precursors [1,4a-$^{3}H_2$]noroxomaritidine, [1,3,4a-$^{3}H_3$]normaritidine, and [3-^{3}H]vittatine in *Narcissus* "Twink" and *Narcissus* "Texas" plants (Fuganti and Mazza, 1971). [2,4-$^{3}H_2$]Crinine and [2,4-$^{3}H_2$]vittatine were made by feeding [2, 4-$^{3}H_2$]4'-*O*-methylnorbelladine to species known for making these two enantiomers, *Nerine bowdenii* and *Pancratium maritimum*, respectively. When fed to *Narcissus* "Twink" and *Narcissus* "Texas", [2,4-$^{3}H_2$]crinine was not incorporated into narciclasine and [2,4-$^{3}H_2$]vittatine was incorporated. This shows the vittatine enantiomer to be the proper substrate and established the absolute configuration of narciclasine (Fuganti and Mazza, 1972b). [2,4-$^{3}H_2$]11-Hydorxyvittatine has been shown to incorporate into narciclasine and haemanthamine (Fuganti, 1973). 11-Hydroxyvittatine is the last theoretical common intermediate between haemanthamine and narciclasine. After formation of 11-hydroxyvittatine, the synthesis of narciclasine requires a series of reactions including a retro-Prins reaction, amine oxidation, 3 hydroxylations, and oxidation of a hydroxyl to a ketone. The retro-Prins reaction would result in the C2-hydroxyl, cleavage of the C10b-11 bond, and immigration of the C1-2 double bond to C10b-1. The amine oxidation to an imine between the nitrogen and C12, in addition to the retro-Prins reaction, would allow the release of the C11-12 carbon skeleton as glyoxal. The hydroxylations would occur at positions 4, 6, and 7. The hydroxyl at position 7 would subsequently be oxidized to a ketone. The Retro-Prins reaction has been implicated as a mechanism for germacradienol/germacrene D

synthase from *Streptomyces coelicolor* (Jiang et al., 2006). If the C2 bond is oxidized and then severed by hydrolysis, SDR and AKR enzymes are prime candidates due to their potential to perform oxidation reactions.

Gracilamine is an unusual dinitrogenous alkaloid skeleton first discovered in 2005 (Ünver and Kaya, 2005). It has been proposed to be a derivative of the *para-para'* phenol-phenol coupling reaction and leucine based on biomimetic schemes that have been devised for the total synthesis gracilamine from the *para-para'* skeleton and leucine (Tian et al., 2012).

The cripowellin skeleton was discovered in 1998 in *Crinum powellii* and resembles a highly oxidized version of the haemanthamine skeleton that has had the C10b-4a carbon-carbon bond severed and replaced with a ketone on the C4a position (Velten et al., 1998). If this is the pathway for generating this alkaloid, it is possible that carbon bond cleavage leading to the formation of a ketone is catalyzed by a cytochrome P450 similar to secologanin synthase that converts loganin into secologanin in an analogous manner with a ketone product (Irmler et al., 2000).

Tazettine biosynthesis proceeds from haemanthidine through a spontaneous ring opening between C6 and the nitrogen of haemanthidine forming a postulated ketone intermediate (Wildman and Bailey, 1969) (Figure 1.5). This is followed by the formation of a hemiacetal between the C6 ketone and the C11 hydroxyl and *N*-methylation (Fales and Wildman, 1964). The resulting pretazettine can spontaneously convert to tazettine on exposure to basic conditions (Wildman and Bailey, 1967).

Several other Amaryllidaceae skeletons are likely biosynthesized by a ring opening similar to the one in tazettine biosynthesis between C6 and the nitrogen, including plicamine, hostasinine, augustamine, and graciline. Plicamine is a dinitrogenous carbon skeleton thought to be derived

from a pathway similar to the tazettine skeleton with the addition of nitrogen from tyramine instead of the oxygen from C11 to the transient ketone intermediate as evidenced by structural similarities (Ünver et al., 1999) (Figure 1.5). Hostasinine is an alkaloid skeleton documented in *Hosta* and listed here because it is potentially derived from the Amaryllidaceae alkaloid haemanthidine. The proposed mechanism is depicted in Figure 1.6 and involves the spontaneous opening of the six membered nitrogen containing ring as in tazettine biosynthesis, reduction of the nitrogen C4a bond, formation of a carbon-carbon bond from C6 to C4, and oxidation of the nitrogen (Wang et al., 2007; Jin, 2009). Augustamine is potentially formed from the haemanthamine skeleton as shown by the presence of a direct connection of the five membered ring to carbon-10b as would be the case if the ether bond in the tazettine skeleton was at the C1 position and not the C11 positon (Machocho et al., 2004). The graciline skeleton is likely a derivative of the haemanthamine skeleton with the ether connected to C4a instead of C11 as in tazettine biosynthesis.

The biosynthesis of montanine proceeds through the intermediate vittatine (Feinstein and Wildman, 1976). One possibility is vittatine is hydroxylated to 11-hydroxyvittatine followed by a ring expansion triggered by a hydroxylation at C2 with subsequent methylation at C2 (Feinstein, 1967; Feinstein and Wildman, 1976).

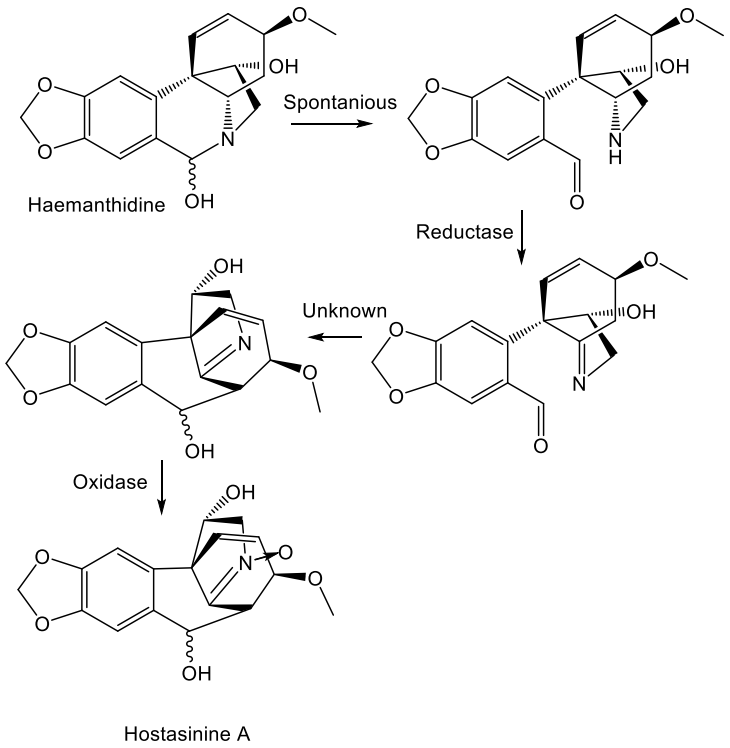

Figure 1.6. Possible mechanism of rearrangement for hostasinine-type alkaloids from haemanthidine. The steps are written in this order for illustration. There is no evidence regarding the order

1.1.7 Lycorine and Homolycorine-Type Alkaloid Biosynthesis

Lycorine was the first Amaryllidaceae alkaloid for which a structure was determined. The ubiquity of lycorine in the Amaryllidaceae made it ideal for radiolabeling experiments to elucidate its biosynthesis. Its early biosynthesis has been determined to be equivalent to that of haemanthamine and galanthamine (Battersby et al., 1964). At the phenol-phenol' coupling step, an *ortho-para'* reaction leads to the production of noroxopluviine. After the initial phenol-phenol' coupling, a re-aromatization of the phenolic rings or an enolization occurs as indicated by the loss of tritium at the C10b-position (Kirby and Tiwari, 1966). A stereospecific carbon nitrogen bond is also formed to create noroxopluviine. If rearomatization or keto-enol tautomerization occurs prior to the carbon nitrogen bond formation, the bond will form on an achiral or racemic intermediate, respectively. This indicates an enzyme is involved in the closure of the nitrogen ring. This chirality may be dictated by the arrangement of the intermediate in the active site of the phenol-phenol' coupling

enzyme or of another independent enzyme. Noroxopluviine is reduced to norpluviine (Battersby et al., 1964). As in haemanthamine biosynthesis, an oxide bridge is formed to make caranine (Battersby et al., 1964; Fuganti and Mazza, 1972a). This is followed by hydroxylation of the 2-position to make lycorine (Battersby et al., 1964; Wildman and Heimer, 1967). Studies in *Zephyranthes candida*, *Narcissus* "Twink" and *Narcissus* "Deanna Durbin" indicate an inversion of configuration of the C2 hydrogen at the hydroxylation step, but a study in *Clivia miniata* indicates the configuration is retained (Wildman and Heimer, 1967; Bruce and Kirby, 1968; Fuganti and Mazza, 1972a). Studies in *Clivia miniata*, *Narcissus* "Twink", and *Narcissus* "Deanna Durbin" agree that the protonation of C2 following phenol-phenol' coupling is stereospecific with addition of hydrogen to the alpha side of the molecule (Bruce and Kirby, 1968; Fuganti and Mazza, 1972a). The disagreement of the data surrounding C2 hydroxylation between these species could imply a difference in the mechanism of the enzyme conducting hydroxylation between these species (Figure 1.3).

Homolycorine biosynthesis is thought to start with the hydroxylation of norpluviine at C6 followed by a ring opening at the C6 nitrogen bond. This is followed by the formation of a hemiacetal connection between the C1 hydroxyl and the C6 ketone and methylations to form the known alkaloid lycorenine. The oxidation of lycorenine makes the representative alkaloid of the skeleton homolycorine. The exact order of the hydroxylation, ring breakage/hemiacetal formation and methylation reactions are poorly determined. The best evidence for order is the presence of lycorenine and homolycorine as products in feeding experiments for norpluviine. This indicates the formation of the homolycorine ester may be the last step in the pathway (Harken et al., 1976; Bastida et al., 2011). This reaction is done by an oxidase that could be an AKR, SDR, or a cytochrome P450 (Figure 1.7).

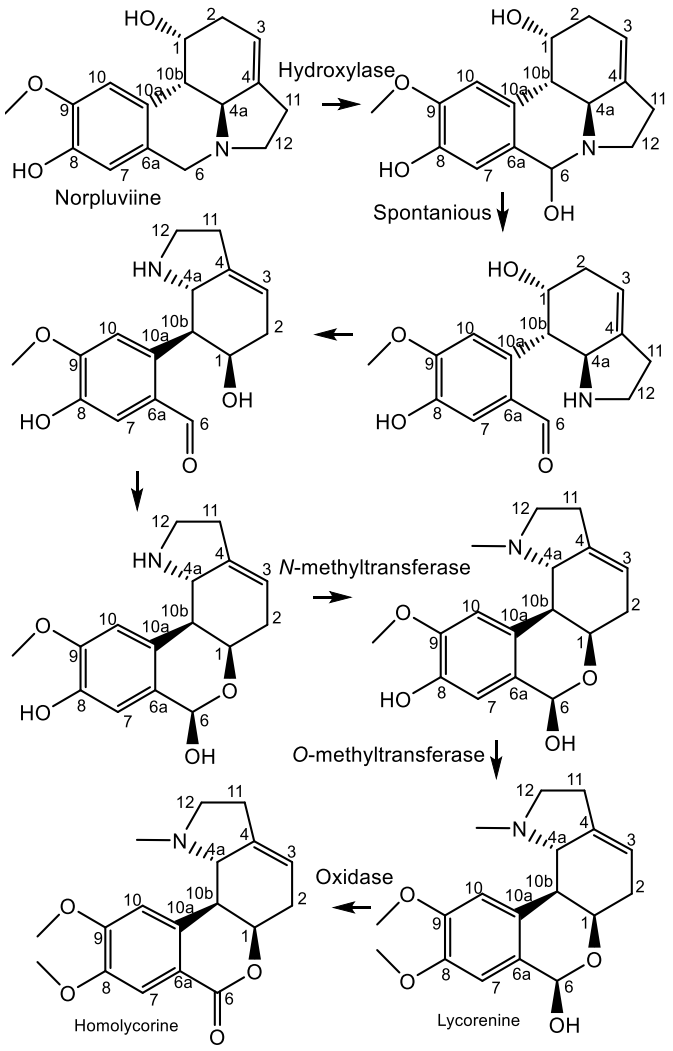

Figure 1.7. Formation of the homolycorine carbon skeleton from norpluviine. The order of reactions between norpluviine and homolycorine are not determined, but are diagramed in a sequence for illustration of the enzyme types that may be involved.

1.1.8 Other Amaryllidaceae Alkaloid Biosynthesis: Cherylline, Galanthindole, Galasine, and Buflavine

Cherylline-type alkaloids have been proposed to have a unique biosynthetic route that makes them and the belladine alkaloids the only Amaryllidaceae alkaloid skeletons to avoid the phenol-phenol coupling steps. This biosynthetic route involves a hydroxylation of the 2-position of 4'-*O*-methylnorbelladine or 4'-*O*-methyl-*N*-methylnorbelladine followed by a cyclization reaction by dehydration (Figure 1.8). An *N*-methylation is required to make cherylline, but its placement

before or after the hydroxylation and cyclization is unclear (Chan, 1973). Galanthindole type alkaloids are indole type alkaloids that were first discovered in 2003 (Unver et al., 2003). The lycorine or haemanthamine skeletons are potential candidates for their origin, but studies have not shown a biosynthetic relation to the Amaryllidaceae alkaloids through radiolabeling or other means (Unver et al., 2003). The galasine alkaloid type is rare and was first isolated in 1995 (Latvala et al., 1995). The galasine skeleton is perhaps derived from a similar biosynthetic route as homolycorine except the ether bond is migrated to position 10b. Buflavine-type alkaloids, including buflavine and apogalanthamine, are potentially of mixed origin. If their hydroxylation patterns are any indication, they may originate from different phenol-coupling reactions: haemanthamine/lycorine in the case of buflavine or galanthamine in the case of apogalanthamine, but having the same end carbon skeleton. There are other Amaryllidaceae alkaloids with diverging carbon skeletons that do not have their own classification due to their small number.

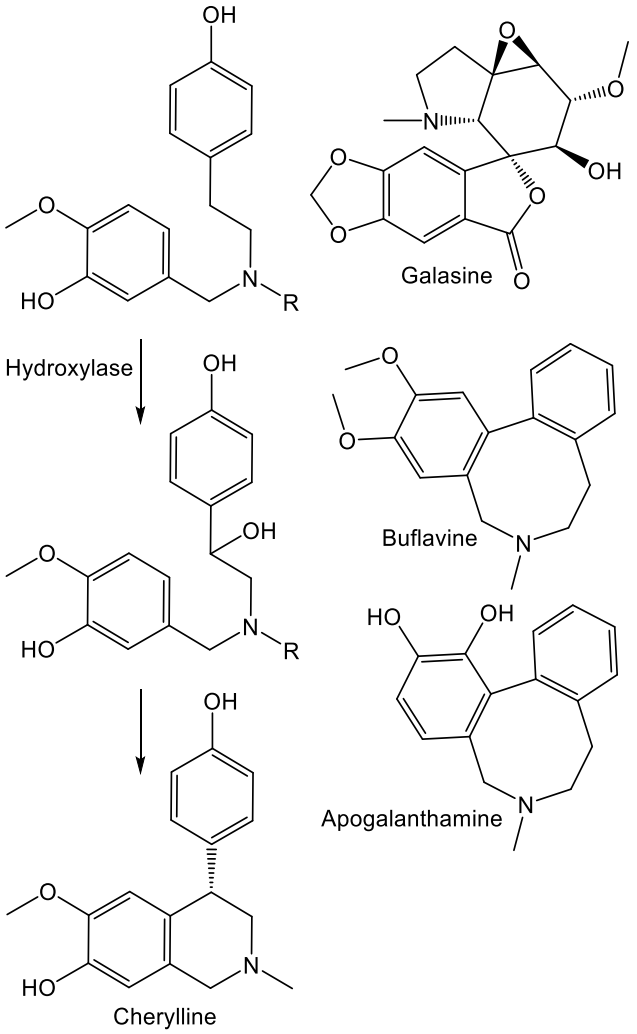

Figure 1.8. The pathway for cherylline and the structures of galasine, buflavine, and apogalanthamine. R indicates an undetermined methyl or hydrogen group.

1.2 Methods of Interest to Pathway Elucidation

1.2.1 Introduction

The themes of miniaturization and increased throughput in methods supporting secondary metabolism research promise to accelerate discovery of biosynthetic enzymes in these systems. These trends are particularly relevant because they enable studies in non-model systems with increased efficiency. How these methods relate to metabolomics has recently been reviewed (Sumner et al., 2015). In this section, advances in methods and theory of potential use to secondary

metabolism research in non-model species is examined including gene discovery, next generation sequencing technologies, NMR and MS.

1.2.2 Gene Clusters and Co-Regulation of Biosynthetic Pathways

Genes involved in the same biosynthetic pathway have recently been found positioned together in the same chromosome region in plant genomes forming gene clusters. Genes clusters have been observed in the secondary metabolism of *Zea mays*, *Avena* spp., *Oryza sativa*, *Arabidopsis thaliana*, *Lotus japonicus*, *Sorghum bicolor*, *Manihot esculenta*, *Papaver somniferum*, and *Solanum* spp. as reviewed recently (Boycheva et al., 2014; Chae et al., 2014). Current theory for gene cluster formation postulates that gene clusters form when a particular set of genes, or alleles of genes, are favored in one environment but disfavored in another and the alleles interact positively together or negatively apart for plant fitness (Takos and Rook, 2012). In secondary metabolism, biosynthetic intermediates are often chemically reactive and, thereby, toxic (Takos et al., 2011). The presence of the entire pathway generates a beneficial compound, but an incomplete pathway may lead to a loss in fitness. When looking at Amaryllidaceae alkaloid biosynthesis, several intermediates are catechols. Catechols can form deleterious reactive oxygen species, DNA adducts, protein adducts, or cause protein-protein cross-linking (Schweigert et al., 2001). Amaryllidaceae alkaloids are thought to function primarily as herbivore deterrents. Under low-nitrogen and/or low-herbivore pressure conditions, the expression of the pathway would have a fitness cost due to use of nitrogen. Corroborating this perspective, *Narcissus rupicola*, one of the only *Narcissus* spp. without Amaryllidaceae alkaloids, grows on rocky soil where nutrients such as nitrogen may be limiting (Berkov et al., 2014). Variability in the composition of alkaloids between *Galanthus elwesii* populations has been observed; this could indicate environment-specific benefits for selected alkaloids (Berkov et al., 2004). Considering that Amaryllidaceae

alkaloids are favorable in particular environments, but perhaps not in others, and that intermediates possess reactive functional groups, all criteria are met that favor the generation of gene clusters (Fisher, 1930; Takos and Rook, 2012). A sequenced genome can be used to discover new genes surrounding known biosynthetic genes and, assuming a gene cluster organization, these genes could be tested for involvement in the biosynthetic pathway (Itkin et al., 2013). Members of gene families with the same theoretical enzymatic mechanism for a proposed reaction are prime candidates for secondary metabolic pathways because evolutionary changes in substrate preference are more likely than changes in the underlying chemistry (Furnham et al., 2012).

Gene clustering information could be combined with co-expression analysis to filter and support candidate gene lists (Itkin et al., 2013). Co-expression is when transcripts express in a manner similar to each other; if one transcript is high the other will probably also be high and *vice versa*. These correlations in expression are frequently scored using correlation statistics including the Pearson correlation or Spearman's rank correlation (Kumari et al., 2012). The correlations can also exist between the transcripts of biosynthetic genes and end products of the biosynthetic pathway as is observed in proanthocyanidin biosynthesis and anthocyanin biosynthesis (Mellway et al., 2009; Ravaglia et al., 2013). The chromatin structural protein H2A.Z has been shown to positively correlate with gene expression in stretches of DNA enriched for this protein, and has been shown to be important for proper expression of these genes by analysis of gene expression in H2A.Z mutants in *Arabidopsis thaliana*. Chromatin modifications such as the content of H2A.Z are thought to contribute to co-expression of genes in gene clusters (Nützmann and Osbourn, 2015). This could be one of the mechanisms for the co-expression observed for many biosynthetic genes in secondary metabolic pathways. Another potential mechanism is a shared transcription factor as noted for the transcription factors *Os*TGAP1 from momilactone biosynthesis, MYB134

from poplar in proanthocyanidin biosynthesis, and bHLH3 from *Prunus persica* in anthocyanin biosynthesis (Mellway et al., 2009; Okada et al., 2009; Ravaglia et al., 2013). Transcription factors can act directly by driving the expression of biosynthetic genes or indirectly by driving expression of other transcription factors. Sharing a transcription factor causes genes to co-express because, barring additional regulatory factors, the transcription factor will influence the expression of its downstream genes in the same way under different conditions and developmental time points. Modification of expression from other transcription factors, transcription factor saturation, chromatin modifications, or DNA modifications, however, can cause deviations from a perfect correlation between genes regulated by the same transcription factor (Finnegan et al., 1996; Kaufmann et al., 2010).

One approach to elucidating biosynthetic pathways is to address correlation between metabolites or known biosynthetic genes and potential biosynthetic genes with regard to expression patterns, presence, absence, or pseudo gene status either within a species or between species. If, for example, genes that co-express with a known biosynthetic gene in multiple species can be found, it is more likely to be related to the function of this known biosynthetic gene than a gene co-expressing in only one species. However, when looking for a biosynthetic gene in this way, the possibility of non-homologous genes doing the same reaction should be considered and misannotation of the known gene should be guarded against. This is possible when considering the history of convergent evolution in secondary metabolism (Pichersky and Lewinsohn, 2011; Takos et al., 2011). This may be the case for the more derivatized Amaryllidaceae alkaloids, because there is not a clear correlation between phylogeny and the presence or absence of particular alkaloids (Rønsted et al., 2012). Due to the limitations of detection methods, however, the absence of these compounds in intervening lineages cannot be determined with certainty.

Examination of alkaloid composition in species closely related to the species in question may be helpful. If the biosynthetic pathway seems to be present in most common lineages as indicated by the presence of the end product, then there is no evidence for convergent evolution and common origin can be assumed in gene discovery workflows.

1.2.3 Sequencing Technologies

Advances in sequencing technologies show great promise to improve *de novo* genome and transcriptome assemblies in non-model systems. These improved datasets will facilitate identification of gene clusters, co-expression analysis, and cloning of candidate genes. Second generation sequencing has improved the efficiency with which genomic and transcriptomic sequence information can be obtained. It is of particular value when studying systems without previous sequencing information as is discussed in Chapter 2 (Kilgore et al., 2014). Early platforms for high-throughput sequencing are Roche 454 sequencing by Life sciences corporation; second generation MiSeq, NextSeq, HiSeq, and HiSeq X Illumina platforms; and SOLiD from life technologies. *De novo* genome assemblies can be made with second generation sequencing data using programs such as ALLPATHS, Velvet, ABySS, and SOAPdenovo (Zerbino and Birney, 2008; Simpson et al., 2009; Li et al., 2010; Gnerre et al., 2011). These genomes can provide information for the majority of the genome including genes, intergenic space, promoters, and introns. If the actual transcripts or the proteins they encode are the primary interest of the study, then *de novo* transcriptome assemblies will be more practical because of the reduced level of information required for sequence coverage and ability to focus only on transcripts that are being transcribed in samples of interest. *De novo* transcriptome assembly can provide a combination of sequence information, alternative splicing, and expression information in one experiment (Wang et al., 2009; Liu et al., 2014). Several prominent *de novo* transcriptome assemblers include Trinity,

Oases, and Trans-ABySS (Robertson et al., 2010; Schulz et al., 2012; Haas et al., 2013). Programs commonly used to align reads back to the transcriptomes and obtain expression estimates in the form of read counts include Bowtie and BWA (Li and Durbin, 2009; Langmead, 2010). This combination of information allows workflows for candidate gene selection based on homology and co-expression to be carried out with a very manageable initial investment and no prior sequence information (Giddings et al., 2011; Yeo et al., 2013; Kilgore et al., 2014). Workflows of this nature are used to identify biosynthetic genes in Chapters 2, 3, and 4. Many transcriptomes have been assembled using second generation sequencing, thereby providing information on a genetic level to previously uncharacterized systems such as *Camptotheca acuminata*, *Catharanthus roseus, Rauvolfia serpentine, Valeriana officinalis,* and *Veratrum californicum,* (Góngora-Castillo et al., 2012; Yeo et al., 2013; Augustin et al., 2015). As more organisms are sequenced, homology-based comparisons become more meaningful because these sequences can be used to prepare phylogenies. When these phylogenies are combined with biochemical validation data for proteins contained in the phylogenies, annotations can be made more accurate through programs like SIFTER (Engelhardt et al., 2009).

Second generation sequencing generates short reads that create fragmented genomes in non-model and model systems alike. The inability to differentiate reads from highly similar transcripts can make *de novo* transcriptome and genome assemblies prone to collapsing these similar sequences into one contig. Sophisticated analysis workflows have been developed to resolve this problem in second generation sequencing (Spannagl et al., 2013). In addition, third and fourth generation sequencing technologies that provide longer sequencing reads are a promising new tool. PacBio from pacific biosciences is an instrument that monitors the incorporation of fluorescently labeled bases into a DNA strand by a polymerase tethered to a pore

on a sequencing cell. The system is able to routinely generate sequences ~5 kb long with up to 50 kb possible. The down side is a ~80% error rate (Lee et al., 2014). Using circular libraries, the polymerase used for sequencing can read the same sequence multiple times and this data can be processed for error reduction. This results in a tradeoff between read length and accuracy determined by the number of times the polymerase can make a cycle around a loop. This system has been used for the sequencing and distinguishing of members of the highly similar vomeronasal receptor class 1 gene family in the non-model lemur *Microcebus murinus* (Larsen et al., 2014). One advantage of this system is its ability to simultaneously detect DNA modifications including the modified bases *N*6-methyladenosine, 5-methylcytosine, and 5-hydroxymethylcytosine. This information could be of value when using genomic DNA as a template and searching for regulatory modifications (Flusberg et al., 2010). A fourth generation system for DNA sequencing that commercializes nanopore technology is the MinION system that measures changes in electrical conductance as a DNA strand passes through a protein pore. This instrument is produced by Oxford Nanopore Technologies. One of the advantages of this technology is its size, measuring only 4 inches long and the ability to be powered by a standard USB 3.0 port making it the first highly portable sequencer. It has been shown to have an average read size of ~5 kb with reads reaching 10 kb. The low accuracy of MinION has improved from ~65% to ~85% since its first appearance, but still makes the technology impractical when used alone (Mikheyev and Tin, 2014; Loman and Watson, 2015). A combination of Illumina sequencing and MinION to make accurate Nanopore Synthetic-long reads prior to assembly has been used, however, to generate an *Acinetobacter baylyi* assembly with 99.99% accuracy (Madoui et al., 2015). A third approach to the generation of long reads is Illumina TruSeq which is a variant of second generation sequencing. This technique shears genomes into 10 kb sections and then performs short read sequencing and

assembly workflow on these 10 kb sections. This has been shown to be effective in the placement of the highly related transposable elements within the *Drosophila melanogaster* genome (McCoy et al., 2014). The longer sequence reads from PacBio, MinION, and Illumina TruSeq will help improve genome assemblies and connect contigs separated by highly repetitive regions. PacBio and MinION should also be able to provide start to end sequencing information of transcripts enabling splice variant analysis. The current downside to PacBio and MinION technologies is the high error rate associated with the raw read data. The Illumina TruSeq technology has low error rates ~0.03% because its reads are built with short reads. As for MinION, the high error rate can be corrected in PacBio with very high coverages with circular libraries or by combining the second generation short reads with third generation long reads as the programs LSC, proovread, and LoRDEC are designed to do (Au et al., 2012; Hackl et al., 2014; Salmela and Rivals, 2014).

1.2.4 Nuclear Magnetic Resonance Spectroscopy

NMR techniques have become more practical for the identification of the small quantities of compound observed in metabolomics surveys and generated during many enzyme assays. This will allow more complete catalogs of compounds in plant species to determine the presence or absence of compounds and their associated metabolic pathways. It will also facilitate identification of unknown products observed in enzyme assays. The usage of solid phase extraction (SPE) to concentrate metabolite fractions coming out of a separation technique like HPLC and the release of metabolite for NMR analysis has been of great utility. This workflow enables the structural elucidation of compounds from complex mixtures. When used in parallel with 2D NMR techniques for deconvolution of co-eluting compounds chromatography issues can be avoided (Mahrous and Farag, 2015). Another improvement of NMR technique is the invention of microprobes and miniaturized coils with current volumes of 10 µl and the potential of nanoliter

sample sizes (Fratila and Velders, 2011). The low volumes required allow for the elucidation of structures with microgram quantities of compound (Aramini et al., 2007). Without these innovations for deconvolution and miniaturization of NMR experiments, the workflows that utilize NMR in high-throughput systems with LC-MS outlined in the following section would not be practical.

1.2.5 Mass Spectrometry

MS has improved in mass accuracy with the development of fourier transform mass spectrometers (FTMS). The increased mass accuracy of FTMS like the Orbitrap can be of immense value in metabolomics by providing the accuracy needed to infer the molecular formula of compounds (Krauss et al., 2010). Combined with MS/MS data the compounds can be searched against databases for the inference of structure. This is of great value in metabolomics studies and in enzyme product identification when the product is a characterized structure for which standard is lacking. Several workflows apply MS^n and HPLC with SPE and NMR for the systematic structure elucidation of components in complex mixtures of metabolites (Castro et al., 2010; Sumner et al., 2015). To obtain quantification of compounds, evaporative light scattering detection (ELSD) can be used in combination with LC-MS (LC-ELSD-MS) (Cremin and Zeng, 2002). Another example of a combination workflow utilizes LC-ELSD-MS detection for initial characterization of complex plant extracts, followed by structural elucidation of select compounds by NMR. A significant effect in a high-throughput screen for a biological activity of interest, such as a greater than 32% reduction in the cancer cell lines MCF7, NCI-H460, or SF-26, is used to prioritize compounds for structural elucidation with NMR (Eldridge et al., 2002). These systematic catalogs of structures are very useful to secondary metabolism research because the associated MS^n and NMR data become available in databases for future analysis of enzyme products or identification of

metabolites from other complex mixtures in workflows lacking MS^n or NMR components. In GC-MS data analysis, the use of libraries such as just described is relatively simple because of the standard ionization settings and resulting reproducibility of spectra. In LC-MS^n the fragmentation of compounds can vary greatly depending on the instrument setup and settings (Hopley et al., 2008). To use existing LC-MS^n libraries the setup and setting used for the library construction should be used and this can be difficult if the instruments used in library construction and available to the researchers are not the same. To deal with this problem there are prediction algorithms that utilize LC-MS^n data with fragmentation rules for ion trees and algorithms for relating MS^n data to databases with different instrument setup and settings including the Mass Frontier software from Thermo Scientific.

1.2.6 Substrate Considerations

Advances in MS, and NMR promise to lower the quantity of product required in enzyme assays. This also lowers the quantity of substrate required which can also be a limiting factor when performing enzyme assays. When trying to acquire substrates for enzyme assays the substrates can be bought, synthesized, or isolated from the source. Purchasing the substrate is frequently not an option in highly specialized pathway intermediates, so the latter two lines of inquiry become necessary. In the case of Amaryllidaceae alkaloids, there is a large diversity of specialized synthesis methods that have regularly been reviewed by Zhong Jin and can be used for production of various pathway intermediates (Jin, 2013). As biosynthetic genes are discovered expression of these genes in a heterologous system for the production of pathway intermediates that can be used in enzyme assays becomes possible (Augustin et al., 2015). Methods for source isolation are provided in the publication of a compound's discovery but the availability of plant material can be a major constraint particularly for endangered species.

1.3 Conclusion

In conclusion, the Amaryllidaceae alkaloids are a diverse group of alkaloids with many undiscovered biosynthetic enzymes. Advances in sequencing will facilitate genomic and transcriptomic analyses of these plants to identify candidate biosynthetic genes. In this thesis transcriptomes for *Narcissus* sp. *aff. pseudonarcissus, Galanthus* sp., *and Galanthus elwesii* are generated and described in Chapters 2 and 3. Previous work has generated a transcriptome for *Lycoris aurea* (Wang et al., 2013). The combination of sequencing with other methods such as proteomics can be a powerful approach for the identification of candidate genes. This combination was implemented during the discovery of *Vp*VAN in *Vanilla planifolia* by selectively looking for transcripts and proteins highly present in the biosynthetic tissue for vanillin, inner bean pod (Gallage et al., 2014). Transcriptomic and genomic sequencing can also provide complementary information during candidate gene selection by allowing the combination of co-expression analysis and gene cluster searches. This combination of transcriptomic and genomic resources have become available in *Catharanthus roseus* through next generation sequencing (Kellner et al., 2015). Testing the candidate enzymes that will be identified through combinations of omics approaches including transcriptomics, proteomics, and genomics will be facilitated through more sensitive detection and structural elucidation of substrates and products by MS and NMR techniques either through plant-to-plant comparisons or direct assay of the candidate enzymes. The discovery of *N4OMT* in Chapter 2 is an example. A *de novo* transcriptome assembly of *Narcissus* sp. *aff. pseudonarcissus* provided both the sequence and expression information needed to identify a candidate methyltransferase. Once the methyltransferase is shown to make a methylated product from norbelladine with MS/MS, small volume NMR technologies are utilized to identify the product, 4'-*O*-methylnorbelladine (Kilgore et al., 2014). Another example of combining

transcriptomics with sensitive techniques such as small volume NMR is the discovery of the first 6 steps of cyclopamine biosynthesis from cholesterol up to the intermediate verazine. In this study a *de novo* transcriptome assembly of *Veratrum californicum* is used to find candidates through co-expression analysis. The candidates are expressed in insect cells and the various products are identified by a combination of MS/MS and NMR (Augustin et al., 2015). These examples of enzyme discovery in different species with different compound classes show these improvements are generally applicable to secondary metabolism in non-model systems when looking for novel enzymes.

1.4 Chapter Summary

1.4.1 Chapter 2

In this thesis, the Amaryllidaceae alkaloid biosynthetic genes are shown to co-express in leaf, bulb and inflorescence tissues with the accumulation of the Amaryllidaceae alkaloid galanthamine, and with each other. In Chapter 2, I show that *N4OMT* transcript co-accumulates with galanthamine in *Narcissus* sp. *aff pseudonarcissus.* This is done using a *de novo* assembled *Narcissus* sp. *aff pseudonarcissus* transcriptome with associated leaf, bulb, and inflorescence expression estimates to find methyltransferase gene transcripts that co-accumulate with galanthamine, and then testing these recombinant methyltransferases for enzyme activity with putative intermediates of galanthamine biosynthesis. N4OMT was shown to have the ability to methylate norbelladine to 4'-*O*-methylnorbelladine. The discovery of this enzyme/gene co-accumulating with galanthamine supports the starting hypothesis that Amaryllidaceae biosynthetic genes will co-express in a tissue–specific pattern that reflects galanthamine accumulation in the plant. This work was done in collaboration with Megan M. Augustin, Courtney M. Starks, Mark O'Neil-Johnson, Gregory D. May, John A. Crow, and Toni M. Kutchan.

Previously published in *PLoS One* 9(7):e103223

Cloning and characterization of a norbelladine 4'-*O*-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in *Narcissus* sp. *aff. pseudonarcissus*

1.4.2 Chapter 3

Chapter 3 demonstrates that co-expression is observed between Amaryllidaceae alkaloid biosynthetic genes and between species through the discovery of CYP96T1. CYP96T1 is a *para-para'* phenol-phenol coupling cytochrome P450 capable of forming noroxomaritidine. The gene was discovered by looking for cytochrome P450 transcripts that co-expressed with the *N4OMT* (discovered in Chapter 2) in *Narcissus* sp. *aff pseudonarcissus, Galanthus* sp., and *Galanthus elwesii*. When characterizing the recombinant enzyme, it was found to catalyze the formation of both enantiomers of noroxomaritidine. This indicates that CYP96T1 is involved in the initial *para-para'* phenol-phenol coupling to form an achiral intermediate and then releases this intermediate prior to a spontaneous nitrogen ring closure, thereby generating both enantiomers. This work was done in collaboration with Megan M. Augustin, Gregory D. May, John A. Crow, and Toni M. Kutchan.

1.4.3 Chapter 4

Chapter 4 demonstrates that the co-expression observed for *N4OMT* and CYP96T1 is shared with all of the core Amaryllidaceae alkaloid biosynthetic genes by identifying a norcraugsodine reductase. The discovery of this enzyme was through a similar work-flow as used in Chapter 3, but interrogating for oxidoreductases instead of cytochromes P450. The discovery of norcraugsodine reductase through co-expression analysis provides all the Amaryllidaceae alkaloid specific biosynthetic genes prior to the phenol-phenol coupling branch point. This enzyme was shown to have norcraugsodine reductase activity, and potentially facilitates the condensation of

tyramine and 3,4-dihydroxybenzaldehyde to form norcraugsodine. Cynthia K. Holland, Joseph M. Jez, and Toni M. Kutchan were key collaborators in this work.

1.4.4 Chapter 5

In Chapter 5, I present the major conclusions of this thesis. I address how this thesis will assist future research on Amaryllidaceae alkaloid biosynthesis, and on non-model species in general. Potential applications of the transcriptomic resources generated during my thesis research, and the enzymes discovered, are also discussed. The discovery of these three biosynthetic genes will facilitate the research on how variation in these genes and their expression leads to the observed variability in Amaryllidaceae alkaloid content across tissues, populations, and species. It will also facilitate discovery of additional biosynthetic genes for Amaryllidaceae alkaloids and regulatory elements, including transcription factors. These discoveries have the potential to be used for improved production of valuable Amaryllidaceae alkaloids and to provide experimental systems for clear examination of the roles these compounds play *in planta* and in the surrounding environment.

Chapter 2: Cloning and Characterization of a Norbelladine 4'-*O*-Methyltransferase Involved in the Biosynthesis of the Alzheimer's Drug Galanthamine in *Narcissus* sp. *aff. pseudonarcissus*

This chapter was previously published in *PLoS One* 9(7):e103223

Cloning and characterization of a norbelladine 4'-*O*-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in *Narcissus* sp. *aff. pseudonarcissus*

Matthew B. Kilgore[1], Megan M. Augustin[1], Courtney M. Starks[2], Mark O'Neil-Johnson[2], Gregory D. May[3], John A. Crow[3], Toni M. Kutchan[1†]

[1]Donald Danforth Plant Science Center, St. Louis, MO, USA

[2]Sequoia Sciences, St. Louis, MO, USA

[3]National Center for Genome Resources, Santa Fe, NM, USA

Galanthamine is an Amaryllidaceae alkaloid used to treat the symptoms of Alzheimer's disease. This compound is primarily isolated from daffodil *(Narcissus* spp.*),* snowdrop (*Galanthus* spp.), and summer snowflake (*Leucojum aestivum).* Despite its importance as a medicine, no genes involved in the biosynthetic pathway of galanthamine have been identified. This absence of genetic information on biosynthetic pathways is a limiting factor in the development of synthetic biology platforms for many important botanical medicines. The paucity of information is largely due to the limitations of traditional methods for finding biochemical pathway enzymes and genes in non-model organisms. A new bioinformatic approach using several recent technological improvements

was applied to search for genes in the proposed galanthamine biosynthetic pathway, first targeting methyltransferases due to strong signature amino acid sequences in the proteins. Using Illumina sequencing, a *de novo* transcriptome assembly was constructed for *Narcissus* sp. *aff. pseudonarcissus*. BLAST was used to identify sequences that contain signatures for plant *O*-methyltransferases in this transcriptome. The program HAYSTACK was then used to identify methyltransferases that fit a model for galanthamine biosynthesis in leaf, bulb, and inflorescence tissues. One candidate gene for the methylation of norbelladine to 4'-*O*-methylnorbelladine in the proposed galanthamine biosynthetic pathway was identified. This methyltransferase cDNA was expressed in *E. coli* and the protein purified by affinity chromatography. The resulting protein was found to be a norbelladine 4'-*O*-methyltransferase (*Np*N4OMT) of the proposed galanthamine biosynthetic pathway.

2.1 Introduction

Amaryllidaceae alkaloids are a group of alkaloids with many documented biological activities. This makes them valuable potential medicines several examples are the anti-cancer compounds haemanthamine and lycorine and the anti-viral compound pancratistatin (Gabrielsen et al., 1992; Liu et al., 2004; Havelek et al., 2014). One example of an Amaryllidaceae alkaloid already used medically to treat Alzheimer's disease is galanthamine. Galanthamine is an alkaloid discovered in 1953 that is produced by members of the Amaryllidaceae family (Uyeo and Kobayashi, 1953). It reduces the symptoms of Alzheimer's disease through acetylcholine esterase inhibition and nicotinic receptor binding. These activities are thought to compensate for reduced acetylcholine sensitivity in Alzheimer's disease by increasing acetylcholine levels and perhaps increasing acetylcholine sensitivity (Irwin and Smith, 1960; Barik et al., 2005). Until now, no committed biosynthetic genes have been identified (Wilcock et al., 2000; Wilcock et al., 2003). Limited

enzyme kinetic characterization has been done on plant protein extracts enriched for the norbelladine 4'-*O*-methyltransferase (N4OMT) of *Nerine bowdenii*, but the underlying gene was never identified (Mann, 1963).

The putative galanthamine biosynthetic pathway has been studied in detail and intermediates in the pathway have been determined. This knowledge is based on radiolabeling experiments. Work on other Amaryllidaceae alkaloids including lycorine and haemanthamine studying steps prior to 4'-*O*-methylnorbelladine can be applied to galanthamine biosynthesis because 4'-*O*-methylnorbelladine is a universal substrate for these alkaloids (Eichhorn et al., 1998). The proposed pathway starts with the amino acids phenylalanine and tyrosine (Barton et al., 1963). In *Narcissus incomparabilis* phenylalanine was established as a precursor that contributes the catechol portion of norbelladine. This was done using radiolabeling experiments to trace incorporation of [3-^{14}C]phenylalanine into lycorine and degradation experiments on the resulting lycorine to determine the location of the ^{14}C label (Suhadolnik et al., 1962). Similar experiments with phenylalanine were performed in *Nerine browdenii* monitoring haemanthamine incorporation (Wildman et al., 1962b). As a follow up radiolabeling experiments were used to determine that phenylalanine probably proceeds sequentially through the intermediates *trans*-cinnamic acid, *p*-hydroxycinnamic acid and 3,4-dihydroxycinnamic acid or *p*-hydroxybenzaldehyde before conversion into 3,4-dihydroxybenzaldehyde (Suhadolnik et al., 1963a). Tyrosine has been established as a precursor of galanthamine that in contrast to phenylalanine contributes only to the non-catechol half of the norbelladine intermediate. This was done by observing [2-^{14}C]tyrosine incorporation into galanthamine and degradation experiments of galanthamine (Barton et al., 1963). Tyrosine decarboxylase converts tyrosine into tyramine and is well characterized in other plant families (Lehmann and Pollmann, 2009). 3,4-

Dihydroxybenzaldehyde and tyramine condense into a Schiff-base and are reduced to form the first alkaloid in the proposed pathway, norbelladine. Norbelladine has been documented to incorporate into galanthamine and all major Amaryllidaceae alkaloid types in ^{14}C radiolabeling studies (Barton et al., 1961; Battersby et al., 1961b; Battersby et al., 1961a; Barton et al., 1963). 4'-*O*-Methylnorbelladine is then formed by *O*-methylation of norbelladine (Barton et al., 1963). A phenol-coupling reaction, followed by spontaneous oxide bridge formation, creates *N*-demethylnarwedine, which is then reduced and *N*-methylated to yield galanthamine (Figure 2.1) (Eichhorn et al., 1998). In one study, Barton et al. fed *O*-methyl[1-^{14}C]norbelladine to flower stalks of King Alfred daffodils, but it was not incorporated into galanthamine. The authors concluded that the intermediate in the pathway must be 4'-*O*-methyl-*N*-methylnorbelladine despite low incorporation of this compound when the equivalent experiment was conducted with 4'-*O*-methyl-[*N*-methyl-^{14}C]norbelladine (Barton et al., 1963). A recent revision of the proposed pathway by Eichhorn et al. contradicted this conclusion and placed the *N*-methylation step at the end of the proposed pathway instead of before the phenol-coupling reaction. In that study, [OC^3H_3]4'-*O*-methylnorbelladine was applied to ovary walls of *Leucojum aestivum.* Incorporation into products indicated that the pathway produced *N*-demethylated intermediates up until the penultimate step to galanthamine. *N*-methylation was proposed as the final step of biosynthesis (Eichhorn et al., 1998).

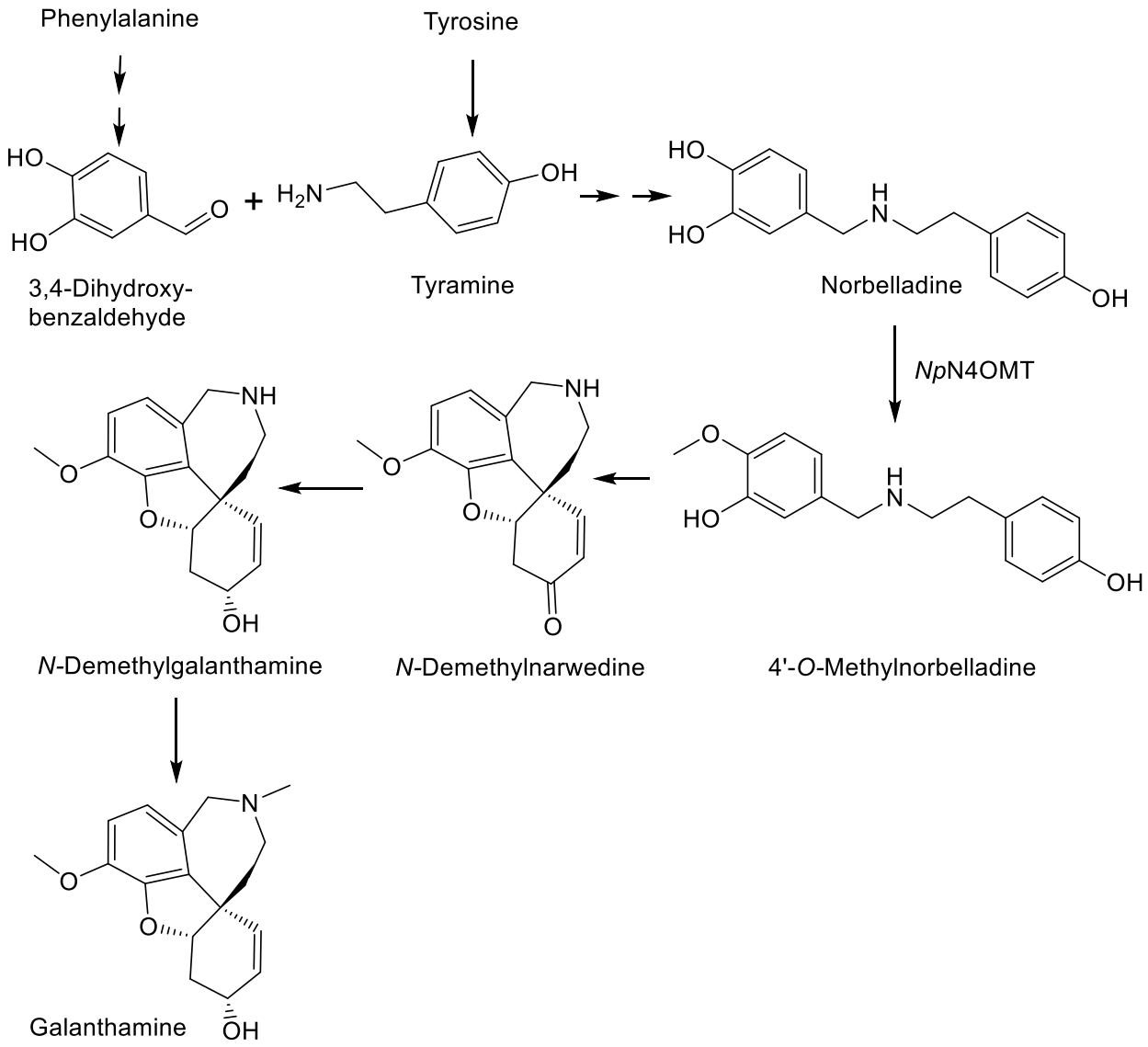

Figure 2.1. Proposed biosynthetic pathway for galanthamine. 3,4-Dihydroxybenzaldehyde derived from phenylalanine and tyramine derived from tyrosine are condensed to form norbelladine. Norbelladine is methylated by *Np*N4OMT to 4'-*O*-methylnorbelladine. 4'-*O*-Methylnorbelladine is oxidized to *N*-demethylnarwedine. *N*-demethylnarwedine is then reduced to *N*-demethylgalanthamine. In the last step, *N*-demethylgalanthamine is methylated to galanthamine.

The discovery of genes involved in metabolism is essential to metabolic engineering and synthetic biology. The elucidation of plant biochemical pathways can take decades. In fact, the biosynthesis of morphine, an important opiate analgesic, is still not completely elucidated at the gene level, even though the first enzyme specific to morphine biosynthesis was discovered more

than 20 years ago in 1993 (Gerardy and Zenk, 1993). Reports on the enzymatic activities of poppy extracts to describe the morphine biosynthetic pathway go even farther back to 1971 (Roberts, 1971). After more than 40 years of enzymology and reverse genetics, the morphine biosynthetic pathway is still incomplete at the gene level. Traditionally, plant biochemical pathway enzymes have been identified either directly by purification from plant extracts or indirectly by examining enriched cDNA libraries and functionally expressing clones (Mann, 1963; Takeshita et al., 1995; Frick and Kutchan, 1999; Morishige et al., 2002; Ounaroon et al., 2003; Raman and Rathinasabapathi, 2003; Nomura and Kutchan, 2010; Widiez et al., 2011). To reduce pathway discovery from a 20+ year process to a more reasonable time frame, new methods must be developed and embraced. The previous work on galanthamine biosynthesis makes the prediction of enzyme classes involved in the proposed pathway possible, thereby rendering the galanthamine pathway a suitable system for development of an omic methodology for biochemical pathway discovery.

In this study, using galanthamine biosynthesis as proof-of-concept, a novel workflow is presented to streamline the identification of biosynthetic pathway genes. A *de novo* transcriptome is created for *Narcissus* sp. *aff. pseudonarcissus* using illumina sequencing. HAYSTACK, a program that utilizes the Pearson correlation, is used to find genes that co-express with galanthamine accumulation in this transcriptome. This set of candidates is interrogated for homologs to methyltransferases. An OMT that converts norbelladine to 4'-*O*-methylnorbelladine (*Np*N4OMT) in the proposed biosynthesis of galanthamine is identified in this manner and characterized.

2.2 Materials and Methods

2.2.1 Plant Tissue and Chemicals

N. sp. *aff. pseudonarcissus* plants were collected from an outdoor plot in St. Louis, MO, with the GPS coordinates (38.659305, -90.410203), during peak flowering and separated into leaf, bulb, and inflorescence tissues. Inflorescence is considered all tissues above the spathe. The plants were collected with the permission of the corresponding author who is the owner of the private property. No endangered species were involved in this collection.

Formic acid, potassium phosphate monobasic, potassium phosphate dibasic, tris(hydroxymethyl)aminomethane, glycerol, sodium acetate, sodium chloride, tetramethylethylenediamine, calcium chloride, magnesium chloride, and β-mercaptoethanol were obtained from Acros Organics. Glycine, papaverine hydrochloride, *S*-adenosyl methionine (AdoMet), cobalt chloride, zinc chloride, and manganese chloride were obtained from Fisher Scientific. Other chemicals include acetonitrile, JT Baker; InstaPAGE, IBI Scientific; ethanol 200 proof, KOPTEC; Bradford reagent, Bio-Rad; *S*-adenosyl-L-homocysteine, Sigma-Aldrich; deoxynucleotide triphosphates (dNTPs), New England BioLabs, Inc. (NEB); and isopropyl β-D-1-thiogalactopyranoside (IPTG), Gold Biotechnology. The norbelladine *N*-methylnorbelladine, 4'-*O*-methyl-*N*-methylnorbelladine, and 4'-*O*-methylnorbelladine were synthesized previously (Eichhorn et al., 1998). NotI, NdeI, T4 DNA ligase, Taq DNA Polymerase, and Phusion High-Fidelity DNA Polymerase enzymes were from NEB. M-MLV reverse transcriptase and RNaseOUT were obtained from Invitrogen.

2.2.2 Alkaloid Extraction and Quantification

N. sp. *aff. pseudonarcissus* leaf, bulb, and inflorescence tissues were extracted by grinding tissue with a mortar and pestle cooled with liquid nitrogen. Each ground sample was split into three

technical replicates. Two volumes of 70 % ethanol were added followed by vortexing 5 min and centrifuging at 14,000 xg for 10 min. The supernatant was filtered through a 0.2 µm low protein binding hydrophilic LCR (PTFE, millex-LG) membrane. For galanthamine quantitation, samples were diluted 1000 fold. Liquid chromatography samples were injected (10 µl) onto an LC-20AD (Shimadzu) with a Waters Nova Pak C-18 (300 X 3.9 mm 4 µm) column coupled to a 4000 QTRAP (AB Sciex Instruments) for MS/MS analysis. The gradient program had a flow rate of 0.8 ml/min; solvent A was 0.1 % formic acid in H_2O and solvent B was 0.1 % formic acid in acetonitrile. At the beginning of the program, solvent B was held at 15 % for 2 min, followed by a linear gradient to 43 % B at 15 min, 90 % B at 15.1 min, 90 % B at 20 min, 15 % B at 21 min, and 15 % B at 26 min. A Turbo Ion Spray ionization source temperature of 500 °C was used with low resolution for Q1 and Q3. All multiple reaction monitoring (MRM) scans were performed in positive ion mode. The ion fragment used for quantitation of galanthamine was 288.00 $[M+H]^+$/213.00 $[M\text{-}OH\text{-}C_3H_7N]^{+\bullet}$ *m/z*. Galanthamine was identified by comparison of retention time and fragmentation pattern to authentic galanthamine standard. The Analyst 1.5 software was used to quantitate galanthamine using a comparison of peak area of the unknown to authentic galanthamine.

2.2.3 Illumina Sequencing and Transcriptome Assembly

The transcriptome was generated via data cleaning, short read assembly, final assembly, and post processing steps. A modified TRIzol RNA isolation method found as protocol number 13 in Johnson et al. was used to obtain RNA for cDNA library preparation (Johnson et al., 2012). Illumina RNA-Seq was used to generate 100 base pair paired-end reads from the cDNA library. The resulting data was monitored for overrepresented reads. Having found no such reads, adaptor sequences and sections of the standard phi X genome were identified and removed. Reads were

then trimmed for quality using the FASTX toolkit version 0.0.8 with a Q value cutoff of 10 as is default for PHRAP.

Reads were assembled in the following manner. ABySS was used to run multiple assemblies of the reads with a range of kmers $24 \leq k \leq 54$. The resulting assemblies were assembled into scaffolds using ABySS scaffolder (Birol et al., 2009). Gaps in the sequences were resolved using GapCloser from the SOAPdenovo suit (Luo et al., 2012). A final assembly was conducted on the resulting synthetic ESTs using Mira in EST assembly mode (Chevreux et al., 2004). All sequences with over 98 % identity were considered redundant and removed using CD-Hit (Li and Godzik, 2006). The resulting contigs >100 base pairs long were included in the final assembly. Protein products for these contigs were predicted using ESTScan (Iseli et al., 1999; Lottaz et al., 2003); all peptides over 30 amino acids were reported. Burrows-Wheeler Aligner was used to align the original reads to the assembled transcriptome to generate relative expression data for the contigs in leaf, bulb and inflorescence tissues (Li and Durbin, 2009). The *N.* sp. *aff. pseudonarcissus* assembly and the raw read data can be found at the MedPlant RNA Seq Database, http://www.medplantrnaseq.org. Anomalies in the number of reads per contig and abnormally long or short contigs were manually curated. To normalize for read depth, each expression value for each contig was divided by the total reads for the respective tissue and multiplied by 1 million.

2.2.4 Candidate Gene Identification

Relative expression data was compared to the levels of galanthamine in *N.* sp. *aff. pseudonarcissus* tissues using HAYSTACK with a background cutoff of 1, correlation cutoff 0.8, fold cutoff 4, and p-value 0.05 (Mockler et al., 2007). Using BLASTP, a list of known methyltransferases were queried against the *N.* sp. *aff. pseudonarcissus* transcriptome peptide list with an E-value of e^{-9} to identify methyltransferase homologs (Altschul et al., 1997). Accession numbers from NCBI for

these methyltransferases are presented in Table C.1. Overlap between the methyltransferase homologs and contigs that pass the HAYSTACK criteria were considered candidate genes. The candidate *N.* sp. *aff. pseudonarcissus* norbelladine 4'-OMT has the designation medp_9narc_20101112|62361 in the contigs.fa file in the Narcissus_spp.tar file on http://www.medplantrnaseq.org.

2.2.5 Phylogenetic Tree

Sequences found in Table C.2 were aligned using MUSCLE in the MEGA 5.2 software with default parameters (Tamura et al., 2011). For the phylogeny, this alignment was provided as input into the Maximum-Likelihood algorithm also found in MEGA 5.2. Default parameters were used except the Gaps/Missing Data treatment was set to partial deletion.

2.2.6 PCR and Cloning

The 5' and 3' ends of the *Np*N4OMT sequence were completed using Rapid Amplification of cDNA Ends (RACE) with the Invitrogen RACE kit. For gene specific primers (GSP) see Table C.3. The same PCR program was used for both 5' and 3'RACE. This applies to both cycles of nested PCR as well. The PCR program parameters were 30 s 98 ^{0}C 1 cycle; 10 s 98 ^{0}C, 30 s 60 ^{0}C, 1 min 72 ^{0}C 30 cycles; 5 min 72 ^{0}C 1 cycle. The outer-primer PCR was a mixture of 4.6 ng/µl RACE ready bulb cDNA, 0.3 mM dNTPs, 0.3 µM GSP primer, 0.9 µM kit provided RACE primer, 1 U NEB Phusion High-Fidelity DNA Polymerase, and Invitrogen recommended quantity of buffer in a 50 µl reaction. The inner-primer PCR used the product of the outer-primer PCR as template with 0.2 µM of the inner RACE GSP and Invitrogen primers and 0.2 mM dNTPs.

Amplification of the *Np*N4OMT open reading frame was performed with 5.1 ng/µl *N.* sp. *aff. pseudonarcissus* bulb oligo(dT) primed cDNA, 0.4 mM dNTPs, 0.4 µM each forward and reverse outer-primer, 1 UNEB Phusion High-Fidelity DNA Polymerase, and recommended buffer

in a 50 µl reaction. With the following PCR program parameters: 30 s 98 ^{0}C 1 cycle; 10 s 98 ^{0}C, 30 s 52 ^{0}C, 1 min 72 ^{0}C for 30 cycles; 5 min 72 ^{0}C 1 cycle. The inner-primer PCR used 1 µl of the outer-primer PCR product and used the inner-primers in Table C.3. The same PCR time program was used except the annealing temperature was increased to 53^{0}C.

*Np*N4OMT was cloned into the pET28a vector with the NotI and NdeI restriction sites that were added to the 5' and 3' ends of the open reading frame using the inner PCR primers. PCR product and pET28a were digested with NotI and NdeI enzymes, followed by gel purification and ligation with the T4 DNA ligase. The resulting construct was transformed into *E. coli* DH5α cells and screened on Luria-Bertani agar plates with 50 µg/ml kanamycin. Resulting colonies were screened by colony PCR with T7 sequencing and T7 terminator primers and Taq DNA Polymerase. The following cycle program was used: 3 min 94 ^{0}C 1 cycle; 30 s 94 ^{0}C, 30 s 52 ^{0}C, 2 min 72 ^{0}C 30 cycles; 7 min 72 ^{0}C 1 cycle. Plasmid minipreps were obtained using the QIAGEN QIAprep Spin Miniprep Kit. After Sanger sequencing of constructs (Genewiz), the desired plasmids were transformed into *E. coli* BL21(DE3) Codon Plus RIL competent cells. The sequences of the resulting five variants have the following accession numbers KJ584561(*Np*N4OMT1), KJ584562(*Np*N4OMT2), KJ584563(*Np*N4OMT3), KJ584564(*Np*N4OMT4), and KJ584565(*Np*N4OMT5).

2.2.7 Protein Purification

Recombinant protein production in 1 L of *E. coli* and purification with TALON resin followed the protocol found in (Higashi et al., 2010). No proteases were added to the protein extract, and desalting was performed with PD-10 columns from GE Healthcare. Protein quantity was determined according to Bradford; purity was monitored by SDS-PAGE. The *E. coli* cell line containing the hexahistidine-tagged methylthioadenosine/*S*-adenosylhomocysteine nucleosidase

(Pfs) construct from Choi-Rhee and Cronan's work was used to purify Pfs protein (Choi-Rhee and Cronan, 2005).

2.2.8 Screening Enzyme Assays

Enzyme assays for initial testing of *Np*N4OMT1 contained 10 µg of pure protein with 200 µM AdoMet, 100 µM norbelladine, and 30 mM potassium phosphate buffer pH 8.0 in 100 µl. The assays were incubated for 2 hr at 30^0C. The vector control was an *E. coli* extract purified with TALON in the same way as the methyltransferase protein. For the vector control assay, an equal volume of the pure vector control extract was substituted for the *Np*N4OMT1 protein in the enzyme assay. These assays were quenched by adjusting the pH to 9.5 with two volumes of sodium bicarbonate and extracted with two volumes ethyl acetate two times. After drying, the extracts were re-suspended in the initial mobile phase of the HPLC program. The HPLC separation of the assays was performed using a phenomenex Luna C8(2) 5 µm 250 x 4.6 mm column with solvent A (0.1 % formic acid in H_2O) and solvent B (acetonitrile). The program started with 10 % solvent B and a flow rate of 0.8 ml/min, a linear gradient began at 2 min to 30 % at 15 min, 90 % at 15.1 min, 90 % at 20 min, 10 % at 21 min, and 10 % at 28 min. Injection volume was 20 µl using a Waters Autosampler. Waters UV detector was set to 277 nm.

2.2.9 Kinetic Characterization

After optimization of the assay, the buffer was changed to 100 µM glycine at pH 8.8, with 5 mM of $MgCl_2$ added and the temperature was increased to 37 ^{0}C in 100 µl total reaction volume. When performing kinetic assays, the *E. coli* enzyme Pfs was added to break down SAH and prevent product inhibition. Product formation in kinetic experiments was quantified by comparing product peak area to a standard curve of the expected product or equivalent. Papaverine was used as an internal standard.

With the same solvent system as for screening enzyme assays, the HPLC program started with 20 % B and a flow rate of 0.8 ml/min, a linear gradient began at 2 min to 25.4 % B at 7 min, 90 % at 7.2 min, 90 % at 9 min, 20 % at 9.1 min, and 20 % at 14 min. A 4000 QTRAP mass spectrometer coupled to the same LC column and time program as used in HPLC was used to collect all compound mass and fragmentation data. For program setting details see Table C.4. For norbelladine kinetics an MRM program in positive ion mode was used to monitor the following fragments 260.0 $[M+H]^+$/138.0 $[M-C_8H_9O]^{+\bullet}$ *m/z*, 260.0 $[M+H]^+$/121.0 $[M-C_7H_8NO_2]^{+\bullet}$ *m/z*, 274.0 $[M+H]^+$/137.0 $[M+H-C_8H_9O_2]^+$ *m/z*, 274.0 $[M+H]^+$/122.0 $[M+H-C_8H_{10}NO_2]^+$ *m/z*. The fragments with 260.0 $[M+H]^+$ *m/z* and 274.0 $[M+H]^+$ *m/z* molecular ions were replaced when looking at *N*-methylnorbelladine for 274.0 $[M+H]^+$/152.1 $[M-C_8H_9O]^{+\bullet}$ *m/z*, 274.0 $[M+H]^+$/121.0 $[M-C_9H_{12}NO_2]^{+\bullet}$ *m/z*, 288.0 $[M+H]^+$/150.1 $[M-C_8H_9O_2]^{+\bullet}$ *m/z*, and 288.2 $[M+H]^+$/137.0 $[M-C_9H_{12}NO]^{+\bullet}$ *m/z*. Papaverine internal standard was monitored with the following fragments 340.4 $[M+H]^+$/324.2 $[M-CH_3]^{+\bullet}$ *m/z* and 340.4 $[M+H]^+$/202.1 $[M-C_8H_9O_2]^{+\bullet}$ *m/z*. When conducting dopamine kinetics, galanthamine was used as the internal standard and samples were not ethyl acetate extracted prior to LC/MS/MS analysis. To remove protein, two volumes of acetonitrile were added followed by 1 hr at -20 ^{0}C and 10 min centrifugation at 16,100 xg, 4 ^{0}C. The supernatant was dried under vacuum and re-suspended in the starting mobile phase before analysis. The HPLC time program was changed to start at 5 % solvent B with solution going to waste until 3.9 min, at 5 min start linear gradient to 25 % B at 25 min, 90 % B at 9.5 min, 90 % B at 11 min, 5 % B at 11.1 min, and 5 % B at 16 min. Ions monitored in the MRM were 168.0 $[M+H]^+$/151.0 $[M+H-OH]^+$ *m/z* and 168.0 $[M+H]^+$/119.0 $[M-OH-OCH_3]^{+\bullet}$ *m/z*. AdoMet steady state kinetic parameters were determined with norbelladine as the saturated substrate. Product was quantitated

using HPLC with the 28 min program used for screening enzyme assays. Product for assays on the additional *Np*N4OMT variants was detected with this same 28 min program on HPLC.

When conducting kinetic experiments the K_m was at least five fold higher than the minimum concentration of substrate and fivefold lower than the maximum concentration of substrate tested. K_m and k_{cat} were calculated by nonlinear regression to the Michaelis-Menten kinetics equation with the GraphPad PRISM 5.0 software.

2.2.10 NMR

NMR spectra were acquired for 4'-*O*-methylnorbelladine in CD_3OD at 600 MHz on a BrukerAvance 600 MHz spectrometer equipped with a BrukerBioSpin TCI 1.7 mm MicroCryoProbe. Proton, gCOSY, ROESY, gHSQC, and gHMBC spectra were acquired; ^{13}C chemical shifts were obtained from the HSQC and HMBC spectra. Chemical shifts are reported with respect to the residual non-deuterated MeOD signal (δ_H 3.31) (Figure C.1, C.2, C.3, C.4, and C.5).

2.2.11 Quantitative Real Time-PCR (qRT-PCR)

cDNA for leaf, bulb, and inflorescence tissues of *N.* sp. *aff. pseudonarcissus* were created using 1 µg RNA from the respective tissues, random primers, and M-MLV reverse transcriptase according to the Invitrogen protocol. qRT-PCR was conducted with a TaqMan designed gene expression assay for the methyltransferase with ribosomal RNA as a reference according to manufacture protocol. Reactions (5 µl) were performed in quadruplicate with outlier exclusion using Applied Biosystems StepOnePlus Real-Time PCR system. Methyltransferase relative expression values were determined by calculating $\Delta\Delta C_T$ values relative to standard ribosomal RNA and leaf tissue (Livak and Schmittgen, 2001).

2.3 Results

The Illumina sequencing of *N.* sp. *aff. pseudonarcissus* leaf, bulb, and inflorescence tissues resulted in 65 million paired-end reads that were used to make the *N.* sp. *aff. pseudonarcissus* transcriptome assembly. The transcriptome assembly consisted of 106,450 sequences (Figure 2.2A) with a mean length of 551 base pairs and a maximum length of 13,381 base pairs. A similar number of >100 base pair sequences were found in the transcriptome of *Chlorophytum borivilianum* (Kalra et al., 2013). This mean length indicates a high number of the sequences are long enough for homology searches and cloning work. Of these sequences, 79,980 were predicted to have open reading frames and were translated into peptides. After determining the reads coming from the three tissues, several homologs of genes with predictable expression patterns were used to evaluate the quality of the expression estimations. The RuBisCO large and small subunits have high amounts of expression in the photosynthetic leaf and inflorescence tissues compared to the non-photosynthetic bulb tissue. A homolog to the MADS62 floral development transcription factor is exclusively expressed in the inflorescence tissue as would be expected (Kwantes et al., 2012). The read counts were thus determined to produce expected expression patterns.

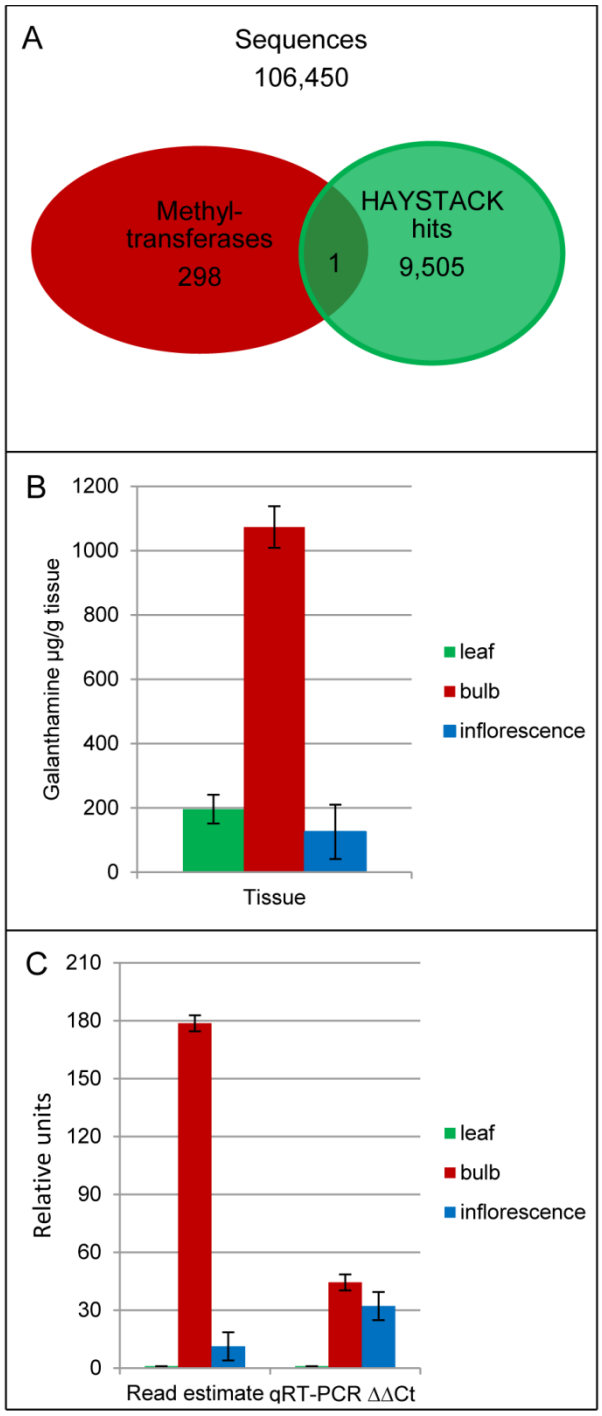

Figure 2.2. Identification of *Np*N4OMT in the *N.* sp. *aff. pseudonarcissus* transcriptome. (A) Venn diagram of all sequences, all OMTs, and all galanthamine correlating sequences according to HAYSTACK. (B) Accumulation level of galanthamine in *N.* sp. *aff. pseudonarcissus* (C) Candidate *Np*N4OMT expression profile in leaf, bulb, and inflorescence with the relative initial read estimate and qRT-PCR ΔΔCt on the y-axis with leaf tissue set to 1.

The LC/MS/MS data for leaf, bulb, and inflorescence tissues resulted in a pronounced accumulation pattern of galanthamine. The largest concentration was found in bulb tissue, with a lower level found in leaf and the lowest level in inflorescence (Figure 2.2B).

Using BLAST to seek homologs to the methyltransferases found in Table C.1 yielded 298 methyltransferase candidate genes (Raman and Rathinasabapathi, 2003). Separately, HAYSTACK identified 9,505 contigs that co-express with galanthamine accumulation. A comparison of the two resulting lists revealed one methyltransferase *Np*N4OMT that fits the HAYSTACK model (Figure 2.2A). This methyltransferase was chosen for functional analysis. After RACE, *Np*N4OMT was found to be a 239 amino acid protein with a predicted molecular weight (MW) of 27 kDa. When expressed using the pET28a vector, the added *N*-terminal Histidine tag increased the MW to 29 kDa (Figure 2.3A). In the course of cloning, five unique clones were obtained with >96 % identity to each other. Due to the two-toned yellow flower color, single flower, and size, the *N.* sp. *aff. pseudonarcissus* variety used in this study is likely Carlton. Based on genome size estimates, Carlton is thought to be a domesticated form of *Narcissus pseudonarcissus* with a genome duplication event that resulted in a tetraploid (Zonneveld, 2010). A high number of paralogs is, therefore, expected. In addition, these bulbs have been propagated vegetatively. For these reasons, the existence of multiple similar sequences is not surprising.

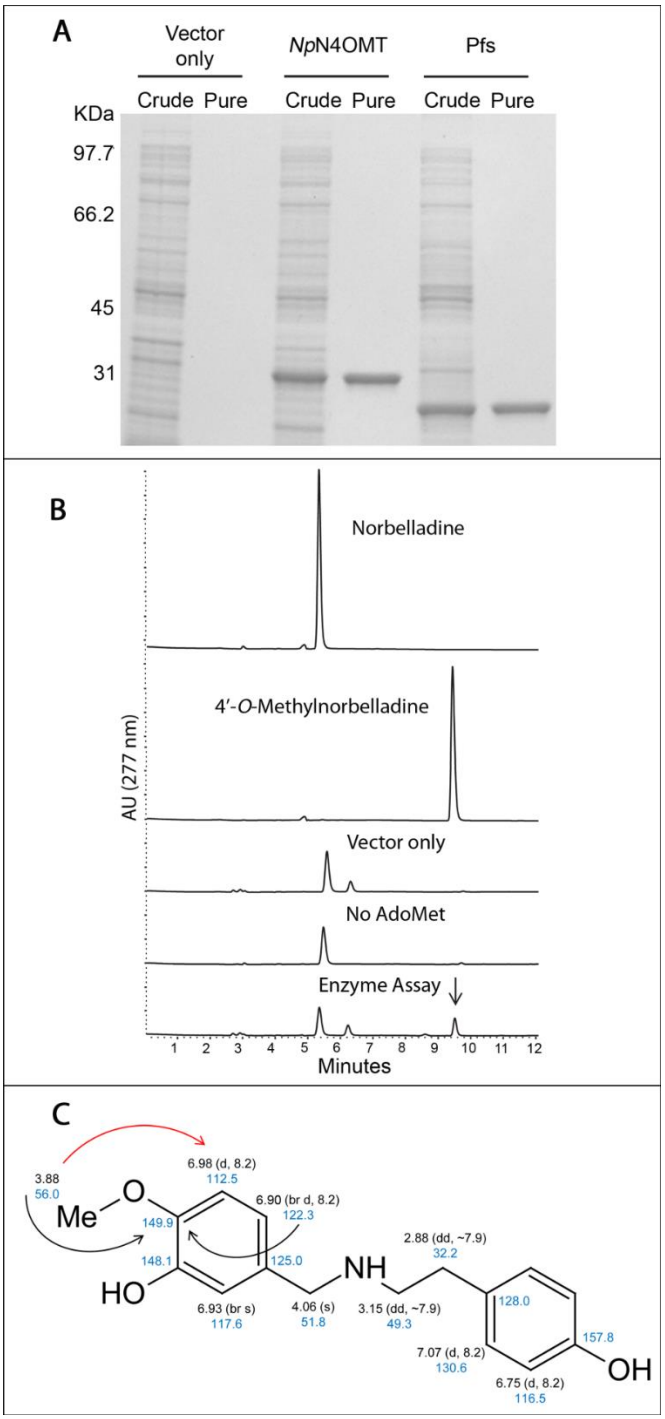

Figure 2.3. Recombinant *Np*N4OMT1 purification, enzyme assay, and NMR structure elucidation of the 4'-*O*-methylnorbelladine product. (A) SDS-PAGE gel 10 % including fractions from crude extract and the desalted protein prep. This is shown for vector only, *Np*N4OMT1, and Pfs preparations. (B) Enzyme assays (top to bottom): Norbelladine standard; 4'-*O*-Methylnorbelladine standard; Assay with *E. coli* vector-only crude extract added; Assay without AdoMet added; Complete methyltransferase assay. (C) NMR structure elucidation: proton chemical shifts are black, carbon chemical shifts are blue, key HMBC correlations are black arrows, and key ROESY correlations are red arrows.

Due to the high similarity of the clones, the first to be cloned was selected for thorough characterization. The clone selected for characterization is 92.5 % identical on the amino acid level to the original sequence in the transcriptome assembly (Figure 2.4). The recombinant protein was purified with a yield of 16.7 mg protein/L *E. coli* culture. SDS-PAGE analysis revealed the protein to be of apparent homogeneity (Figure 2.3A). Initial enzyme assays with *Np*N4OMT1 yielded, upon HPLC analysis, a peak with the retention time of 4'-*O*-methylnorbelladine. The vector only control lacks *Np*N4OMT1 but has all other assay components. Therefor the absence of product in the vector control assay excludes the possibility of a background reaction. The absence of product in the assay lacking AdoMet shows that the methyltransferase uses AdoMet as a co-substrate and cannot form product without AdoMet (Figure 2.3B). The pH optimum was found to be 8.8 and the temperature optimum 45 ^{0}C (Figure C.6B-C). The Pfs protein, shown purified in Figure 2.3A, was added to prevent SAH inhibition in kinetic enzyme assays, through the Pfs catalyzed hydrolysis of SAH to adenine and *S*-ribosyl-homocysteine (Choi-Rhee and Cronan, 2005).

Figure 2.4. Protein sequence alignment of *Np*N4OMT variants. Five unique variants of the *Np*N4OMT sequence are aligned against the original sequence predicted by the *de novo* assembled transcriptome using CLC software. Dots are identical residues.

An alternative methylation product, 3'-*O*-methylnorbelladine, has the same retention time on HPLC, the same UV profile, and MS/MS fragmentation pattern as 4'-*O*-methylnorbelladine. Thus, NMR analysis was performed to determine the regiospecificity of *O*-methylation. HMBC correlations from both the methoxyl protons (δ_H 3.88) and H-6' (δ_H 6.90) to the same carbon (δ_C 149.9) placed the methoxyl group at C-4'. Its location was further supported by a ROESY correlation from the methoxyl protons to H-5' (δ_H 6.98). The NMR data thus confirmed that 4'-*O*-methylnorbelladine is the product of the enzyme reaction (Figure 2.3C).

To determine the substrate specificity of this methyltransferase, several similar substrates were tested. Activity comparable to that found with norbelladine was observed using *N*-methylnorbelladine as the substrate. Dopamine also served as a substrate. Under the assay conditions used, product was not detected with caffeic acid, vanillin, 3,4-dihydroxybenzaldehyde, and tyramine as substrates (Table 2.1). To determine if the other 4 variants show similar activity, they were purified, and enzymatic activity was confirmed for all variants using norbelladine as the

substrate. When monitoring *Np*N4OMT1 norbelladine assays allowed to proceed to completion, no sign of double methylation products were observed as expected.

Table 2.1. Substrate specificity of *Np*N4OMT1

Substrate	K_m (μM)	k_{cat} (1/min)	k_{cat}/K_m (1/μM*min)
Norbelladine	1.6±0.3	1.3±0.06	0.8
AdoMet	28.5±1.6	4.5±0.01	0.16
N-Methylnorbelladine	1.9±0.4	2.6±0.15	1.3
Dopamine	7.3±2.7	3.6±0.15	0.5
Caffeic acid	ND	ND	ND
Vanillin	ND	ND	ND
3,4-Dihydroxybenzaldehyde	ND	ND	ND
Tyramine	ND	ND	ND

ND, Not detected

±, Standard error

Phylogenetic analysis of the *Np*N4OMT1 placed it in the class I OMT group (Figure 2.5). *Np*N4OMT1 has a length consistent with the 231-248 amino acid range found in class I OMTs. This is in contrast to other known plant catechol 4-OMTs which all group in the class II OMTs as their length and cofactor requirements reported in previous work would predict. All these methyltransferases are significantly longer than the standard class I OMTs and none are reported to have the characteristic divalent cation dependence of class I OMTs (Joshi and Chiang, 1998; Gang et al., 2002; Akashi et al., 2003; Schröder et al., 2004; Kim et al., 2005). When testing *Np*N4OMT1 for cation dependence, enzymatic activity improved upon the addition of cobalt. Enzymatic activity increased four-fold more with the addition of magnesium instead of cobalt (Figure C.6A). This preference for magnesium over other divalent cations is also to be expected from a class I OMT (Joshi and Chiang, 1998). It is, furthermore, consistent with previous work on enzyme extracts enriched for this OMT (Mann, 1963).

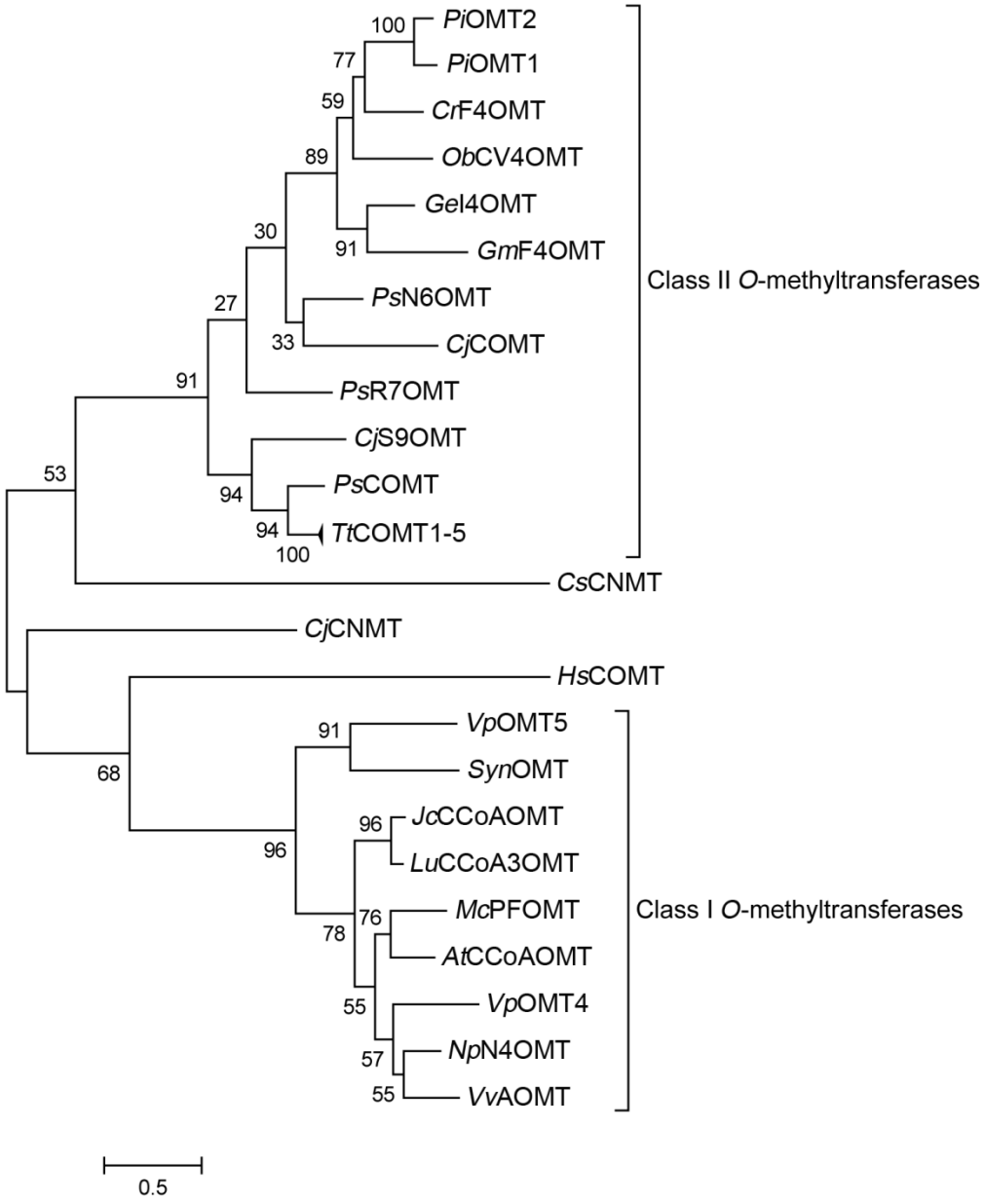

Figure 2.5. Phylogenetic analysis of *Np*N4OMT1. A maximum-likelihood phylogenetic tree of characterized methyltransferases listed in Table C.2. Alignment constructed using MUSCLE.

To validate the expression profiles predicted based on read counts; qRT-PCR was conducted with the same RNA preparation used to prepare the cDNA libraries for Illumina sequencing. The resulting expression profile is slightly different from that obtained from Illumina sequencing. The qRT-PCR expression profile has a higher quantity of inflorescence transcript

relative to bulb transcript (Figure 2.2C). This minor difference is potentially due to cross amplification, during qRT-PCR, with other close homologs in the plant.

2.4 Discussion

The expression pattern, product formation, and low K_m for norbelladine all indicate *Np*N4OMT methylates norbelladine in the proposed galanthamine biosynthetic pathway. Two differing orders of methylation have been proposed for galanthamine biosynthesis (Eichhorn et al., 1998). The methylation of *N*-methylnorbelladine was tested to determine if a preference for the *N*-methylation state could be observed at *O*-methylation. Similar K_m and k_{cat} values for *N*-methylnorbelladine and norbelladine indicate that a preference for the *N*-methylation state does not occur at *O*-methylation. The results presented here support both proposed galanthamine biosynthetic pathways. Future work on additional enzymes in the pathway will be needed to enzymatically validate one pathway or the other. The lack of enzymatic activity when testing 3,4-dihydroxybenzaldehyde suggests that methylation does not occur prior to formation of norbelladine. The methylation of dopamine is expected considering structural similarity to the methylated moiety of norbelladine. Tyramine was not methylated; this is as expected for a class I OMT (Table 2.1).

Several aspects of the candidate gene selection approach proved important for this successful identification. One is the selection of a variety of methyltransferases for the homology search. If only the known 4-OMTs had been used in the homology search, the gene would have been missed due to the large difference in sequence between known 4-OMTs and *Np*N4OMT. It has been shown that catechol 4'-OMT and catechol 3'-OMT can differ by as little as one amino acid (Wils et al., 2013). Because of this potential for a conversion form catechol 3'-*O*-methylation to 4'-*O*-methylation though evolution, OMTs of both positions were used in the homology search. Also, both class I and class II OMTs were used in the search because both classes are known to

methylate catechols. Considering the multiple branches of the *N*-methyltransferases in the OMT phylogeny, it is worth investigating enzymes that annotate as *N*-methyltransferases (Raman and Rathinasabapathi, 2003). For these reasons, the sequences used in the initial BLAST search consisted of representatives of known *O*- and *N*-methyltransferases of small metabolites. The *Np*N4OMT turned out to be a member of the class I OMTs. Class I OMTs show closer homology to human catechol OMT than to all known plant catechol 4-OMTs in class II, as demonstrated in Figure 2.5 (Ibdah et al., 2003). The closest known catechol 4-OMT to *Np*N4OMT is bacterial, has 34 % identity to *Np*N4OMT, and is a class I OMT from *Cyanobacterium Synechocystis* sp. Strain PCC 6803 (*Syn*OMT) (Kopycki et al., 2008). Many 3-OMTs show even higher homology to *Np*N4OMT than *Syn*OMT. It is probable that the 4-OMT activity of *Np*N4OMT was acquired independently of *Syn*OMT (Figure 2.5).

The second selection criterion, co-expression with galanthamine accumulation, was also of great value. It reduced the number of candidate OMTs from hundreds to one. There are a variety of methods for the prioritization of candidate genes (Moreau and Tranchevent, 2012). Many of these methods are oriented towards species and systems for which there are extensive databases or prior knowledge regarding a gene involved in the pathway or process. In one study, a collection of ~500 microarrays was used to demonstrate the co-expression of genes in the same pathway in Arabidopsis (Wei et al., 2006; Saito et al., 2008). However, extensive gene expression data are typically not available for non-model systems. There have been several studies that use co-expression analysis to find genes in a pathway and produce promising candidate gene lists. These studies often lack biochemical validation of the in *silico* candidates (Sun et al., 2011). If a novel function is proposed, this type of analysis is incomplete without biochemical validation of enzyme activity. Enzymes that are homologous to functionally equivalent enzymes in a different species

can be corroborated by co-expression analysis (Sun et al., 2011). There are several studies that use a simple differential expression model and microarrays to find biosynthetic genes by comparing biosynthetically active and inactive accessions in rose and strawberry (Aharoni et al., 2000; Guterman et al., 2002). Differential expression analysis lacks algorithms to use data with differing levels of metabolism occurring in more than two samples. The Pearson correlation can compare data from multiple samples. Mercke et al. have used a Pearson correlation-based method to identify gene expression with microarrays that correlate with levels of specific terpenes in cucumber (Mercke et al., 2004). Illumina-based transcriptomes are, however, more sensitive to minor variants in the sequences and to splice variants. Illumina-based gene expression data also have a far greater dynamic range, limited only by sequence depth, than microarrays (Wang et al., 2009). Subtleties in the sequences that could be missed with microarrays can now be detected with Illumina sequencing.

The use of HAYSTACK as a platform to use the Pearson correlation is ideal because it is designed to receive a hypothesis for gene expression and look for gene expression that correlates with that hypothesis. This is in contrast to an approach in which gene expression patterns are clustered based on similarity to each other. The search for a defined pattern in the data allows the number of required expression data points to be reduced compared to an approach that needs to define clusters of gene expression patterns based on similarity. In HAYSTACK, the shared expression pattern is already defined. HAYSTACK applies additional screening criteria including a p-value test for significance, a fold cutoff, and background cutoff. The approach chosen in our study used knowledge of known chemical intermediates, a transcriptome with expression profiles for three tissues, and metabolite levels to identify a candidate gene product to validate with *in vitro*

enzyme activity. Little prior knowledge of a pathway is required to use this approach, making this workflow ideal for the identification of genes in unknown biochemical pathways.

There are several modifications to this approach that could be used to improve its power. It could be applied to more tissues, environmental conditions or time points to provide even greater statistical power to correlate co-expression of biosynthetic genes with the biosynthesis of their products. The method could also be modified to include analysis of multiple end products. If the pathway in which the enzyme participates branches, several end products could be equally important to co-expression analysis. This combined consideration of multiple end products could lead to more informative models (Yamazaki et al., 2013). Another potential source of information on the metabolite level could be the concentrations of intermediates made during synthesis. Correlations between biosynthetic gene expression, and perhaps the accumulation of metabolites as well, tend to decrease as distance in a pathway increases (Wei et al., 2006). Experiments that quantitate metabolic intermediates could be useful for finding biosynthetic genes if the flux through the pathway is not so high that intermediates do not accumulate. The latter would be the case in a pathway assembled into a metabolon.

The discovery of this enzyme enables the future elucidation of other enzymes in the proposed galanthamine biosynthetic pathway and other novel pathways. Genes that co-express with *Np*N4OMT can be identified and used as candidate genes for other steps in the proposed galanthamine biosynthetic pathway. This will potentially be useful for earlier steps in the pathway, considering the tendency of expression correlations to decrease as distance in metabolic pathways increase (Wei et al., 2006). This enzyme discovery also validates the utility of this workflow to characterize metabolic pathways and provides a valuable method for pathway discovery in orphan species.

2.5 Acknowledgments

We thank Judy Coyle for figure preparation, Robin S. Kramer (National Center for Genome Resources) for contributing to the transcriptome assembly, Bradley S. Evans (Donald Danforth Plant Science Center) for assistance in mass spectral interpretation and John E. Cronan (Department of Microbiology, University of Illinois) for the *E. coli* culture with the pfs construct.

Chapter 3: CYP96T1 of *Narcissus* sp. *aff. pseudonarcissus* Catalyzes Formation of the *Para-Para' C-C* Phenol Couple in the Amaryllidaceae Alkaloids

This chapter has been submitted for publication as:

CYP96T1 of *Narcissus* sp. *aff. pseudonarcissus* Catalyzes Formation of the *Para-Para' C-C* Phenol Couple in the Amaryllidaceae Alkaloids

Matthew B. Kilgore[1], Megan M. Augustin [1], Gregory D. May[2], John A. Crow[2], Toni M. Kutchan[1]

[1]Donald Danforth Plant Science Center, St. Louis, MO, USA

[2]National Center for Genome Resources, Santa Fe, NM, USA

The Amaryllidaceae alkaloids are a family of amino acid derived alkaloids with many biological activities; examples include haemanthamine, haemanthidine, galanthamine, lycorine, and maritidine. Central to the biosynthesis of the majority of these alkaloids is a *C-C* phenol-coupling reaction that can have *para-para'*, *para-ortho',* or *ortho-para'* regiospecificity. Through comparative transcriptomics of *Narcissus* sp. *aff. pseudonarcissus*, *Galanthus* sp., and *Galanthus elwesii* we have identified a *para-para' C-C* phenol coupling cytochrome P450, CYP96T1, capable of forming the products (10b*R*,4a*S*)-noroxomaritidine and (10b*S*,4a*R*)-noroxomaritidine from 4'-*O*-methylnorbelladine. CYP96T1 was also shown to catalyzed formation of the *para-ortho'* phenol coupled product, *N*-demethylnarwedine, as less than 1 % of the total product. *CYP96T1* co-expresses with the previously characterized norbelladine 4'-*O*-methyltransferase.

The discovery of CYP96T1 is of special interest because it catalyzes the first major branch in Amaryllidaceae alkaloid biosynthesis. CYP96T1 is also the first phenol-coupling enzyme characterized from a monocot.

3.1 Introduction

The Amaryllidaceae alkaloids are produced by species of Amaryllidaceae including *Narcissus* spp. (daffodil) and *Galanthus* spp. (snowdrop). Alkaloids from all major structural classes of Amaryllidaceae alkaloids have biological activities. Some of these alkaloids have potential pharmaceutical applications or are already established medicines. The alkaloid skeleton types, haemanthamine, narciclasine, tazettine, and montanine are derived from the *para-para' C-C* phenol coupled, (10b*R*,4a*S*)-noroxomaritidine, biosynthetic precursor (Wildman and Bailey, 1969; Fuganti et al., 1971; Feinstein and Wildman, 1976) (see Figure 3.1 and Figure 3.2 for representative structures). Specific examples of alkaloids derived from (10b*R*,4a*S*)-noroxomaritidine include haemanthamine, maritidine, vittatine, and pretazettine. Haemanthamine has been shown to have antiproliferative and apoptotic effects on cancer cell lines and antioxidant activity in a 2,2-diphenyl-1-picrylhydrazyl scavenging assay (Oloyede et al., 2010; Van Goietsenoven et al., 2010; Havelek et al., 2014). Crinine and its derivatives are also derived from a *para-para' C-C* phenol coupling, however, the phenol-coupled product is the enantiomer, (10b*S*,4a*R*)-noroxomaritidine. Antibacterial activities have been noted for the derivatives of the (10b*S*,4a*R*)-noroxomaritidine skeleton including buphanidrine and distichamine (Cheesman et al., 2012). An example of an *ortho-para*' *C-C* phenol-coupling product is lycorine, derived from noroxopluvine. Lycorine has been documented to cause apoptosis in leukemia and multiple myeloma cancer cell lines (Liu et al., 2004; Li et al., 2007; Liu et al., 2009). Galanthamine is a representative derivative of the *para-ortho' C-C* phenol-coupled product *N*-demethylnarwedine

and is used as an Alzheimer's treatment drug (Wilcock et al., 2003). It acts through acetylcholine esterase inhibition and nicotinic receptor binding (Irwin and Smith, 1960; Barik et al., 2005). The limited supply of some Amaryllidaceae alkaloids and diversity of biological activities make the biosynthesis of Amaryllidaceae alkaloids a topic of interest for biotechnology.

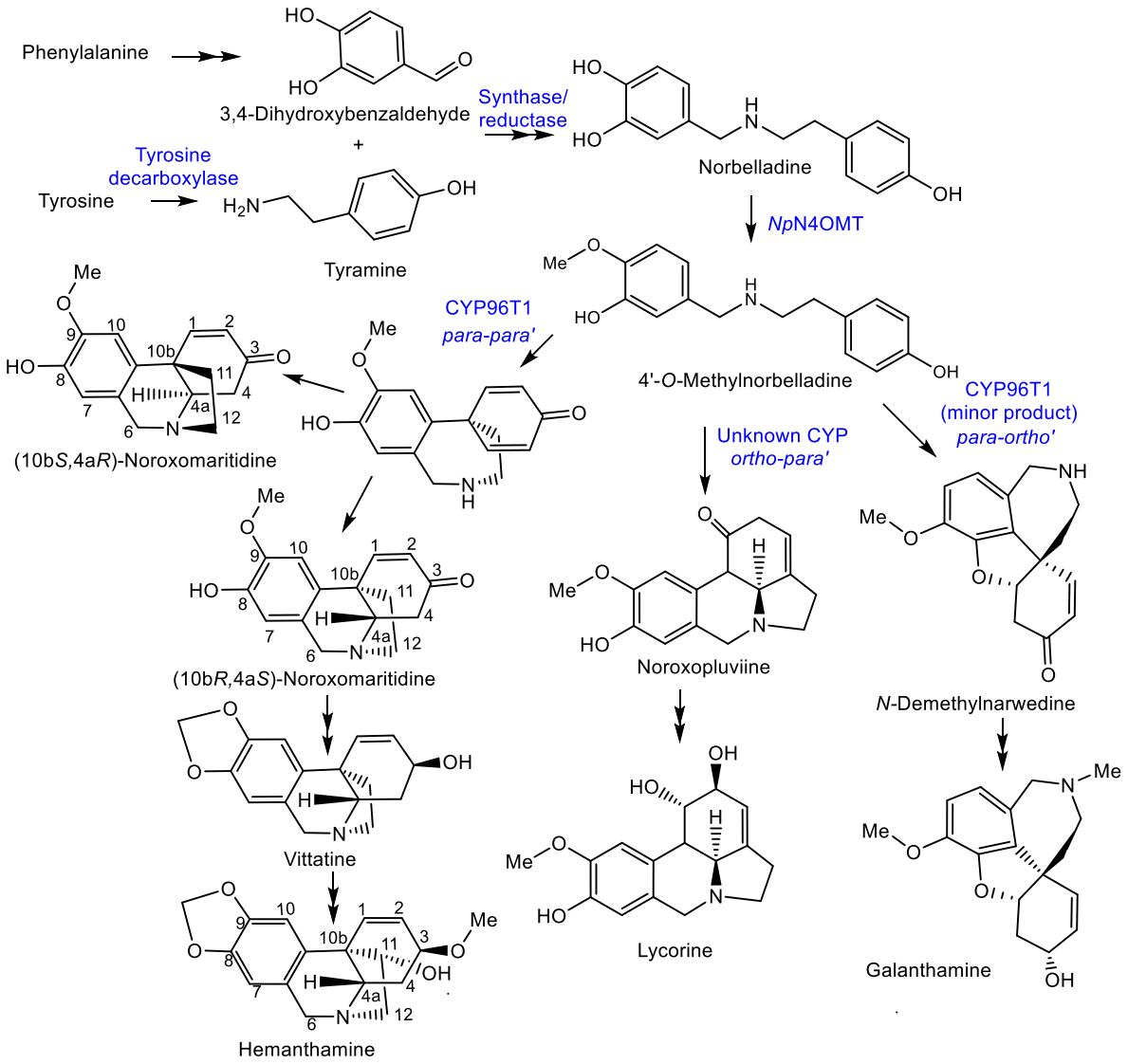

Figure 3.1. Proposed biosynthetic pathways for representative Amaryllidaceae alkaloids directly derived from *C*-*C* phenol coupling. The previously discovered *Np*N4OMT, the CYP96T1 discovered in this study, and potential enzyme classes involved in each step of the pathways are in blue.

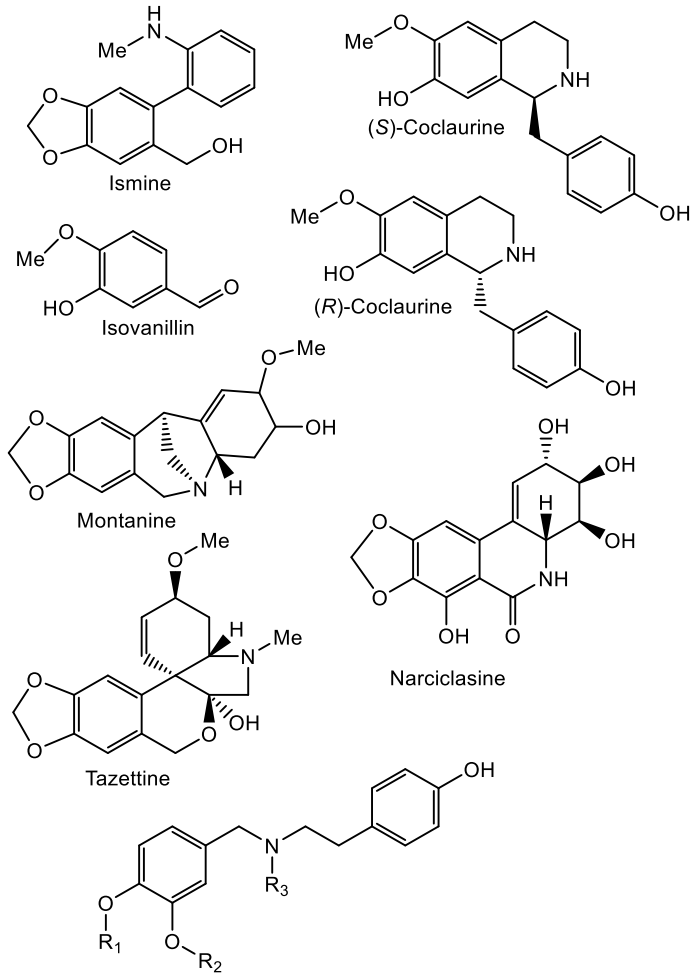

Figure 3.2. Structures of relevant compounds.

Haemanthamine is derived from the amino acids phenylalanine and tyrosine. Phenylalanine was established as a precursor by incorporation of [3-^{14}C]phenylalanine into haemanthamine in *Nerine bowdenii* (Wildman et al., 1962b). The conversion of phenylalanine to 3,4-dihydroxybenzaldehyde in haemanthamine biosynthesis was clarified by feeding [3-^{14}C]*trans*-cinnamic acid, [3-^{14}C]*para*-hydroxycinnamic acid, [7-^{14}C]benzaldehyde, [7-^{14}C]*para*-hydroxybenzaldehyde, [^{3}H]3,4-dihydroxybenzaldehyde, and [^{3}H]threo-DL-phenylserine to *Narcissus pseudonarcissus*. Incorporation into haemanthamine from the precursors [3-^{14}C]*trans*-

cinnamic acid, [3-^{14}C]*para*-hydroxycinnamic acid, [^{3}H]3,4-dihydroxybenzaldehyde, and trace incorporation of [7-^{14}C]*para*-hydroxybenzaldehyde, but not other administered compounds, lead to the conclusion that the pathway proceeds as follows: phenylalanine is converted to *trans*-cinnamic acid, then to *para*-hydroxycinnamic acid, then to 3,4-dihydroxycinnamic acid or *para*-hydroxybenzaldehyde followed by conversion to 3,4-dihydroxybenzaldehyde (Suhadolnik et al., 1963a). Observation of radiolabeled [3-^{14}C]tyrosine incorporation into haemanthamine established tyrosine as a precursor (Jeffs, 1962). Tyrosine was demonstrated to contribute to C11 and C12 of haemanthamine by feeding [β-^{14}C]tyrosine to *Sprekelia formosissima* and [α-^{14}C]tyrosine to *Narcissus* 'Twink' daffodil followed by haemanthamine degradation experiments (Battersby et al., 1961a; Wildman et al., 1962a) (Figure 3.1). These results indicate tyrosine also contributes C1-4, C4a, and C10b because of their ring shape and proximity to the C11 and C12 of haemanthamine. Equivalent sections of the galanthamine and lycorine carbon skeleton also originate from tyrosine (Battersby and Binks, 1960; Barton et al., 1963). Tyrosine is converted into tyramine by tyrosine decarboxylase, a well characterized enzyme in other secondary metabolite pathways (Lehmann and Pollmann, 2009).

3,4-Dihydroxybenzaldehyde and tyramine are condensed to a Schiff-base and reduced to norbelladine. The central role of norbelladine in Amaryllidaceae alkaloid biosynthesis was demonstrated by incorporation of [1-^{14}C]norbelladine into haemanthamine, lycorine, and galanthamine (Barton et al., 1961; Battersby et al., 1961b; Battersby et al., 1961a). Next, norbelladine is methylated to 4'-*O*-methylnorbelladine. In 1963, crude enzyme extracts of *N. bowdenii* were used to perform a preliminary characterization of the 4'-*O*-methyltransferase conducting this methylation (Mann, 1963). This cation-dependent norbelladine 4'-*O*-methyltransferase (*N4OMT*) was identified in *Narcissus* sp. *aff. pseudonarcissus*, and

enzymatically characterized by heterologous expression in *E. coli* in Chapter 2 (Kilgore et al., 2014). (10b*R*,4a*S*)-Noroxomaritidine is formed from the *para-para' C-C* phenol coupling of 4'-*O*-methylnorbelladine. The biosynthesis of haemanthamine deviates from alkaloids with *ortho-para'* and *para-ortho*' carbon skeletons at this branch point. The next step is thought to be a reduction of the ketone group to synthesize 8-*O*-demethylmaritidine followed by an oxide bridge formation to form vittatine. Conversion of vittatine to haemanthamine is thought to occur through hydroxylation followed by methylation (Figure 3.1). The conversion of vittatine to haemanthamine has been demonstrated by radiolabeling studies. The order of hydroxylation and methylation in this conversion is inferred from the presence of the hydroxylated 11-hydroxyvittatine in the *N. bowdenii* plants under investigation and the absence of the methylated (10b*R*,4a*S*)-buphanisine (Feinstein and Wildman, 1976). Haemanthamine accumulates *in planta* and is modified further to compounds such as haemanthidine and pretazettine in some Amaryllidaceae. The proposed biosynthesis of galanthamine from the *ortho-para'* product *N*-demethylnarwedine through the reduced intermediate *N*-demethylgalanthamine has been reviewed in Chapter 1 and discussed in Chapter 2 (Eichhorn et al., 1998; Kilgore et al., 2014).

Cytochrome P450 enzymes are a diverse enzyme family with numerous functions. Reactions catalyzed include hydroxylation, *C-C* and *C-O* phenol coupling, oxide bridge formation, carbon-carbon bond cleavage, demethylation, and rearrangements of carbon skeletons (Mizutani and Sato, 2011). Previously documented *C-C* phenol coupling by cytochrome P450 enzymes that synthesize salutaridine (CYP719A1), (*S*)-corytuberine (CYP80G2), and cyclodipeptide cyclo(l-Tyr-l-Tyr) (CYP121), suggest the *C-C* phenol coupling reactions found in Amaryllidaceae alkaloid biosynthesis are cytochrome P450 dependent (Ikezawa et al., 2008; Belin et al., 2009; Gesell et al., 2009). *C-C* phenol coupling reactions have also been documented in the human cytochromes

P450 CYP2D6 and CYP3A4 with the substrate (*R*)-reticuline (Grobe et al., 2009). In addition to cytochromes P450 peroxidases and laccases are documented phenol-coupling enzymes (Schlauer et al., 1998; Davin et al., 2008; Constantin et al., 2012)

Orphan plant species are frequently of interest due to their unique metabolism. Study of this metabolism is problematic due to scarcity of genetic information, limited mutant libraries, and lack of efficient transformation methods. In addition, secondary metabolites can be phylogenetically restricted. Method development for efficient metabolic pathway elucidation in orphan species is therefore desirable. An efficient work-flow for the identification of biosynthetic genes has been previously developed and applied to Amaryllidaceae alkaloid biosynthesis. Methyltransferase transcripts correlating with galanthamine accumulation in *N.* sp. *aff. pseudonarcissus* were targeted and tested for norbelladine 4'-*O*-methyltransferase activity, leading to the discovery of the biosynthetic gene *N4OMT* in Chapter 2 (Kilgore et al., 2014). In this study, a similar work-flow is applied utilizing transcriptomic data from multiple species to identify cytochrome P450 genes that co-express with *N4OMT*. This led to the isolation and characterization of CYP96T1, which catalyzes formation of the *para-para'* and a small quantity of the *para-ortho' C-C* phenol couple with 4'-*O*-methylnorbelladine.

3.2 Materials and Methods

3.2.1 Plant Tissue and Chemicals

Leaf, bulb, and inflorescence tissues were collected from adult blooming *N.* sp. *aff. pseudonarcissus* and *Galanthus* sp. plants in St. Louis, MO and *Galanthus elwesii* in Pullman, WA. Chemicals acquired from Sigma Aldrich include ammonium acetate 97 % A.C.S. reagent, HPLC grade ethanol, and tyramine 99 %. Other chemicals purchased include ammonium acetate extra pure 25 % solution in water from Acros Organics, ampicillin from GoldBio, and vanillin

from Merck. Several compounds were obtained previously by synthesis including: veratraldehyde, (*R*)-coclaurine, and (*S*)-coclaurine. Haemanthamine was previously isolated from *Narcissus pseudonarcissus*. Additional materials are as described previously in Chapter 2 and (Gesell et al., 2009).

3.2.2 Transcriptome Assembly and Transcript Abundance Estimation

The transcriptomes assembled using ABySS and MIRA for *Galanthus* sp. and *G. elwesii* were assembled in the same manner as in Chapter 2 for the previously described ABySS and MIRA *N.* sp. *aff. pseudonarcissus* transcriptome (Kilgore et al., 2014), but with 50 base paired-end reads with leaf, bulb, and inflorescence tissues. Alternative transcriptomes were made using Trinity. For these transcriptomes the same raw reads were assessed using FastQC followed by trimming with the FASTX tool kit. The fastx_trimmer was used to remove the first 13 bases and fastq_quality_trimmer was used to remove all bases on the 3' end with a Phred quality score lower than 28. Sequences below 30 bases or without a corresponding paired-end read were removed from the trimmed data set. Cleaned reads were input into the Trinity pipeline with default parameters for each data set (Haas et al., 2013). The unprocessed reads and Trinity assemblies were used with the Trinity tool RNA-Seq by Expectation-Maximization (RSEM) to obtain the transcripts per million mapped reads (TPM) for all transcripts in each tissue (leaf, bulb, and inflorescence) for each Trinity assembly. To assess quality, the following parameters were considered: the size of the resulting assembly and identification of homologues to the conserved genes *Zea mays* MADS6 (NP_001105153.1), *Arabidopsis thaliana* ribulose bisphosphate carboxylase small chain 1A (NP_176880.1), and the *Oryza sativa* ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit (AAB02583.1). Assemblies and transcript expression data are deposited in the MedPlant

RNA Seq Database, http://www.medplantrnaseq.org. ESTScan trained against *A. thaliana* open reading frames was used to predict peptides encoded in all Trinity assemblies (Iseli et al., 1999).

3.2.3 Candidate Gene Identification

BLASTP with an e-value cut off of 1×10^{-4} was used to find homologs to known cytochrome P450 enzymes in all transcriptomes. A list of 472 unique, curated plant cytochrome P450 sequences from Dr. David Nelson, University of Tennessee, was used as a query against the ESTScan predicted peptides for each assembly (supplemental File 1). HAYSTACK was used to find correlations between the appropriate *N4OMT* expression model for each assembly (Table 3.1) and the transcripts in each assembly. All *Galanthus* models were based on the expression estimates for the closest *NpN4OMT1* homologue in the assembly being used. The *N.* sp. *aff. pseudonarcissus* model was based on the RT-PCR data for *NpN4OMT1* expression obtained in Chapter 2 (Kilgore et al., 2014). HAYSTACK parameters are as follows: correlation cutoff ≥ 0.8, background cutoff ≥ 1, fold cutoff ≥ 4, and p-value cutoff ≤ 0.05 (Mockler et al., 2007). Homologues to annotated cytochrome P450 enzymes that were correlating with the *N4OMT* models were identified using BLASTN with an e-value cut off of 1×10^{-50} queried against the *N4OMT* co-expressing candidates in every other assembly. For each cytochrome P450 candidate, the total number of assemblies with an *N4OMT* co-expressing BLASTN hit were determined. Candidates present in five of the five comparable lists were considered top priority candidate genes and were cloned (Figure 3.3).

Table 3.1. Models used in HAYSTACK analysis

Model name	Leaf	Inflorescence	Bulb
[N] *N.* sp. *aff. pseudonarcissus N4OMT* (relative units)	1	30	45
[N] *Galanthus* sp. *N4OMT* (RPM)	0.01	33.34	139.79
[N] *Galanthus elwesii N4OMT* (RPM)	2.24	22.59	71.71
[TC] *N.* sp. *aff. pseudonarcissus N4OMT* (TPM)	NA	NA	NA
[T] *Galanthus* sp. *N4OMT* (TPM)	2.42	29.02	94.73
[T] *Galanthus elwesii N4OMT* (TPM)	15.95	49.32	201.97

[N] AbySS and MIRA assembly
[C] homologue not found
[T] Trinity assembly
RPM=reads per million
NA=not applicable

3.2.4 Polymerase Chain Reaction (PCR) and Cloning

After cloning *CYP96T1*, designated as narcissus-20101112|22907 in the ABySS and MIRA *N.* sp. *aff. pseudonarcissus* assembly, a close homologue of *CYP96T1* with the designation narcissus-20101112|13079 was identified with a complete ORF. Primers based on the 5' sequence of narcissus-20101112|13079 and the 3' sequence of *CYP96T1*. Design of inner primers was validated by sequencing (GENEWIZ Inc.) of the outer PCR product using the following primers *CYP96T* forward outer (5'-ACATCCCCCCCAAAAAAATCATAAC-3'), *CYP96T* reverse outer (5'-AGACCATCAATGTGATCACCA-3'), and *CYP96T* reverse sequencing (5'-TGGTGAAATCGTTGAATTGGTTGT-3'). *N.* sp. *aff. pseudonarcissus* bulb cDNA was prepared as described in Chapter 2 (Kilgore et al., 2014). The outer PCR reaction contained 0.2 µM each of *CYP96T* forward outer and *CYP96T* reverse outer primers, 1 µl of bulb cDNA, 1X Phusion HF reaction buffer, 0.3 mM dNTPs, and 1 U NEB Phusion High-Fidelity DNA Polymerase in a 50 µl reaction. The PCR program parameters were as follows: 30 s 98 ºC 1 cycle, 15 s 98 ºC, 30 s 52 ºC, 2 min 72 ºC 35 cycles, and 5 min 72 ºC 1 cycle. Inner PCR parameters were identical except primers used were *CYP96T* forward inner (5'-aattGCGGCCGCATGGCCACTTCTTCTTCAGCA-3') and *CYP96T* reverse inner (5'-

aattTCTAGATCACATGACTGATCTCTTTCT-3') adding NotI and XbaI restriction sites, respectively (underlined), the outer PCR reaction was added as template as opposed to cDNA, and the PCR cycles were reduced to 25.

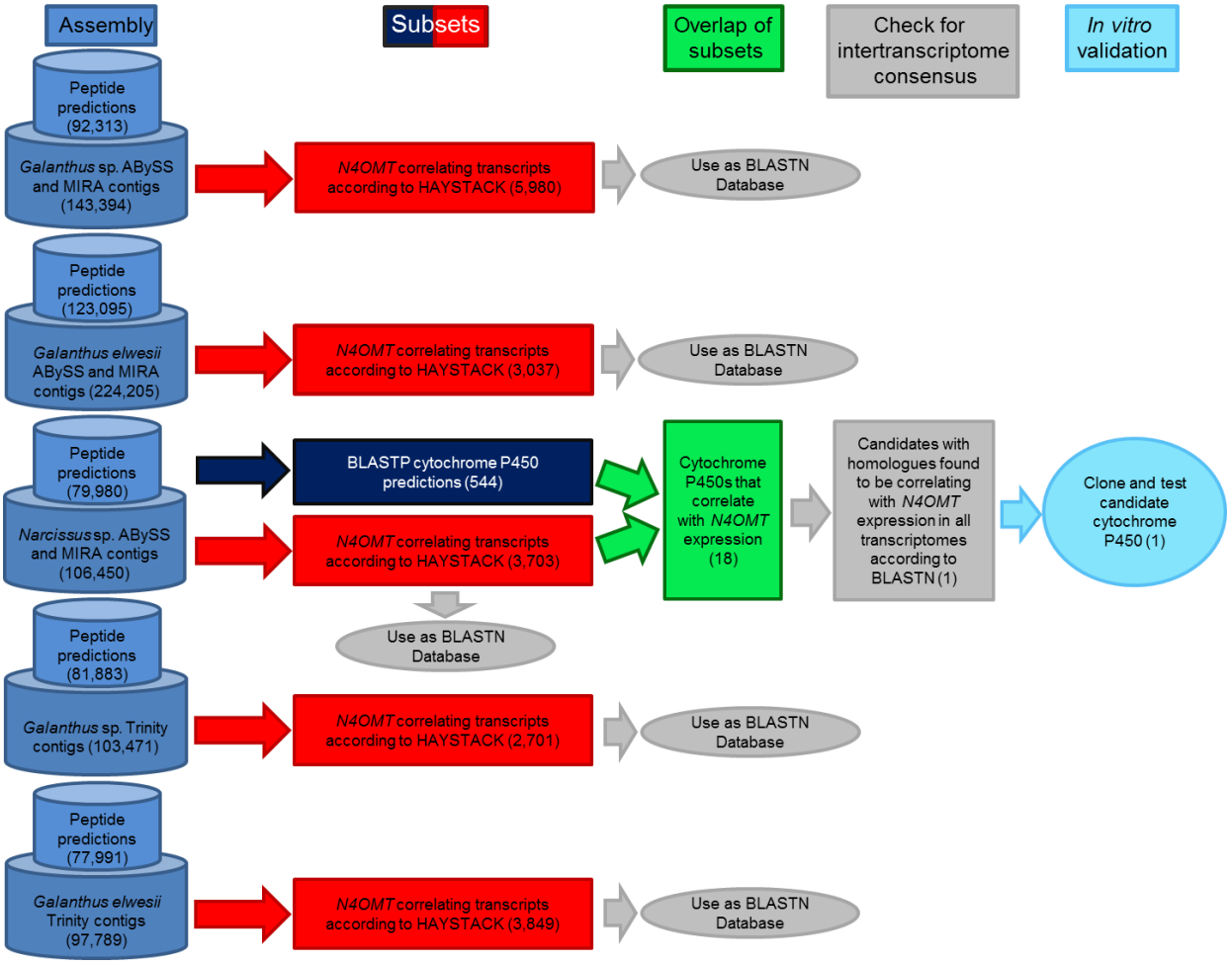

Figure 3.3. Work-flow for identification of candidate cytochrome P450 enzymes. Following the generation of transcriptome assemblies, cytochrome P450 enzymes were identified with BLASTP (Navy blue) and genes correlating with *N4OMT* were identified with HAYSTACK (Red). The genes present in both lists makeup the initial candidate gene list (Green). Homologues of these genes were identified in the *N4OMT* correlating lists of the other transcriptomes using BLASTN (Gray). Candidates with homologues in all five *N4OMT* correlating lists were cloned from *N.* sp. *aff. pseudonarcissus*, *Narcissus* sp. (light blue). The analysis for the *N.* sp. *aff. pseudonarcissus* ABySS and MIRA assembly is completely diagrammed to illustrate the process followed in every assembly. The number of transcripts selected in each step is in parentheses. The *N.* sp. *aff. pseudonarcissus* Trinity assembly is excluded from this work-flow due to its poor quality.

The inner PCR product and pVL1392 vector were digested with NEB NotI and XbaI followed by PCR purification with the QIAquick PCR Purification Kit and ligation with NEB T4

DNA ligase according to manufacturer's instructions. The resulting construct was transformed into DH5α *E. coli* chemically competent cells. Recombinant bacteria were selected on Luria-Bertani 1.5 % agar plates with 50 µg/ml ampicillin.

Transformants were screened by colony PCR with the following components: 0.67 µM polyhedrin forward (5'-AAAATGATAACCATCTCGC-3') and polyhedrin reverse (5'-GTCCAAGTTTCCCTGTAGA-3') primers, ThermoPol Reaction buffer 1X, 0.2 mM dNTPs, and 1 U of NEB Taq DNA polymerase in a volume of 20 µl. The PCR parameters were as described in Chapter 2 (Kilgore et al., 2014). Plasmid minipreps of 5 ml cultures in LB supplemented with 50 µg/ml ampicillin were prepared using the QIAGEN QIAprep Spin Miniprep Kit according to manufacturer's instructions and sequenced with polyhedrin forward and reverse primers by Eurofins genomics. All reproducible sequences were named by Dr. David Nelson, University of Tennessee. The closest biochemically characterized homologue to the resulting *CYP96T1* clone, CYP96A15, was identified in the UniProt database with BLASTP. CYP97T1 (KT693311), CYP96T2 (KT693312), CYP96T3 (KT693313), the original CYP96T1 sequence, and CYP96A15 were aligned with MUSCLE in the CLC main workbench version 6.9.1 (Edgar, 2004).

3.2.5 Protein Expression

Co-transfection of *CYP96T1* in pVL1392 and Baculogold baculoviurus (BD Biosciences), viral amplification, protein expression, and microsome preparation in *Spodoptera frugiperda* Sf9 cells was performed as previously described (Gesell et al., 2009). Microsomes of Sf9 cells expressing CYP96T1 were solubilized with 0.17 % emulgen 913 at 4 ºC for 15 min followed by centrifugation (15,000 X g for 15 min) before obtaining the CO difference spectra. The resulting CYP96T1 concentration was used to calculate concentration of CYP96T1 in all subsequent Sf9 cell cultures.

CYP96T1 was always co-expressed with *Eschscholzia californica* cytochrome P450 reductase (*CPR*) (Gesell et al., 2009). For a negative control CPR was expressed without *CYP96T1*.

3.2.6 3'-*O*-Methylnorbelladine and 3',4'-*O*-Dimethylnorbelladine Synthesis

For the synthesis of 3'-*O*-methylnorbelladine sodium cyanoborohydride (48.5 mM), vanillin (5.66 mM), and tyramine (5.15 mM) were mixed in 2.5 ml anhydrous methanol for 2 days at room temperature. For synthesis of 3',4'-*O*-dimethylnorbelladine, sodium cyanoborohydride (65.18 mM), veratraldehyde (9.40 mM), and tyramine (7.29 mM) were mixed in 2.5 ml anhydrous methanol for 2 days at room temperature. The reaction mix was then taken to dryness under N_2. The resulting material was suspended in 200 µl of 1 M $NaCO_3$ pH 9.5 and extracted twice with 400 µl of ethyl acetate by vortexing for 1 min, followed by centrifugation at 16.1 x *g* for 2 min at room temperature. Ethyl acetate extractions were pooled and dried under vacuum. Extractions were re-suspended in 10 % acetonitrile and 0.1 % formic acid and purified by fractionation using a Waters fraction collector III and Waters 1525 binary HPLC pump as described in Chapter 2 (Kilgore et al., 2014). 3'-*O*-methylnorbelladine was collected at 9 min for ~1 min and 3',4'-*O*-dimethylnorbelladine at 11.5 min for ~1 min. Purified compound was dried under vacuum, re-suspended in H_2O and quantified with a 4'-*O*-methylnorbelladine standard curve using peak area by HPLC with the method stated above.

3.2.7 Enzyme Assays

Screening assays contained 30 mM KPO_4 pH 8.0, 1.25 mM NADPH, 10 µM substrate, and 70 µl of virus infected Sf9 cell suspension in 200 µl total volume. The assays were incubated for 2-4 hr at 30 ºC. 4'-*O*-metylnorbelladine was tested for all CYP96T variants. CYP96T1 was used for substrate specificity tests on norbelladine, *N*-methylnorbelladine, 4'-*O*-methyl-*N*-methylnorbelladine, 3'-*O*-methylnorbelladine, 3',4'-*O*-dimethylnorbelladine, haemanthamine,

(*S*)-coclaurine, (*R*)-coclaurine, and mixed (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine (see Figure 3.1 and Figure 3.2 for chemical structures). Assays derivatized with sodium borohydride were incubated 2 hr at 30 ºC followed by addition of 0.5 volumes 0.5 M sodium borohydride in 0.5 M sodium hydroxide and incubated 30 min at RT. The CYP96T1 assay resolved on a Chiral-CBH column and assays measured with HPLC used fresh *CYP96T1* and *CPR* expressing Sf9 cell protein prepared using re-amplified virus. Kinetic assays were run in the linear time range for each substrate in 200 mM KPO_4 pH 6.5 buffer with 40 µl assays. Product for overnight enzyme assays of (*R*)-coclaurine, (*S*)-coclaurine, and 4'-*O*-methyl-*N*-methylnorbelladine were quantified at 277 nm against a noroxomaritidine standard curve with the same HPLC method and setup used for 3'-*O*-methylnorbelladine and 3',4'-*O*-dimethylnorbelladine isolation. These products were subsequently used as standards for quantifying kinetic assays. K_m and k_{cat} values were estimated using R version 3.2.0 with nonlinear fitting.

3.2.8 LC-MS/MS

Enzyme assays on all substrates were extracted as previously described and run on a QTRAP 4000 coupled to a IL-20AC XR prominence liquid auto sampler, 20AD XR prominence liquid chromatograph and Phenomenex Luna 5 µm C8(2) 250 x 4.60 mm column. HPLC gradient and MS settings were as described in Chapter 2 (Kilgore et al., 2014). Assay specific MS/MS parameters are presented in Table 3.2. Multiple Reaction Monitoring (MRM) parameters for relative quantification of (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine, *N*-demethylnarwedine, narwedine, and the two unknown compounds are presented in Table 3.3. For analysis of product chirality, a Chrom Tech, Inc. Chiral-CBH 100 x 4.0 mm, 5 µM column was used with a 30 min isocratic flow of 2.5 % HPLC grade ethanol and 10 mM ammonium acetate with pH adjusted to 7.0 with ammonium hydroxide. Kinetic assays were quantified with an isocratic flow 20 %

acetonitrile and 0.08 % formic acid with the Phenomenex Luna 5 µm C8(2) 250 x 4.60 mm column connected to the same QTRAP 4000 setup. MRM transitions used in kinetics were 284.1/223.0 *m/z* for the (*S*)-coclaurine and (*R*)-coclaurine products, 286.1/271.0 *m/z* for the 4'-*O*-methyl-*N*-methylnorbelladine *para-para'* product, and 272.3/229.0 *m/z* for noroxomaritidine.

Table 3.2. MS/MS parameters for substrate tests

Substrate	Product specific parameters (CE)(DP)(Q1 *m/z*)	Substrate specific parameters (CE)(DP)(Q1 *m/z*)
4'-*O*-Methylnorbelladine	(35)(70)(272.30)	(20)(60)(274.30)
4'-*O*-Methyl-*N*-methylnorbelladine	(35)(70)(286.20)	(20)(60)(288.30)
3'-*O*-Methylnorbelladine	(35)(70)(272.30)	(35)(60)(274.30)
3',4'-*O*-Dimethylnorbelladine	(35)(70)(286.20)	(20)(60)(288.30)
Norbelladine	(35)(60)(258.00)	(15)(50)(260.00)
N-Methylnorbelladine	(35)(70)(272.30)	(20)(60)(274.30)
Haemanthamine	(35)(70)(300.12)/ (35)(70)(318.13)^{HO}	(35)(70)(302.14)
(10b*S*,4a*R*)- and (10b*R*,4a*S*)-Noroxomaritidine	(35)(70)(270.30)/ (35)(70)(288.30)^{HO}	(35)(70)(272.30)
Isovanillin and tyramine	(20)(40)(290.30)[a]/ (20)(60)(272.20)[b]/ (35)(70)(270.20)[c]	(20)(60)(138.20)/ (20)(50)(153.20)
(*S*)-Coclaurine	(35)(70)(284.30)/ (30)(60)(570.60)^{dim}	(20)(70)(286.30)
(*R*)-Coclaurine	(35)(70)(284.30)/ (30)(60)(570.60)^{dim}	(20)(70)(286.30)
4'-*O*-Methylnorbelladine assays followed by sodium borohydride derivatization	(20)(60)(274.30)	(20)(60)(274.30)

^{HO} hydroxylation monitored

^{dim} dimer formation monitored

[a] *C-C* phenol coupling with no amine aldehyde condensation

[b] amine aldehyde condensation/amine aldehyde condensation with *C-C* phenol coupling and a reduction.

[c] amine aldehyde condensation with *C-C* phenol coupling

Table 3.3. MS/MS parameters used in MRM studies

Compound(*C-C* phenol coupling type)	MRM parameters(CE)(DP)(Q1 *m/z*)(Q2 *m/z*)(RT min)
Noroxomaritidine(*para'-para*)	(35)(70)(272.3)(229.0)(5.3)
N-Demethylnarwedine(*para'-ortho*)	(35)(70)(272.3)(201.0)(7.9)
4'-*O*-Methyl-*N*-methylnorbelladine assay unknown 1(potential *para'-para* product)	(35)(70)(286.1)(271.0)(4.7)
4'-*O*-Methyl-*N*-methylnorbelladine assay unknown 2 (potential *ortho'-para* product)	(30)(70)(286.1)(243.0)(7.5)
Narwedine(*para'-ortho*)	(30)(70)(286.1)(229.1)(8.1)

3.3 Results

3.3.1 Transcriptome Assembly and Transcript Abundance Estimation

Key statistics for each transcriptome including total number of transcripts, maximum transcript length, and average transcript length are summarized in Table 3.4. ABySS and MIRA assemblies were found to have a high number of incomplete ORFs. This was problematic for cloning and highlighted the potential problem of unannotated transcripts and inaccurate expression estimates in transcripts with short assemblies. Quality processed reads were reassembled with Trinity to provide alternate information on the same transcripts. These assemblies provided additional sequence information with comparable expression estimates. The *N.* sp. *aff. pseudonarcissus* Trinity assembly resulted in a large number of contigs but lacked well-characterized genes, such as ribulose bisphosphate carboxylase small chain 1A and *NpN4OMT1*. In addition, the maximum contig length was 73,933, well above the expected size range. For these reasons, further analysis of the *N.* sp. *aff. pseudonarcissus* Trinity assembly was abandoned. The other assemblies have comparable statistics regardless of assembly method. Because these assemblies are complementary to each other, both sets of *Galanthus* assembles were used for subsequent analysis (Table 3.4).

Table 3.4. Transcriptome statistics

	N. sp. *aff. pseudonarcissus* AbySS and MIRA*[!]	*Galanthus* sp. AbySS and MIRA*	*Galanthus elwesii* AbySS and MIRA*	*N.* sp. *aff. pseudonarcissus* Trinity[#]	*Galanthus* sp. Trinity[#]	*Galanthus elwesii* Trinity[#]
Sequences (bp)	106,450	143,394	224,205	608,439	103,471	97,789
Longest (bp)	13,381	15,365	19,356	40,450	13,629	13,055
N50 (bp)	1,130	1,418	1,330	931	1,044	1,139
Mean (bp)	551	664	602	671	723	777
Median (bp)	248	271	245	430	481	528

*assemblies were set to have a minimum cut off of 100 bp.

[#]assemblies were set to have a minimum cut off of 201 bp.

[!]assembly reported in Chapter 2 (Kilgore et al., 2014)

3.3.2 Candidate Gene Identification and Cloning

The pattern-matching algorithm HAYSTACK was used to identify transcripts that co-express with *N4OMT*. *N4OMT* is the only validated gene involved in Amaryllidaceae alkaloid biosynthesis to date. Its position in the pathway is just prior to the *C-C* phenol-coupling step therefore, *N4OMT* gene expression is a suitable choice to serve as a model for analysis of co-expressing transcripts encoding additional Amaryllidaceae alkaloid biosynthetic genes. Since the *C-C* phenol-coupling enzyme is targeted herein, BLASTP was used to find transcripts that encode putative cytochrome P450 enzymes. The resulting 544 *N.* sp. *aff. pseudonarcissus* cytochrome P450 protein sequences were compared to the list of 3,704 *N4OMT* co-expressing transcripts identified by HAYSTACK. This resulted in the identification of 18 *N4OMT* co-expressing cytochrome P450 transcripts in the *N.* sp. *aff. pseudonarcissus* assembly. The *Galanthus* assemblies were interrogated using these 18 sequences to identify close homologues. This allowed for selection of the cytochrome P450 transcripts that consistently co-expressed with *N4OMT* across species in all assemblies. One candidate (*CYP96T1*) co-expressed with *N4OMT* in all assemblies and was investigated further in *N.* sp. *aff. pseudonarcissus* where its correlation was 0.9995. A close homologue to *CYP96T1* with

99 % identity in shared ORF sequence and the first 67 bases of the 3’ UTR was identified. In contrast to *CYP96T1*, this transcript was complete at the 5’ end of the ORF and contained 5’ UTR sequence information. This allowed the incomplete 5’ region of *CYP96T1* to be predicted by comparison. The PCR product generated with outer primers was sequenced and the inner primer sequences were found not to deviate from the assembly prediction. A clone was acquired with no conflicts to the previously known *CYP96T1* sequence and was used for functional characterization. Two additional variants were cloned reproducibly. The closest biochemically characterized homologue to *CYP96T1* was *CYP96A15* from *A. thaliana* (Q9FVS9) (Figure 3.4).

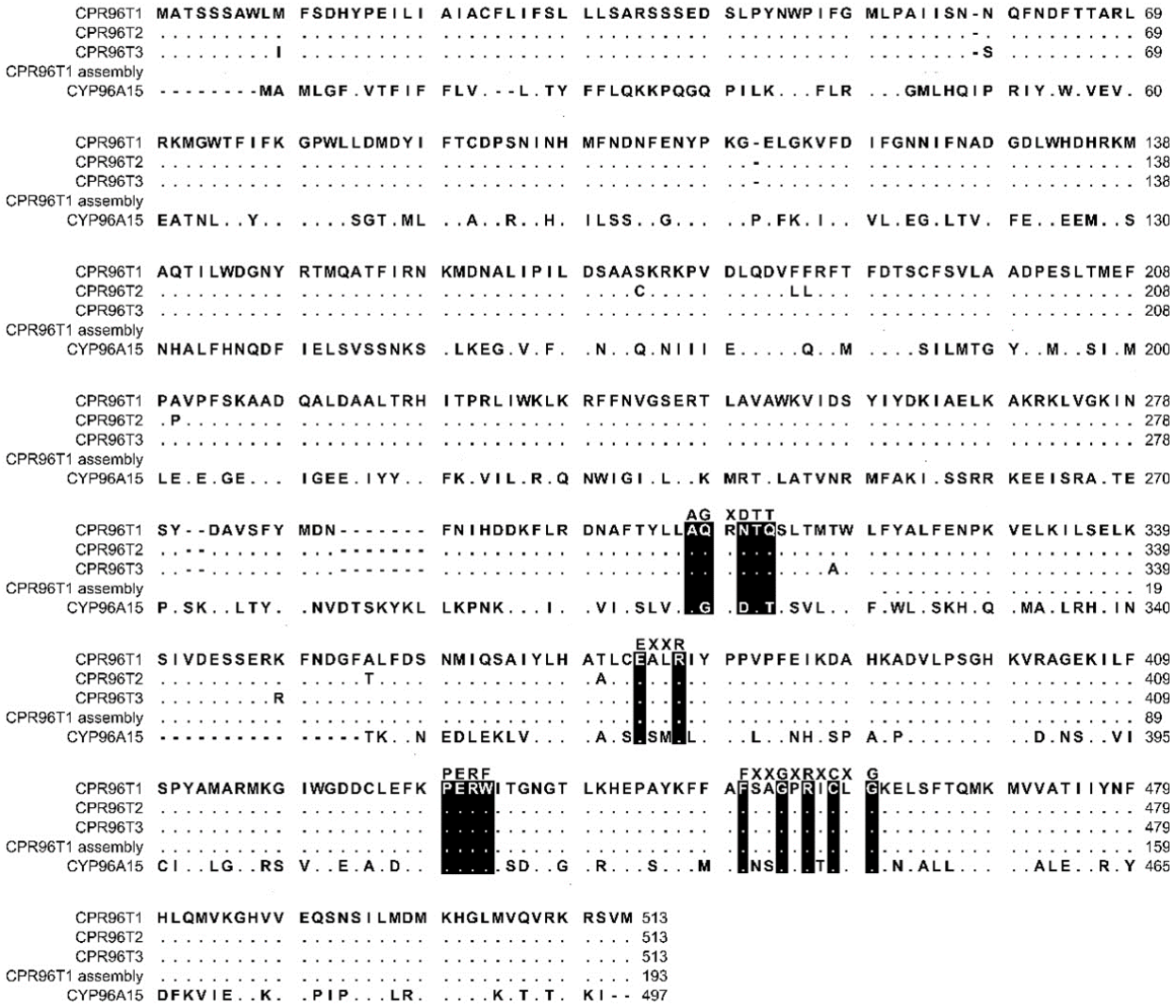

Figure 3.4. MUSCLE alignment of protein sequences for CYP96T1, CYP96T2, CYP96T3, the CYP96T1 sequence from the *N.* sp. *aff. pseudonarcissus* ABySS and MIRA assembly and CYP96A15 from *A. thaliana* (Q9FVS9). Simplified consensus motifs for cytochrome P450 enzymes are placed above the corresponding color inverted CYP96T1 sequence. Dots are exact matches to CYP96T1 and dashes are gaps.

3.3.3 Enzyme Assays and Analysis by LC-MS/MS

The concentration of CYP96T1 in Sf9 cell culture was determined to be 2.5 nM by CO-difference spectra. The temperature and pH optima for 4'-*O*-methylnorbelladine substrate were determined to be 30°C (half height ±5-10°C) and 6.5 (half height ±1), respectively. Testing of the CYP96T1 enzyme demonstrated that several structurally related alkaloids were *C-C* phenol coupled as detected by LC-MS/MS. These reactions were accompanied by a background reaction catalyzed

by the Sf9 cells. 4'-*O*-Methylnorbelladine was *C*-*C* phenol coupled into *N*-demethylnarwedine, (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine in CYP96T1 assays. (10bS,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine were identified by their identical liquid chromatographic retention times (Figure 3.5A) and mass spectrometric fragmentation pattern with (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine mixed standard (Figure 3.5C and D). To determine the chirality of the noroxomaritidine product, 4'-*O*-methylnorbelladine assays with CYP96T1 were analyzed with a chiral-CBH column by LC-MS/MS. Chromatographic separation of (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine standards was achieved preceding MS/MS analysis. Equal amounts of each enantiomer were observed (Figure 3.6A). A mass spectrometric comparison of standards (Figure 3.6B and C) and enzymatically formed (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine (Figure 3.6D and E) yielded identical MS/MS fragmentation patterns. The enzyme is, therefore, producing both (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine. A minor *N*-demethylnarwedine product was also detected in assays analyzed by HPLC on the Luna C8 column. The relative quantity of (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine and *N*-demethylnarwedine formed in assays with CYP96T1 are quantified in Figure 3.7A and B. HPLC was used to measure the relative contribution of these compounds to total product. (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine account for ~99 % of the total product in CYP96T1 assays. (10b*S*,4a*R*)- and/or (10b*R*,4a*S*)-noroxomaritidine and *N*-demethylnarwedine are also produced in assays containing only Sf9 cells and 4'-*O*-methylnorbelladine, but not in an enzyme-free control, indicating Sf9 cells have the ability to catalyze the *C*-*C* phenol couple with 4'-*O*-methylnorbelladine (Figure 3.5 A). In addition, the *N*-methylated form of 4'-*O*-methylnorbelladine, 4'-*O*-methyl-*N*-methylnorbelladine, was shown to produce several *C*-*C* phenol-coupled products when assayed with Sf9 cells alone, as indicated by the detection of products with a mass reduction of 2 *m/z*, including narwedine and

two unknown products (Figure 3.5B and 3.7D). Unknown 1 is enzymatically produced from 4'-*O*-methyl-*N*-methylnorbelladine by CYP96T1, as indicated by the increase of product in assays containing CYP96T1 as compared to the CPR-only control (Figure 3.5B). Unknown 2 production can be explained by the endogenous activity of Sf9 cells only expressing CPR on 4'-*O*-methylnorbelladine (Figure 3.7E). These observations were confirmed by an MRM-based relative quantification of selected transitions of these three products (Figure 3.7C, D, and E). The LC-MS/MS fragmentation pattern of unknown 1 is a mixture of masses found in the *para'-para* products (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine (165.1 *m/z*, 184.2 *m/z*, 195.0 *m/z*, 212.2 *m/z*, 229.0 *m/z*) and masses +14 *m/z* (120.1 *m/z*, 149.1 *m/z*, 243.2 *m/z*, 258.1 *m/z*, 271.0 *m/z*), representing the addition of a methyl moiety (Figure 3.5E). For this reason, it appears the enzyme is capable of catalyzing formation of the *para-para' C-C* phenol-couple regardless of *N*-methylation state (Figure 3.7A and C). To examine the ability of CYP96T1 to *C-C* phenol couple substrates with an altered carbon linker between the phenol groups, (*S*)-coclaurine and (*R*)-coclaurine were also tested. Assays on ether (*S*)-coclaurine or (*R*)-coclaurine yield products with a mass -2 m/z, which is consistent with a *C-C* phenol coupling. Product formation is not observed when norbelladine or *N*-methylnorbelladine is used as substrate. These results indicate the 4'-*O*-methylation state of norbelladine may be important for substrate-enzyme binding. The substrates 3'-*O*-methylnorbelladine and 3',4'-*O*-dimethylnorbelladine were tested to determine the relevance of 3'-*O*-methylation; products were not detected (Table 3.5).

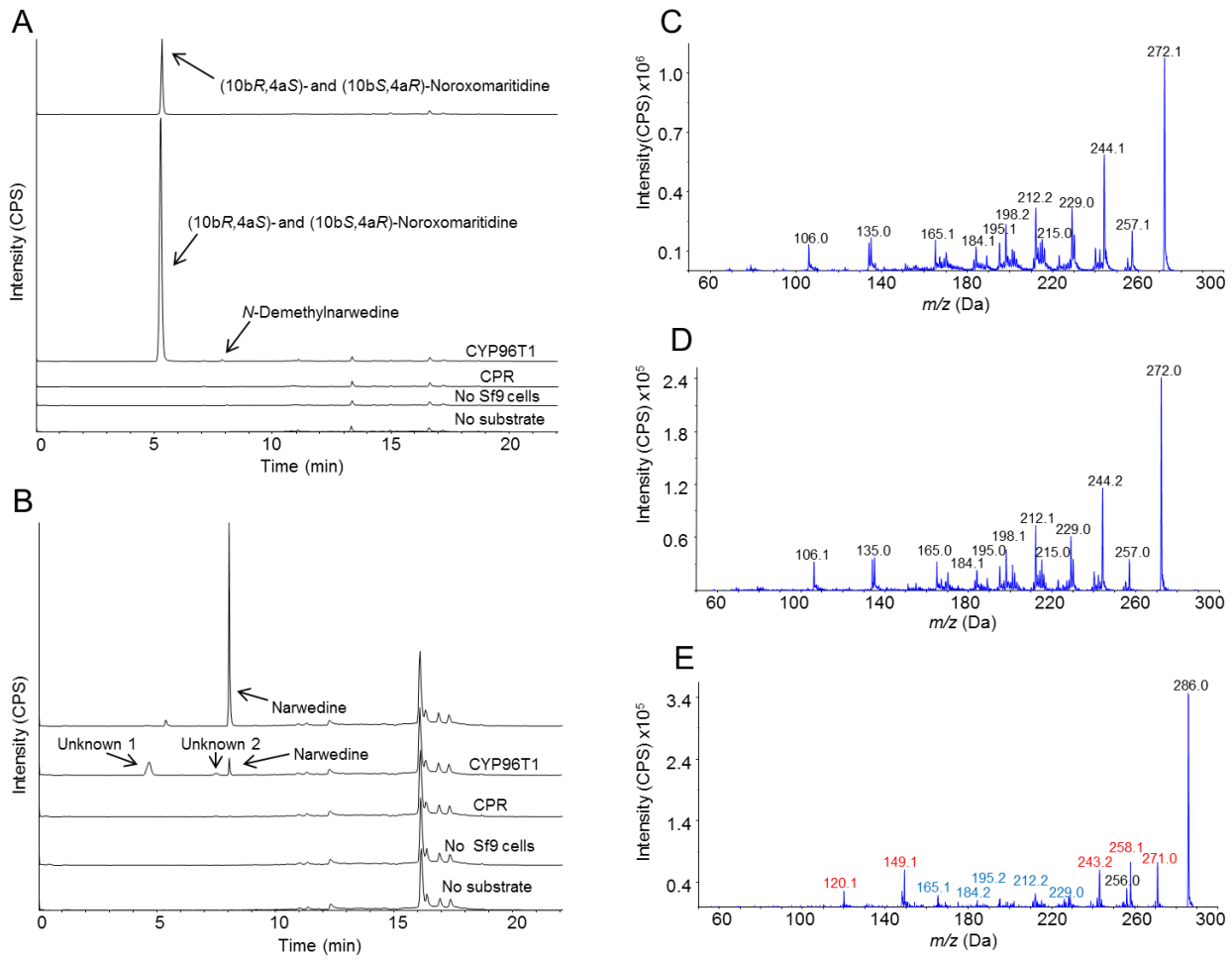

Figure 3.5. LC-MS/MS enhanced product ion scan (EPI) monitoring the *C-C* phenol coupling of 4'-*O*-methylnorbelladine and 4'-*O*-methyl-*N*-methylnorbelladine in CYP96T1 assays. Arrows indicate peaks unique to Sf9 cell containing assays with substrate present. (A) Standards and assays with 4'-*O*-methylnorbelladine as the substrate. Sample runs top to bottom (10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine standard (1 µM), CYP96T1 assay, CPR assay, CYP96T1 assay without 4'-*O*-methylnorbelladine, and assay without Sf9 cells. (B) Standards and assays with 4'-*O*-methyl-*N*-methylnorbelladine as the substrate. Top to bottom narwedine standard, CYP96T1 assay, CPR assay, assay without 4'-*O*-methylnorbelladine, and assay without Sf9 cells. (C) EPI of the (10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine standard. (D) EPI of the CYP96T1 (10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine product with 4'-*O*-methylnorbelladine as substrate. E. EPI of the CYP96T1 *para-para'* product (Unknown 1) with 4'-*O*-methyl-*N*-methylnorbelladine as substrate. Red fragments indicate the addition of one methyl group, 14 *m/z*, relative to (10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine and blue fragments indicate the same *m/z* as (10bR,4aS)- and (10b*S*,4a*R*)-noroxomaritidine fragments. Intensity is presented in counts per second (CPS).

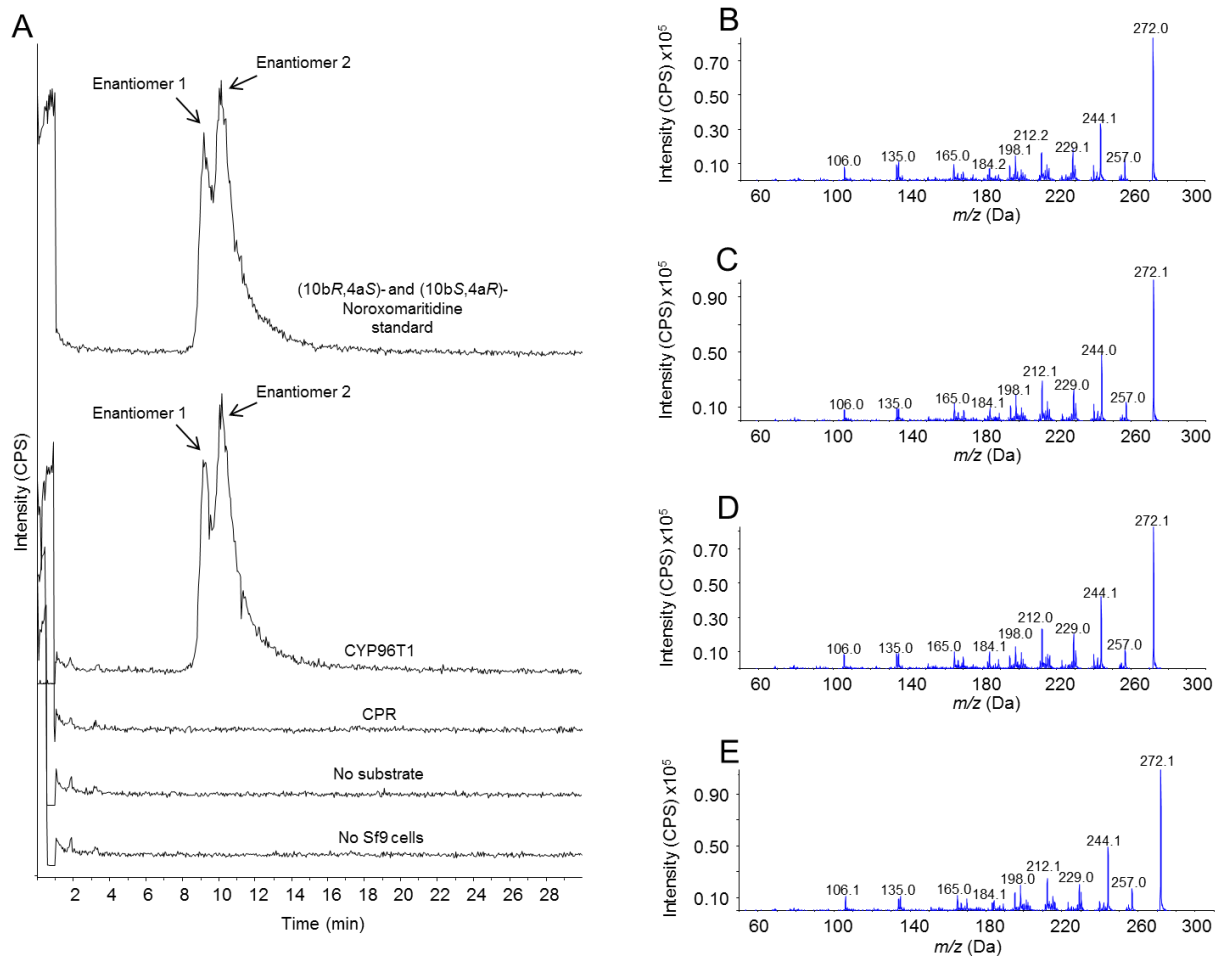

Figure 3.6. Chromatographic separation and MS/MS analysis of the primary 4'-*O*-methylnorbelladine products (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine. The enantiomers (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine were chromatographically separated with a chiral-CBH column and analyzed by MS/MS using an enhanced product ion (EPI) scan. (A) Samples, top to bottom: (10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine standard, CYP96T1 assay, CPR assay, CYP96T1 assay without 4'-*O*-methylnorbelladine substrate and no Sf9 cells assay. (B) EPI fragmentation pattern for enantiomer 1 of (10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine. (C) EPI fragmentation pattern for enantiomer 2 of (10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine. (D) EPI fragmentation pattern for enantiomer 1 in the CYP96T1 assay with 4'-*O*-methylnorbelladine as substrate. (E) EPI fragmentation pattern for enantiomer 2 in the CYP96T1 assay with 4'-*O*-methylnorbelladine as substrate. Intensity is presented in counts per second (CPS).

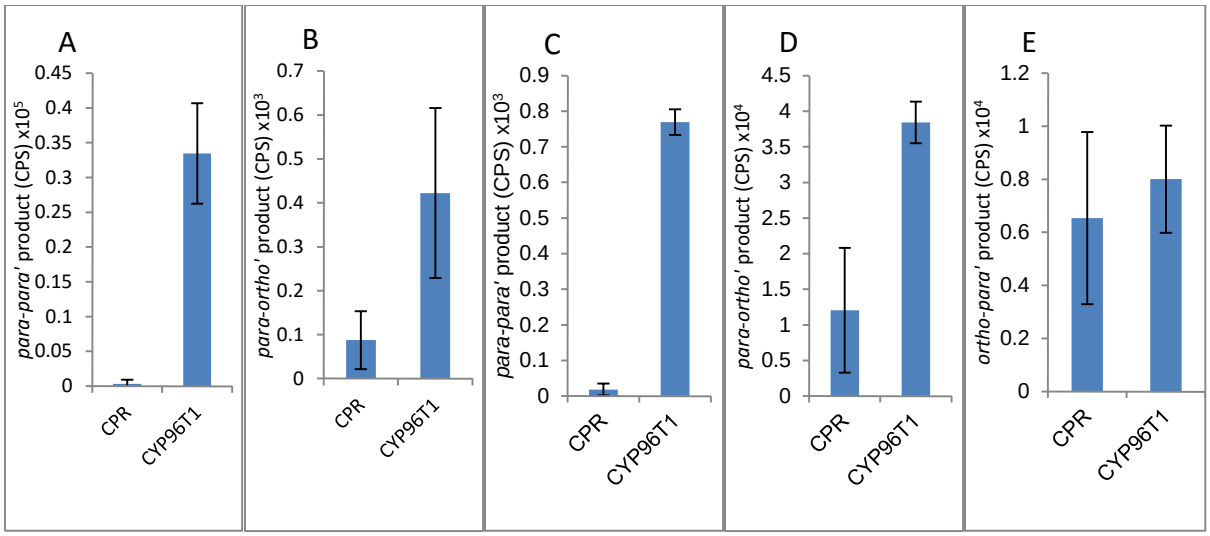

Figure 3.7. Relative product formed in assays with 4'-*O*-methylnorbelladine (A and B) or 4'-*O*-methyl-*N*-methylnorbelladine (C, D, and E) as substrate. Assays are performed in triplicate only expressing *CPR* or with *CPR* in combination with *CYP96T1*. (A) *para-para'* ((10b*R*,4a*S*)- and (10b*S*,4a*R*)-noroxomaritidine) product. (B) *para-ortho'* (*N*-demethylnarwedine) product. (C) Potentially *para-para' C-C* phenol coupling (unknown 1) product. (D) *para-ortho'* (Narwedine) product. (E) Potentially *ortho-para' C-C* phenol coupling (unknown 2) product.

The K_m of (*S*)-coclaurine, 636.7 μM, and (*R*)-coclaurine, 658.8 μM, are several orders of magnitude higher than the K_m values for 4'-*O*-methylnorbelladine, 1.13, and 4'-*O*-methyl-*N*-methylnorbelladine, 3.28 (Table 3.5). Substrate inhibition was observed in 4'-*O*-methylnorbelladine and 4'-*O*-methyl-*N*-methylnorbelladine with K_i values of 64.34±26.36 μM and 173.7±140.0 μM respectively. No substrate inhibition was observed in (*R*)-coclaurine or (*S*)-coclaurine with concentrations up to 1000 μM. The k_{cat} of 4'-*O*-methylnorbelladine was higher than observed for 4'-*O*-methyl-*N*-methylnorbelladine, (*R*)-coclaurine, or (*S*)-coclaurine. The k_{cat}/K_m value 4'-*O*-methylnorbelladine is at least one order of magnitude larger than the 4'-*O*-methyl-*N*-methylnorbelladine, (*R*)-coclaurine, or (*S*)-coclaurine values consistent with the role of 4'-*O*-methylnorbelladine as the native substrate.

Table 3.5. Substrate specificity tests for CYP96T1

Substrate	K_m (μM)	k_{cat} (1/min)	k_{cat}/K_m (1/μM*min)	K_i	Modifications monitored
4'-*O*-Methylnorbelladine	1.13±0.54	15.0±2.03	13	64.3±26.4	*C-C* phenol coupling
4'-*O*-Methyl-*N*-methylnorbelladine	3.28±2.27	2.44±0.54	0.742	174±140	*C-C* phenol coupling
(*S*)-Coclaurine	637±156	1.34±0.15	$2.11x10^{-3}$	NA	Intramolecular phenol coupling and Intermolecular coupling
(*R*)-Coclaurine	659±104	2.07±0.14	$3.15x10^{-3}$	NA	Intramolecular phenol coupling and Intermolecular coupling
3'-*O*-Methylnorbelladine	NA	ND	NA	NA	*C-C* phenol coupling
3',4'-*O*-Dimethylnorbelladine	NA	ND	NA	NA	*C-C* phenol coupling
Norbelladine	NA	ND	NA	NA	*C-C* phenol coupling
N-Methylnorbelladine	NA	ND	NA	NA	*C-C* phenol coupling
Haemanthamine	NA	ND	NA	NA	Methoxy bridge formation and hydroxylation
(10b*S*,4a*R*)- and (10b*R*,4a*S*)-Noroxomaritidine	NA	ND	NA	NA	Methoxy bridge formation and hydroxylation
Isovanillin and tyramine	NA	ND	NA	NA	*C-C* phenol coupling, amine-aldehyde condensation, amine-aldehyde condensation, and *C-C* phenol coupling

ND=not detected
NA=not applicable

3.3.4 Sodium Borohydride Assays and Analysis by LC-MS/MS

Enzymatically formed *N*-demethylnarwedine from enzyme assays with CYP96T1 was converted to *N*-demethylgalanthamine by sodium borohydride reduction and detected by LC-MS/MS (Figure 3.8A). Sodium borohydride selectively reduced the ketone group on (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine and *N*-demethylnarwedine to yield a stereoisomeric mixture of the corresponding alcohols 8-*O*-demethylmaritidine and *N*-demethylgalanthamine. Confirmation of *N*-demethylgalanthamine in these assays is demonstrated by the identical retention time (Figure 3.8A,) and fragmentation pattern (Figure 3.8B and C) with *N*-demethylgalanthamine standard.

Another peak is also present with a different retention time (Figure 3.8A) and very similar fragmentation pattern (Figure 3.8D) and is likely the diastereomer *epi*-*N*-demethylgalanthamine formed by non-stereospecific ketone reduction. Stereoisomeric 8-*O*-demethylmaritidine is present in sodium borohydride reduced CYP96T1 4'-*O*-methylnorbelladine assays as the largest product peak (Figure 3.8A). This is validated by a comparison of the LC-MS/MS fragmentation pattern of (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine reduced by sodium borohydride to the corresponding peak in the CYP96T1 assay (Figure 3.8E and F).

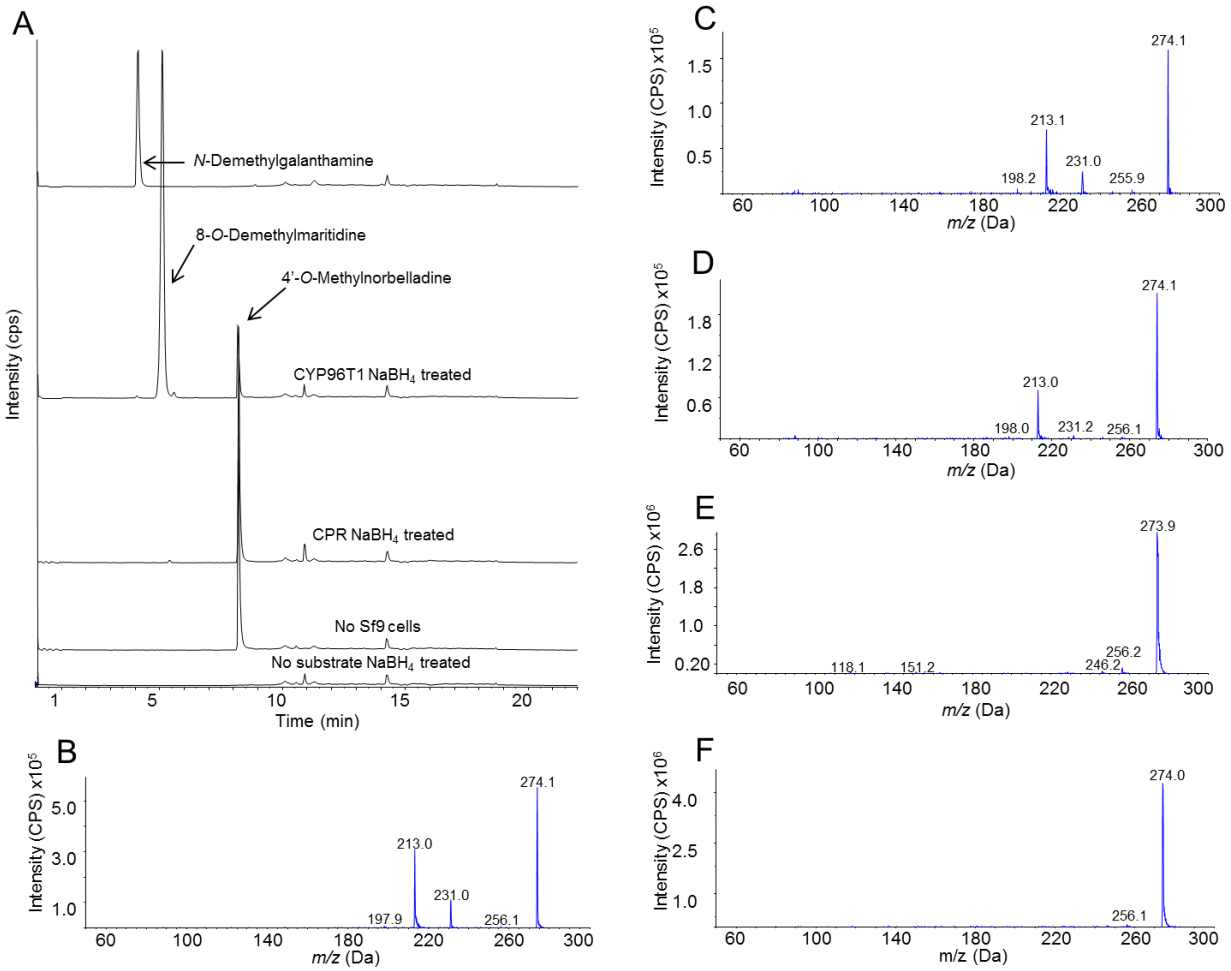

Figure 3.8. LC-MS/MS Enhanced Product Ion (EPI) scan of sodium borohydride ($NaBH_4$) treated CYP96T1 assays with 4'-*O*-methylnorbelladine substrate. (A) Chromatograph with the following sample runs top to bottom: *N*-demethylgalanthamine standard, CYP96T1 assay, CPR assay, assay with no Sf9 cells and CYP96T1 assay without 4'-*O*-methylnorbelladine. (B) EPI fragmentation pattern of the *N*-demethylgalanthamine standard peak eluting at 4 min. (C) EPI fragmentation pattern of the *N*-demethylgalanthamine product in the CYP96T1 assay. (D) EPI

fragmentation pattern of *epi*-*N*-demethylgalanthamine from the CYP96T1 assay. E. EPI fragmentation pattern of (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine standard reduced to stereoisomeric 8-*O*-demethylmaritidine. F. EPI fragmentation pattern of reduced (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine product from CYP96T1 assays.

3.4 Discussion

CYP96T1 converts 4'-*O*-methylnorbelladine to the products (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine indicating that this enzyme is involved in the biosynthesis of (10b*R*,4a*S*)-noroxomaritidine-derived alkaloids such as haemanthamine. Because (10b*S*,4a*R*)-noroxomaritidine derivatives have not been previously reported from *Narcissus* spp., the enantiomeric mixture of (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine made by CYP96T1 is interesting. It is possible the CYP96T1 enzyme is only making the achiral intermediate that later spontaneously forms the different enantiomeric forms of noroxomaritidine. If this is the case the absence of the (10b*S*,4a*R*)-noroxomaritidine derivatives in *Narcissus* spp. may result from another enzyme perhaps associated with CYP96T1 directing the chirality of the ring closure.

The production of *N*-demethylnarwedine by CYP96T1 is of interest to galanthamine biosynthesis. The low amount produced relative to (10b*S*,4a*R*)- and (10b*R*,4a*S*)-noroxomaritidine indicates that under the assay conditions used *N*-demethylnarwedine is not the enzyme's primary product. Kinetic analysis shows a clear preference for 4'-*O*-methylnorbelladine over all other tested substrates (Table 3.5).

A diradical mechanism has been proposed for formation of the *C*-*C* phenol coupled product of (*R*)-reticuline and 4'-*O*-methylnorbelladine (Eichhorn et al., 1998; Grobe et al., 2009) (Figure 3.9A and B). A radical is formed on a hydroxyl group *ortho* or *para* to the position for formation of a carbon bond. To determine if the 3'(*para'*) hydroxyl group is important to *C*-*C* phenol coupling, 3'-*O*-methylnorbelladine and 3',4'-*O*-dimethylnorbelladine were tested for enzymatic activity; product formation was not observed. The lack of activity with a methoxy group at the

para' position indicates that a free hydroxyl moiety is important at this position to enable extraction of a hydroxyl radical by the enzyme (Figure 3.9A). These results support the proposed mechanism for *C*-*C* phenol coupling of 4'-*O*-methylnorbelladine.

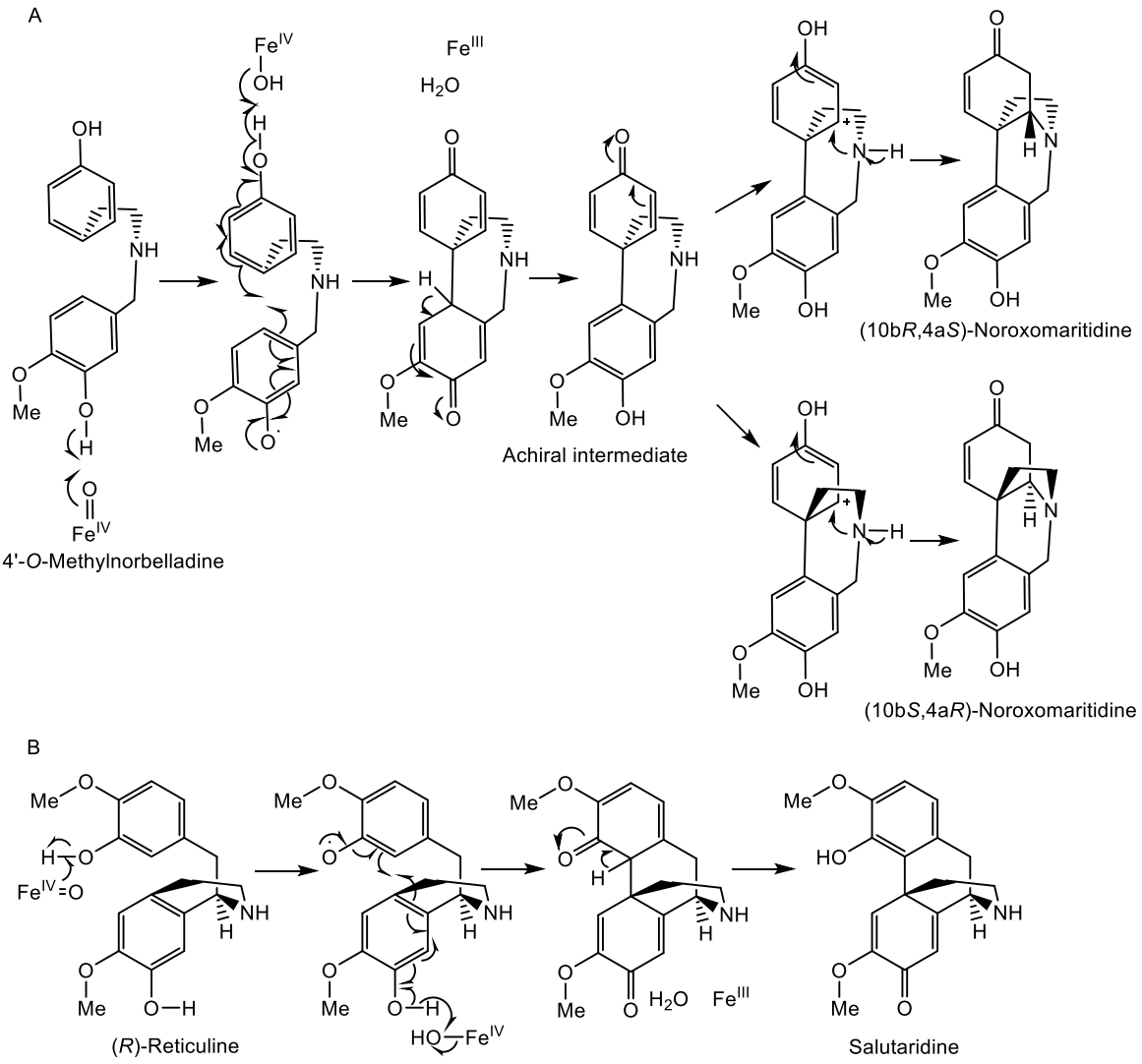

Figure 3.9. Proposed *C*-*C* phenol coupling mechanisms. (A) 4'-*O*-Methylnorbelladine *para-para' C-C* phenol coupling mechanism followed by spontaneous nitrogen ring closure to form noroxomaritidine. (B) (*R*)-Reticuline *para-ortho' C-C* phenol coupling mechanism to form salutaridine panel adapted from (Grobe et al., 2009).

The oxygen binding and activation motif (A/G)GX(D/E)TT is substantially different between CYP96T1 (and its variants) when compared to cytochrome P450 enzymes that catalyze hydroxylation reactions. The substitutions G322Q, D324N and T326Q replace key hydrophobic,

acidic and alcoholic groups with neutral hydrophilic amide groups. This is consistent with the proposal of Mizutani and Sato that cytochrome P450 enzymes not performing hydroxylation reactions can have a significantly altered oxygen binding and activation site (Mizutani and Sato, 2011). The highly conserved (387-389) EXXR, PERF (430-433) PXRX, and heme binding (464-473) FXXGXRXCXG motifs are present (Syed and Mashele, 2014). These motifs are thought to have more universal functions than substrate hydroxylation including maintenance of proper structural integrity and heme placement in cytochrome P450 enzymes (Hasemann et al., 1995; Hatae et al., 1996).

Presented herein is the first documented *C-C* phenol coupling cytochrome P450 enzyme in monocots. It is in the CYP96 family of cytochrome P450 enzymes, which falls into the CYP86 clan. The CYP96A15 from *A. thaliana* has been previously documented to be a midchain alkane hydroxylase involved in wax synthesis (Greer et al., 2007). Previously documented members of the CYP86 clan have shown activity towards fatty alcohols, fatty acids, alkanes, and derivatives thereof (Nelson and Werck-Reichhart, 2011). This makes this phenolic alkaloid a novel substrate class for this clan of cytochrome P450 enzymes. All other documented *C-C* phenol coupling plant cytochrome P450 enzymes are in the CYP71 clan (Nelson and Werck-Reichhart, 2011). This indicates the *C-C* phenol coupling activity of CYP96T1 was acquired independently from other known *C-C* phenol coupling cytochrome P450 enzymes. This independent origin of *C-C* phenol coupling could help direct the search for new *C-C* phenol coupling cytochrome P450 enzymes. The independent evolution of CYP96T1 shows that future searches for novel *C-C* phenol coupling enzymes should look broadly across the cytochrome P450 families because lineages of cytochrome P450 enzymes responsible for these reaction activities have likely not all been identified. Other phenol-phenol coupling reactions potentially performed by cytochromes P450 include the

intramolecular coupling of 4'-*O*-methylnorbelladine to oxonorpluvine in lycorine biosynthesis, (*S*)-autumnaline to isoandrocymbine in colchicine biosynthesis, and the intermolecular *C*-*C* phenol coupling of dioncophylline A biosynthesis (Bringmann et al., 2000; Herbert, 2003).

3.5 Acknowledgments

This work was supported by the National Institutes of Health award number 1RC2GM092561 (NIGMS). This material is based upon work supported by the National Science Foundation under Grant No. DBI-0521250 for acquisition of the QTRAP LC-MS/MS. *G. elwesii* collection and identification was done thanks to Laurence Davin from Washington State University. Thank you to J. Steen Hoyer for critiquing this manuscript.

Chapter 4: The Identification of a Noroxomaritidine/Norcraugsodine Reductase in the Core Amaryllidaceae Alkaloid Biosynthetic Pathway

Amaryllidaceae alkaloids are a large group of plant natural products with over 300 documented structures and diverse biological activities. An approved medication from this group is the Alzheimer's treatment drug galanthamine. The first committed steps in the core Amaryllidaceae alkaloid biosynthetic pathway are the condensation of 3,4-dihydroxybenzaldehyde and tyramine to form the Schiff-base norcraugsodine and subsequent reduction of the imine to norbelladine. In this study, oxidoreductase genes that co-express with the previously discovered norbelladine 4'-O-methyltransferase gene in *Narcissus* sp. and *Galanthus* spp. were cloned and expressed in *E. coli*. One of these genes named noroxomaritidine/norcraugsodine reductase encodes an enzyme that forms norbelladine from a mixture of 3,4-dihydroxybenzaldehyde and tyramine. Considering the ability of norcraugsodine to form spontaneously in solution, the protein was crystallized and shown to bind tyramine and an analogue of 3,4-dihydroxybenzaldehyde, piperonal, individually. This binding indicates the enzyme may be able to facilitate norcraugsodine formation by binding the substrates, thereby providing a scaffold for the Schiff-base condensation prior to imine bond reduction.

4.1 Introduction

Norbelladine is an intermediate in the biosynthesis of all Amaryllidaceae alkaloids (Battersby et al., 1961b; Battersby et al., 1961a; Wildman et al., 1962c). norbelladine can act as a reactive

oxygen species scavenger and suppress enzymes associated with inflammation including CAS1, CAS2, and NF-κB making it a potential drug for mitigating the effects of obesity (Park, 2014). It is synthesized from the condensation and reduction of 3,4-dihydroxybenzaldehyde and tyramine (Figure 4.1). Tyrosine decarboxylase catalyzes the formation of tyramine from tyrosine (Lehmann and Pollmann, 2009). 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-hydoryxbenzaldehyde (Suhadolnik et al., 1963a). 3,4-Dihydroxycinnamic acid could be directly converted to 3,4-dihydroxybenzaldehyde by a paralogue of vanillin synthase. Vanillin synthase converts ferulic acid to vanillin (Gallage et al., 2014). This reaction is similar to that proposed for the reduction of 3,4-dihydroxycinnamic acid to 3,4-dihydroxybenzaldehyde by a hydratase/lyase type mechanism. An analogous, non-oxidative pathway is proposed for benzaldehyde biosynthesis. Reactions similar to the beta-oxidative and non-oxidative CoA-dependent pathways of benzaldehyde biosynthesis are also a possibility (Widhalm and Dudareva, 2015). The condensation to the Schiff-base norcraugsodine is analogous to known amine-aldehyde condensing enzymes including norcoclaurine synthase and strictosidine synthase (Kutchan et al., 1994; Samanani and Facchini, 2002). Considering the simple chemistry required for the formation of norcraugsodine, it is possible that the condensation occurs either nonenzymatically or in the active site of the same enzyme that reduces norcraugsodine to norbelladine. Enzyme superfamilies capable of catalyzing reductions similar to the reduction of norcraugsodine to norbelladine include aldo-keto reductases (AKRs) and short-chain dehydrogenase/reductases (SDRs) (Jörnvall et al., 1995; Penning, 2015). An example of an imine reductase from the alcohol dehydrogenase (ADH) branch of the SDR

superfamily is tetrahydroalstonine synthase from *Catharanthus roseus*, which reduces strictosidine aglycone to tetrahydroalstonine (Stavrinides et al., 2015).

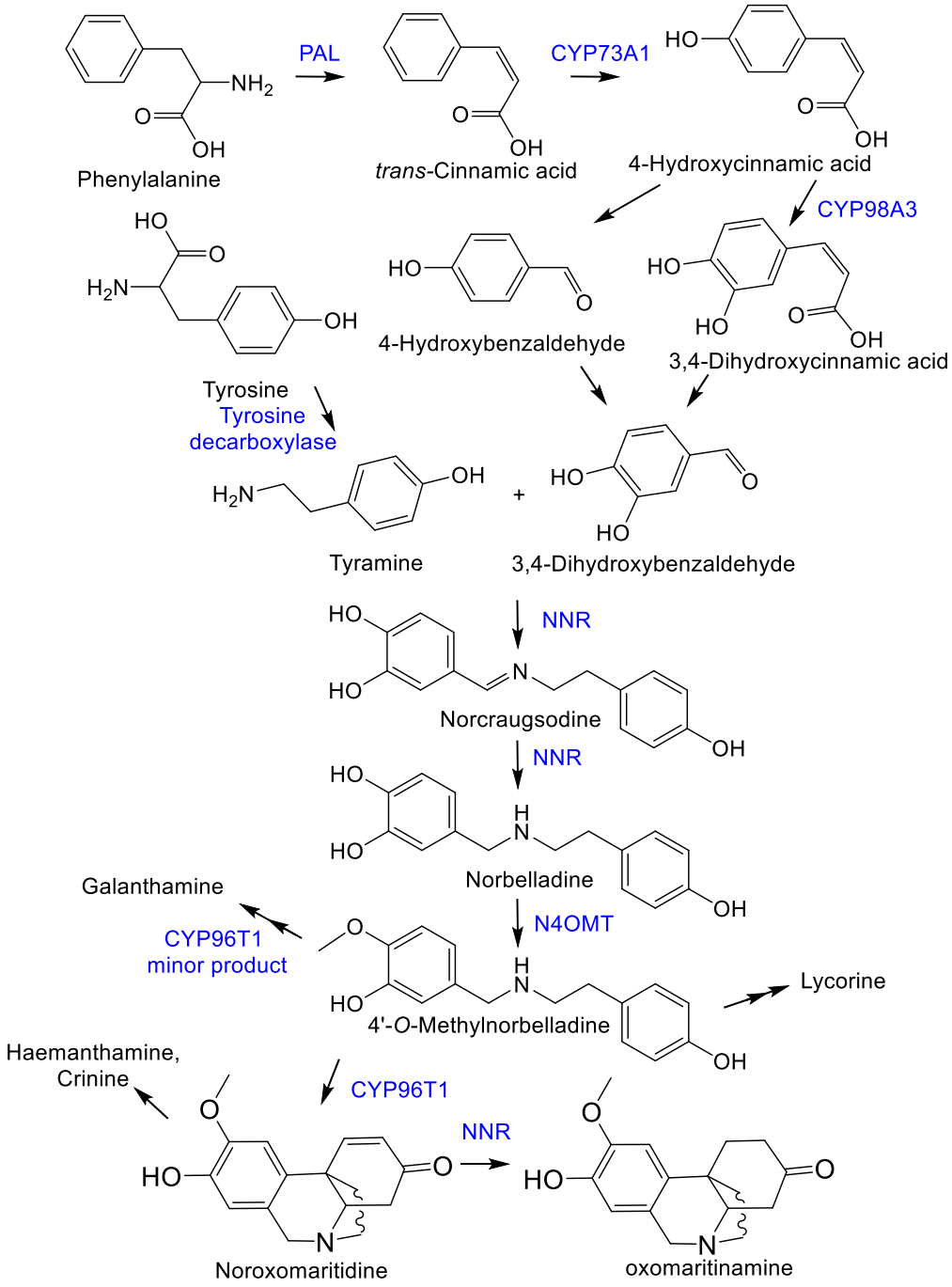

Figure 4.1. Amaryllidaceae alkaloid biosynthesis. Enzymes involved in the pathway are indicated in blue. Abbreviations phenylalanine ammonia lyase (PAL), norbelladine 4'-*O*-methyltransferase (N4OMT), and noroxomaritidine/norcraugsodine reductase (NNR).

Norbelladine is methylated and phenol-coupled to produce the vast diversity of Amaryllidaceae alkaloids, including the representative alkaloids for the primary skeletons haemanthamine, crinine, galanthamine, and lycorine. Several downstream Amaryllidaceae alkaloid biosynthetic enzymes have been characterized on a molecular level including norbelladine 4'-*O*-methyltransferase (N4OMT) and CYP96T1. N4OMT is responsible for the 4'-*O*-methylation of norbelladine to 4'-*O*-methylnorbelladine (Kilgore et al., 2014). CYP96T1 is a *para-para'* phenol-phenol coupling cytochrome P450 that generates an enantiomeric mixture of noroxomaritidine with a small quantity of *N*-demethylnarwedine from 4'-*O*-methylnorbelladine (Chapter 3). The two enantiomers of noroxomaritidine are further modified into haemanthamine- and crinine-type alkaloids. *N*-demethylnarwedine from CYP96T1 may contribute a small amount to the galanthamine type alkaloids (Figure 4.1). Several of these compounds are interesting to medicine including the anticancer compound haemanthamine and the Alzheimer's disease treatment galanthamine (Wilcock et al., 2003; Havelek et al., 2014). These gene discoveries facilitate co-expression analysis to identify additional Amaryllidaceae alkaloid biosynthetic genes.

In this study, previously described in Chapters 2 and 3 *Narcissus* sp. *aff. pseudonarcissus*, *Galanthus* sp., and *Galanthus elwesii* transcriptomes are interrogated for reductase genes that consistently co-express with the previously discovered *N4OMT* in 4 or more of these transcriptomes (Kilgore et al., 2014). These reductase cDNAs were cloned, expressed in *E. coli*, and the proteins affinity purified. Assays monitoring the formation of norbelladine from a solution containing 3,4-dihydroxybenzaldehyde and tyramine show formation of norbelladine with the SDR, which we designated noroxomaritidine/norcraugsodine reductase (NNR). Competition experiments and crystal structures indicate this enzyme acts directly on the 3,4-dihydroxybenzaldhyde and tyramine and acts as a norcraugsodine synthase. In addition, the

enzyme reduces a carbon-carbon double bond of the Amaryllidaceae alkaloid noroxomaritidine to form one of the enantiomeric forms of oxomaritinamine. This enzyme is responsible for the first committed biosynthetic step in the Amaryllidaceae alkaloids and will be key to future biological production of these compounds.

4.2 Materials and Methods

4.2.1 Plant Tissue and Chemicals

Plant tissue and chemicals are as described for isolation and characterization of N4OMT and CYP96T1 in Chapters 2 and 3 (Kilgore et al., 2014). In addition, the following chemicals were used: piperonal from Sigma; 8,9-*O*-didemethyloxomaritidine and noroxomaritidine were from our natural product collection; DMSO from New England BioLabs; nicotinamide adenine dinucleotide phosphate disodium salt ($NADP^+$) from MP Biomedicals, Inc; citric acid anhydrous reagent ACS 99.5% from ACROS; sodium borohydride 99% from Acros Organics; and sodium cyanoborohydride reagent grade 95 % from Aldrich.

4.2.2 Candidate Gene Identification

Candidates were selected as in the discovery of CYP96T1 (Chapter 3), but instead of a list of cytochromes P450, a list of oxidoreductases were used as a BLASTP query (Table D.1 and Table D.2). Contigs found to co-express with *N4OMT* in 4 or more of the 5 available Amaryllidaceae transcriptomes were cloned and tested for enzymatic activity (Chapter 3).

4.2.3 cDNA Cloning and Recombinant Enzyme Purification

cDNA was prepared as previously described in Chapter 3. The outer PCR reaction contained the following components: 25 ng *Narcissus* sp. *aff. pseudonarcissus* bulb cDNA, 1X Phusion HF reaction buffer, 1 U NEB Phusion High-Fidelity DNA polymerase, 3% DMSO, 0.4 mM dNTPs, 0.4 µM NNR forward outer, and 0.4 µM NNR reverse outer primers (Table D.3). The parameters

for the PCR were 98 ^{0}C 30 sec for 1 cycle, 98 ^{0}C 10 sec, 50 ^{0}C 30 sec, 72 ^{0}C 60 sec for 35 cycles, 72 ^{0}C 5 min 1 cycle, 4 ^{0}C until removed. The inner PCR solution was the same except 0.2 µM NNR forward inner and 0.2 µM NNR reverse inner primers were used and the outer PCR was used as template instead of cDNA (Table D.3). The inner PCR program had a T_m of 52 ^{0}C instead of 50 ^{0}C and 25 cycles instead of 35 cycles. Primers for inner PCR have BamHI and NotI restriction sites appended to the start and stop sites of the open reading frame. These were used for ligation into pET28a and subsequent transformation into *E. coli*. Expression and subsequent affinity purification of hexahistidine-tagged protein was as for *N4OMT* (Chapter 2).

4.2.4 Enzyme Assays

Initial screening assays contained 100 mM sodium phosphate buffer pH 7.0, 1 mM NADPH, 1 mM tyramine, 1 mM 3,4-dihydroxybenzaldehyde, and 10 µg pure protein in 100 µl. Reactions were incubated at 30 ^{0}C for 2 hr. Assays were extracted with ethyl acetate at pH 9.5 as previously described (Kilgore et al., 2014). The extracts were re-suspended in mobile phase matching the solvent composition at the beginning of the HPLC program. Samples were run on the same LC-MS/MS instrument and with the same program as for CYP96T1 (Chapter 3). MS/MS parameters used to specifically monitor norbelladine are collision energy (CE) 15, declustering potential (DP) 50, and *m/z* 260.0. Other compounds screened were tested with the parameters listed in Table D.4. IC_{50} and specific activity assays were performed with 100 mM citrate buffer pH 6.0 at 35 ^{0}C in 40 µl reactions. Substrates used in specific activity assays were all 500 µM with 1 mM NADPH. For temperature optima and pH optima determination, the MRM parameters used to monitor norbelladine were *m/z* 260.0/138.0 and *m/z* 260.0/121.0 with the same CE and DP as for screening. For IC_{50} assays, oxomaritinamine was monitored with the following MRMs *m/z* 274.3/136.1 and *m/z* 274.3/219.1 with CE (35) and DP (70). For specific activity measurements, a QTRAP 6500

was used for MRM analysis with the same isocratic LC set up as previously described (Chapter 3). The MRM parameters for specific activity measurements were norbelladine *m/z* 260.1/238.0 and *m/z* 260.1/121.0 CE (60) and DP (10), 4'-*O*-methylnorbelladine *m/z* 274.1/137.0 and *m/z* 274.1/122.0 CE (20) and DP (60), and oxomaritinamine *m/z* 274.1/136.1 and *m/z* 274.1/219.1 CE (30) and DP (70). Oxomaritinamine was quantified by incubating noroxomaritidine and NNR overnight and equating the quantity of noroxomaritidine consumed to the quantity of oxomaritinamine produced using the same HPLC program used to analyze enzyme assays during *N4OMT* characterization at 288 nm (Kilgore et al., 2014). This standardized oxomaritinamine was then used in specific activity experiments on the LC-MS/MS to determine product quantity.

4.2.5 Substrate Synthesis

4-(2-((1,3-Benzodioxol-5-ylmethyl)amino)ethyl)phenol, 3'-*O*-methylnorbelladine, and 4'-*O*-methylnorbelladine were made by mixing ~100 mM sodium cyanoborohydride and ~10 mM tyramine with ~10 mM piperonal, vanillin, or isovanillin respectively in 2.5 ml anhydrous methanol. After incubation overnight at room temperature the reactions were dried to 0.5 ml with subsequent addition of 2 ml 1 M sodium carbonate pH 9.5 and extracted with 2 ml ethyl acetate two times. After drying, all extracts were dissolved in 1 ml water, diluted 1/20 and purified by fraction collection with the same HPLC program used to analyze enzyme assays during *N4OMT* characterization (Kilgore et al., 2014).

4.2.6 Protein Expression and Purification for Crystallography

NNR in pET28a was transformed into *E. coli* Rosetta II (DE3) cells (EMD Millipore). Cells were cultured in Terrific broth up to A_{600nm}~0.6-0.8. Induction of protein production used a final concentration of 1 mM IPTG overnight at 16^0C. Cells were pelleted by centrifugation and resuspended in lysis buffer (50mM Tris, pH 8.0, 500 mM NaCl, 20 mM imidazole, 10% glycerol,

and 1% Tween). Following sonication, cell debris were removed by centrifugation, and the resulting lysate was passed over a Ni^{2+}-nitriloacetic acid (Qiagen) 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. Prior to crystallization the His-tag was removed by overnight dialysis at 4 ^{0}C using thrombin (1:2000 total protein). Dialyzed protein was loaded onto a mixed Benzamidine-Sepharose/Ni^{2+}-nitrilotriacetic acid column. The flow-through 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.

4.2.7 Protein Crystallography

Purified NNR 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 buffer). Diffraction quality crystals of the NNR·$NADP^+$ complex were obtained at 4 ^{0}C with a crystallization buffer of 20% PEG-8000 and 100 mM Hepes, pH 7.5. Crystals of the NNR·tyramine complex formed at 4 ^{0}C with a crystallization buffer of 2 M ammonium sulfate, 100 mM CAPS, pH 10.5, 200 mM lithium sulfate, 1.5 mM $NADP^+$, and 6 mM tyramine. For the NNR·piperonal complex, 35% MPD, 100 mM sodium acetate, pH 4.5 and 3 mM piperonal was the crystallization buffer. Crystals were flash-frozen in liquid nitrogen the 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 implemented in PHASER (de La Fortelle and Bricogne, 1997) using *Datura stramonium* tropioine reductase-II (PDB: 2AE2; (Yamashita et

al., 1999) as a search model to determine the structure of the NNR·tyramine complex. Iterative rounds of manual model building and refinement, which included translation-libration-screen (TLS) models, used COOT (Emsley and Cowtan, 2004) and PHENIX (Adams et al., 2010). The crystal structures of NNR·$NADP^+$ and NNR·piperonal were determined by molecular replacement using PHASER (de La Fortelle and Bricogne, 1997) and the NNR·tyramine structure with ligands removed as the search model. Data collection and refinement data are summarized in Table D.5. The final model of the NNR·$NADP^+$ complex included residues Ser-16- Gly-271, the $NADP^+$ ligand, and 136 waters. The final model of the NNR·tyramine complex included residues Ser-16–Gly-271 for chain A, residues Met15–Gly-271 for chain B, residues Leu-17–Asn-270 for chain C, and residues Leu-17–Gly-271 for chain D, the $NADP^+$ ligand in chains A-C, the tyrosine ligand in chains B and D, and 443 waters. The final model of the NNR·piperonal complex included Met-15–Gly-271, the $NADP^+$ ligand, the piperonal ligand in chain A, and 644 waters.

4.3 Results

4.3.1 Identification and Initial Characterization of NNR

When interrogating for *N4OMT* co-expressing oxidoreductase homologues, 36 transcripts were identified in the *Narcissus* sp. *aff. Pseudonarcissus* transcriptome. Of these 36 transcripts, medp_9narc_20101112|58880 was found to co-express with *N4OMT* in the previously described ABySS and MIRA based assemblies for *Narcissus* sp. *aff. pseudonarcissus, Galanthus elwesii* and *Galanthus* sp. (Kilgore et al., 2014). It also co-expressed with *N4OMT* in the previously described Trinity based assembly for *Galanthus* sp. (Chapter 3). The medp_9narc_20101112|58880 was incomplete in the 3’ and 5’ regions. The medp_9narc_20101112|12438 sequence, a close homologue of apparent full-length, was used to design primers for PCR amplification. A clone with the genbank accession number KU295569 was obtained that differed from the

medp_9narc_20101112|58880 at four nucleic acids with one nonsynonymous change (see Figure D.1 for details).

Expression in *E. coli* followed by affinity purification on a cobalt column resulted in a prominent ~31 kDa band on SDS-PAGE. When this enzyme was tested for norcraugsodine reductase activity, it was found to produce norbelladine (Figure 4.2A). The identity of norbelladine was ascertained by comparison of the MS fragmentation pattern to that of authentic standard (Figure 4.2B and C). The gene was therefore named noroxomaritidine/norcraugsodine reductase (NNR). A norbelladine background is present in all assays with 3,4-dihydroxybenzaldehyde and tyramine due to spontaneous formation of norcraugsodine. The enzyme was found to be NADPH dependent. This is consistent with the prediction that this will be a NADPH-dependent SDR based on the residues Lys-21, Ser-42, and Arg-43 (Kallberg et al., 2002). NADH did not serve as reductant for NNR. To simplify analysis quantity of enzymatic product relative to chemical background was maximized for pH and temperature. This maximum was found to be at pH 6.0 and 35 ^{0}C (Figure D.2 A-D).

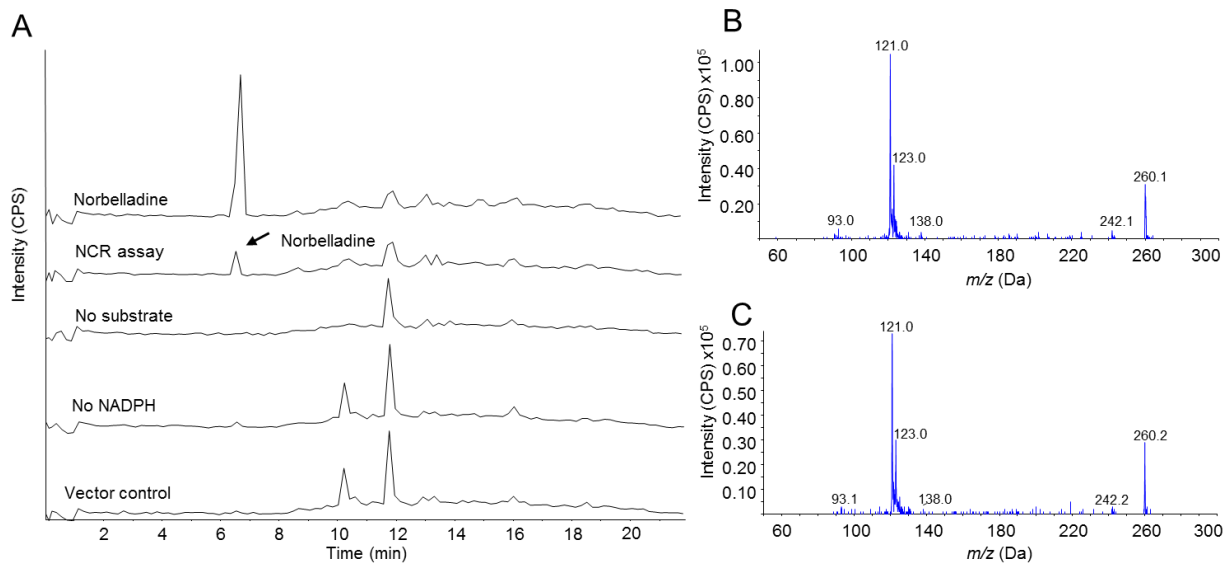

Figure 4.2. LC-MS/MS enhanced product ion scan (EPI) of the enzyme assays testing noroxomaritidine/norcraugsodine reductase (NNR) *m/z* 260.0. (A) Top to bottom Norbelladine standard, functioning NNR assay, assay without tyramine and 3,4-dihydroxybenzaldehyde, assay without NADPH, and assay with His-tag purified *E. coli* vector control protein extract but no NNR protein. (B) EPI of norbelladine standard. (C) EPI of NNR product.

4.3.2 Substrate Specificity

Substrate specificity tests identified vanillin, isovanillin, and piperonal as substrates. These substrates, paired with tyramine, generated the products 3'-*O*-methylnorbelladine, 4'-*O*-methylnorbelladine, and 4-(2-((1,3-benzodioxol-5-ylmethyl)amino)ethyl)phenol, respectively (Table 4.1). The structures of the enzymatic products were confirmed by comparing to the products produced by chemical condensation and reduction of vanillin, isovanillin, or piperonal with tyramine in methanol containing sodium cyanoborohydride and subsequent examination with LC-MS/MS. The resulting LC-MS/MS spectra were identical whether the products were chemically or enzymatically generated. To examine the ability of NNR to operate in other points of Amaryllidaceae alkaloid pathways noroxomaritidine was tested for reductase activity. This substrate *in planta* is thought to be reduced at a ketone group to yield the alcohol normaritidine that is further modified into haemanthamine or potentially at a 1-2 carbon-carbon double bond to

give oxomaritinamine. Noroxomaritidine was reduced by NNR (Figure 4.3A). When noroxomaritidine reduced by NNR was compared to noroxomaritidine reduced with sodium borohydride, the fragmentation patterns were not the same (Figure 4.3B and C). Aldehyde and ketone double bonds are the typical substrates of sodium borohydride (Banfi et al., 2001). Carbon-carbon double bonds can be reduced, but the ketone will be reduced as well. For this reason, the product generated by the sodium borohydride reduction is likely normaritidine. The 1-2 carbon-carbon double bond is, therefore, likely reduced by NNR into oxomaritinamine. The noroxomaritidine substrate is an enantiomeric mixture. The enantiomers were therefore resolved by chiral chromatography as previously described (Chapter 3). In overnight assays only one of the noroxomaritidine enantiomers was notably consumed. More than half of the enantiomeric mixture was never consumed, indicating that the enzyme acts preferentially on one of the two enantiomers. Optically pure standards of known configuration are lacking, so the absolute configuration of the preferred substrate remains unknown. To examine the ability of NNR to take different substrates like noroxomaritidine the demethylated form 8,9-*O*-didemethyloxomaritidine was tested for enzymatic activity. 8,9-*O*-Didemethyloxomaritidine was reduced by NNR yielding a product highly similar to oxomaritinamine in its mass spectrum indicating that the product is also reduced at the carbon-carbon double bond and therefore is 9-*O*-demethyloxomaritinamine. Fragments that are identical from these two molecules include *m/z* 91.1, *m/z* 94.1, *m/z* 108.1, and *m/z* 136.1. Fragments that represent the presence of an additional methyl group in oxomaritinamine compared to 9-*O*-demethyloxomaritinamine are *m/z* 246.1 and *m/z* 232.1, *m/z* 219.1 and *m/z* 205.1, *m/z* 204.1 and *m/z* 190.1, *m/z* 190.1 and *m/z* 176.1, and *m/z* 165.0 and *m/z* 151.0.

Table 4.1. Substrate specificity of NNR

Substrate	Product monitored	Specific activity (standard deviation) (pmol min^{-1} mg^{-1})
3,4-Dihydroxybenzaldehyde and tyramine	Norbelladine	21.2 (1.2)
Isovanillin and tyramine	4'-*O*-Methylnorbelladine	5.33 (0.21)
Noroxomaritidine	Oxomaritinamine	8600 (1250)
Vanillin and tyramine	3'-*O*-Methylnorbelladine	+
Piperonal and tyramine	4-(2-((1,3-Benzodioxol-5-ylmethyl)amino)ethyl)phenol	+
8-*O*-,9-*O*-Didemethyloxomaritidine	9-*O*-Demethyloxomaritinamine	+

[+]product made but not quantified

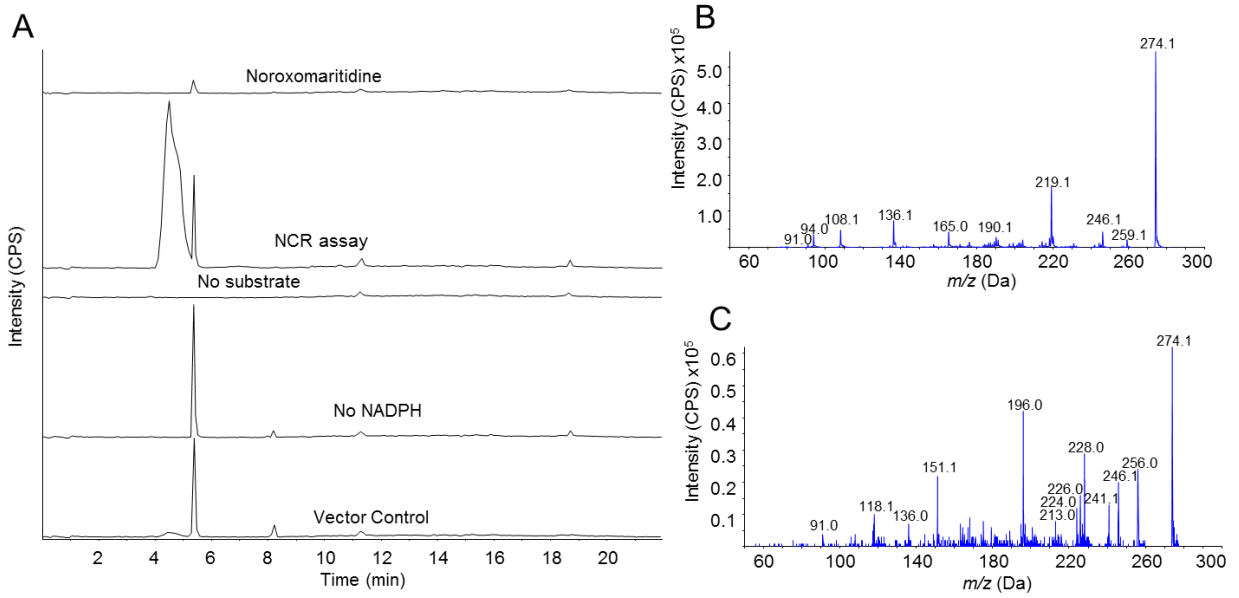

Figure 4.3. LC-MS/MS enhanced product ion scan (EPI) for NNR enzyme assays m/z 274.3 monitoring reduction of noroxomaritidine. (A) Top to bottom substrates noroxomaritidine standard, NNR assay, assay without noroxomaritidine, assay without NADPH, and assay with *E. coli* empty vector protein extract substituted for NNR protein. (B) EPI of NNR product. (C) EPI of noroxomaritidine reduced with $NaBH_4$.

An analysis of the relative reaction rates for substrates potentially involved in Amaryllidaceae alkaloid biosynthesis indicated several NNR preferences. Assays on 3,4-dihydroxybenzaldehyde-tyramine or noroxomaritidine show preference for the noroxomaritidine

reduction by several orders of magnitude, 21.2(1.2) μmol min^{-1} mg^{-1} verses 8600(1250) μmol min^{-1} mg^{-1} (Table 4.1). This is not surprising considering the condensation reaction that is prerequisite to the reduction of norcraugsodine. The specific activity for the isovanillin-tyramine reaction, 5.33(0.21) μmol min^{-1} mg^{-1}, was determined to be about 4-fold less than 3,4-dihydroxybenzaldehyde-tyramine. This difference supports the proposed pathway with 3,4-dihydroxybenzaldehyde as the substrate for condensation and reduction with subsequent methylation by N4OMT to 4'-*O*-methylnorbelladine.

4.3.3 Inhibition of Noroxomaritidine Reduction

The reduction of noroxomaritidine provided the opportunity to test the occupancy of the active site of NNR by 3,4-dihydroxybenzaldehyde and tyramine. If the enzyme active site binds these substrates, it could facilitate the formation of norcraugsodine prior to reduction to norbelladine. Also, if the enzyme binds these substrates then the individual components could act as competitive inhibitors for the reduction of noroxomaritidine. If previously formed norcraugsodine is the substrate NNR acts on than this inhibition would not be expected. The IC_{50} values of tyramine, 3,4-dihydroxybenzaldehyde, and piperonal are greater than 300 μM (Table 4.2). Piperonal has a lower IC_{50} value (330 μM) than 3,4-dihydroxybenzaldehyde (1,050 μM). This could result from the increased stability of piperonal relative to 3,4-dihydroxybenzaldehyde or a stronger binding to the active site. These values are high, however, strictosidine synthase has been shown to have a K_m of even greater magnitude K_m ~4 mM for tryptamine and secologanin (Hampp and Zenk, 1988).

Table 4.2. NNR noroxomaritidine reduction inhibition

Inhibitor	IC_{50} (μM)
Tyramine	1,050
3,4-Dihydroxybenzaldehyde	3,870
Piperonal	330

4.3.4 Structure of NNR

The crystal structures of *Narcissus* sp. *aff. pseudonarcissus* NNR complexed with $NADP^+$, $NADP^+$ and tyramine, and $NADP^+$ and piperonal were determined to 1.7 Å, 1.8 Å and 1.5 Å, respectively. NNR can be classified as a classical short-chain dehydrogenase reductase (SDR), containing the conserved dinucleotide binding motif TGxxxGxG at the N-terminus (Persson and Kallberg, 2013). NNR is a tetramer with each chain being 271 amino acids in length and contains the classic Rossmann fold typical of the SDR enzyme family (Rossmann et al., 1975). An alternating α/β structure formed by seven parallel β-sheets sandwiched between two layers of three α-helices shapes the NADPH binding domain (Figure 4.4A). NNR uses NADPH as an electron donor to catalyze the reduction of norcraugsodine to norbelladine. NADPH binds the active site pocket of NNR through polar interactions and nonpolar interactions (Figure 4.4B). The adenine ring, which is in the anti conformation, interacts with the side chain carboxylate of Val-83 through its exocyclic N6 and a terminal hydroxyl group of Asp-82 through hydrogen bonding with N1. The 2′-phosphate of the adenine ribose interacts with Arg-57 through hydrogen bonds to η_1N and εN of the side chain and to the backbone amide-nitrogen. Additionally, the adenine ribose phosphate interacts with the hydroxyl side chain of Ser-56. These interactions are key structural features that distinguish the NADPH SDRs from the NADH-specific enzymes (Deacon et al., 2000; Filling et al., 2002). The adenine binding pockets of all SDRs are stabilized by a highly conserved DxxD motif (Filling et al., 2002). The nicotinamide ribose has the C3′-endo conformation, and its C3

hydroxyl group hydrogen bonds to the Thr-34 hydroxyl. The 2′-hydroxyl of the nicotinamide ribose interacts with Tyr-175, a key residue for substrate recognition, and the 3′-hydroxyl of the nicotinamide ribose interacts with another key substrate-recognition residue, Lys-199, and the carboxylate of Asn-110. A water molecule is responsible for hydrogen bonding to two of the nicotinamide ribose phosphate oxygens, while a third hydrogen bond forms between an oxygen on the first nicotinamide ribose phosphate and Thr-210. The syn conformation of the nicotinamide ring is stabilized by a hydrogen bond involving the nicotinamide carboxamide group to the backbone amide-nitrogen Ile-208. The A-face of the nicotinamide ring packs against several aliphatic side chains (Ile-208, Pro-225, Gly-226, and Ala-227) and directs the B-face toward the binding pocket for the imine substrate. The NNR·$NADP^+$ complex was intended to be an apo enzyme but the binding affinity of $NADP^+$ to the NNR active site prevented its removal during purification from *E. coli.*

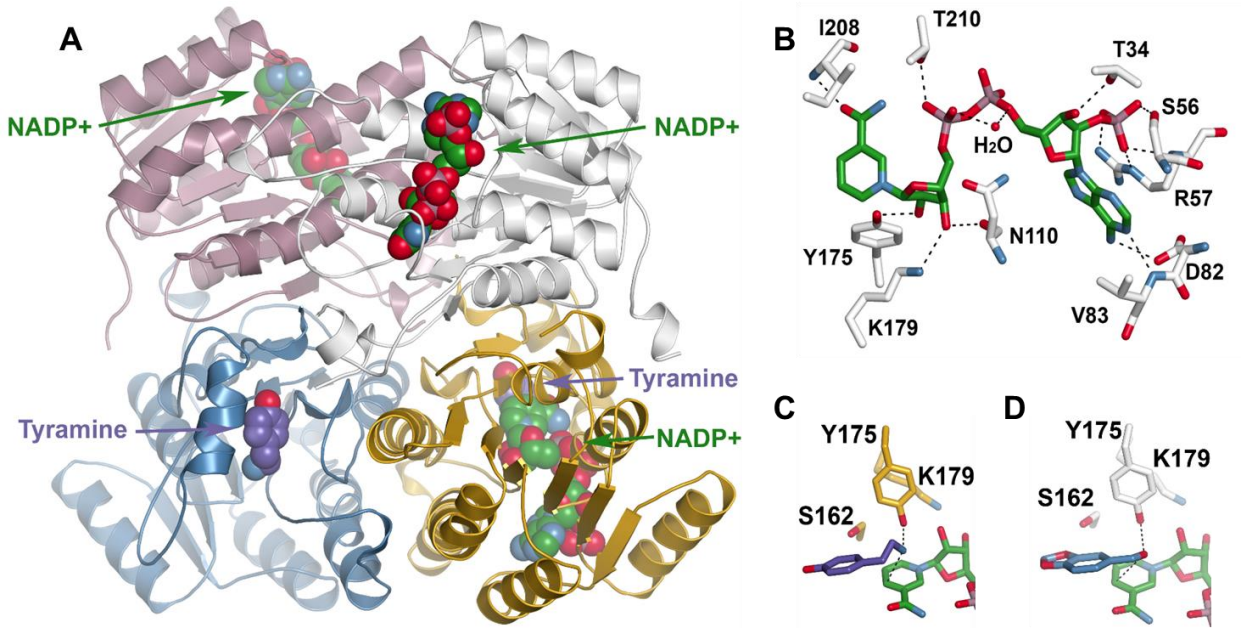

Figure 4.4. Crystal structure of NNR. (A) Ribbon diagram of the NNR tetramer complexed with $NADP^+$ and tyramine. (B) Polar contacts between $NADP^+$ and NNR. (C) Tyramine co-crystallized with NNR. (D) Piperonal co-crystallized with NNR.

Tyramine and piperonal were co-crystallized with NNR with $NADP^+$ (Figure 4.4C and D). These crystals demonstrate that these two compounds can bind to the NNR active site prior to condensation into a Schiff-base. In the context of the noroxomaritidine reduction both these compounds were demonstrated to inhibit the reduction and this is consistent with their position in the active site (Table 4.2). Both compounds interact with residues Tyr-175 and Lys-179, blocking the transfer of the hydride from C4 of the nicotinamide ring, and would prevent the reduction reaction of noroxomaritidine. The amine group of tyramine hydrogen bonds to Tyr-175 at a distance of 3.2 Å, while the distance between tyramine and C4 of the nicotinamide is 3.36 Å.

4.4 Discussion

NNR belongs to the SDR superfamily. It is closely related to the tropinone reductase-like reductases that are frequently involved in the metabolism of secondary metabolites (Brock et al., 2008; Reinhardt et al., 2014). This is the second example of a member of the SDR superfamily in plants reducing an imine functional group (Stavrinides et al., 2015). This makes SDRs promising candidates to target for identifying additional imine reductase activities.

NNR shows two separate enzymatic activities. One is the formation of norbelladine from a mixture of tyramine and 3,4-dihydroxybenzaldehyde. The norbelladine forming reaction is one of the first committed steps in Amaryllidaceae alkaloid biosynthesis. The reaction consists of two steps: condensation of tyramine and 3,4-dihydroxybenzaldehyde to an imine, and subsequent reduction of the imine. NNR was shown to have norcraugsodine reductase activity, and potentially facilitates the condensation of tyramine and 3,4-dihydroxybenzaldehyde to form norcraugsodine. The competition experiments designed to determine the enzyme's ability to bind 3,4-dihydroxybenzaldehyde and tyramine alone resulted in high IC_{50} values. However, actual concentrations of tyramine can be 3 mM in bulb tissue of *Narcissus* sp. *aff. pseudonarciss.* In

addition, enzymes for secondary metabolism can have high K_m values, such as that for strictosidine synthase (Hampp and Zenk, 1988). Crystal structures also showed binding of the individual components tyramine and piperonal to the NNR active site. This confirms the binding of these uncondensed molecules to the active site.

A second activity of the enzyme is the reduction of one of the enantiomeric forms of noroxomaritidine to oxomaritinamine. Depending on the enantiomer being reduced, this second activity could contribute to the biosynthesis of maritinamine or elwesine (Boit and Döpke, 1961; Pabuççuoglu et al., 1989). The double bond is reduced in both maritinamine or elwesine, but are derived from (10b*R*,4a*S*)-noroxomaritidine and (10b*S*,4a*R*)-noroxomaritidine, respectively.

When using co-expression analysis to search for additional Amaryllidaceae alkaloid biosynthetic genes individual genes will vary in their expression pattern over diverse environmental, developmental, tissue, and time points. Because of these differences the addition of *NNR* to a co-expression analysis with *N4OMT* and CYP96T1 in systems other than what was initially used to identify them will enrich efforts to identify additional biosynthetic genes and provide a deeper understanding of the relationships between Amaryllidaceae alkaloid biosynthesis and accumulation of participating enzymes.

The discovery of *NNR* now makes it possible to produce Amaryllidaceae alkaloids through bioengineering. Tyrosine decarboxylase, a well-known enzyme in other systems, has been identified in *Narcissus* sp. *aff. pseudonarcissus* and the recombinant enzyme has been biochemically validated to have tyrosine decarboxylase activity (Appendix A). While some aspects of 3,4-dihydroxybenzaldehyde synthesis remain unknown, a synthetic biology system for the production of vanillin produces 3,4-dihydroxybenzaldehyde as an intermediate (Hansen et al.,

2009). It is possible that the addition of tyrosine decarboxylase, the enzymes required for 3,4-dihydroxybenzaldehyde production, and NNR to a heterologous system would be sufficient for the production of norbelladine. Even if the synthase activity is not sufficient in NNR to drive the condensation reaction, a heterologous system may make sufficient quantities of the two substrates that the spontaneous condensation proceeds and NNR can mediate reduction. The previously discovered *N4OMT* and CYP96T1 could also be added to the system to potentially obtain haemanthamine, crinine, and some galanthamine skeleton derivatives (Kilgore et al., 2014). This could provide a more efficient production system for these anticancer compounds and Alzheimer's disease treatments (Wilcock et al., 2003; Havelek et al., 2014).

4.5 Acknowledgments

This work was supported by the National Institutes of Health award number 1RC2GM092561 (NIGMS). This material is based upon work supported by the National Science Foundation under Grant No. DBI-0521250 for acquisition of the QTRAP LC-MS/MS. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1143954. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Crystallography results shown in this report are derived from work performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by UChicago Argonne, LLC, for the U.S. Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357.

Chapter 5: Conclusions and Perspectives

5.1 Thesis Conclusions

This thesis examines how Amaryllidaceae alkaloids are biosynthesized using an effective workflow for identifying pathway genes in non-model systems. The hypothesis from which this workflow evolved is that genes involved in the biosynthesis of particular Amaryllidaceae alkaloids are co-regulated and this leads to correlations between biosynthetic gene expression and corresponding end-product accumulation. The discovery of the Amaryllidaceae alkaloid biosynthetic genes norbelladine 4'-*O*-methyltransferase (*N4OMT*), CYP96T1, and noroxomaritidine/norcraugsodine reductase (*NNR*) in Chapters 2, 3, and 4 through correlations with galanthamine accumulation (*N4OMT*) and co-expression with *N4OMT* (CYP96T1 and *NNR*) confirms this underlying hypothesis. The confirmation of this hypothesis is in keeping with previous work demonstrating co-expression of other pathways with end metabolites and their biosynthetic genes (Augustin et al., 2015). These core pathway enzyme discoveries also provide biosynthetic information relevant to many biologically active Amaryllidaceae alkaloids, including the Alzheimer's treatment galanthamine and the anti-cancer compound haemanthamine (Wilcock et al., 2003; Havelek et al., 2014).

The three newly discovered enzymes were found to have the following activities: NNR reduces norcraugsodine to norbelladine, N4OMT methylates norbelladine to 4'-*O*-methylnorbelladine, and CYP96T1 phenol-phenol couples 4'-*O*-methylnorbelladine to noroxomaritidine. When characterizing N4OMT, CYP96T1, and NNR, previously unknown aspects of Amaryllidaceae alkaloid biosynthesis were discovered. N4OMT and CYP96T1 were both shown to accept *N*-methylated versions of norbelladine and 4'-*O*-methylnorbelladine as substrate. Detailed radiolabeling experiments by Eichhorn et al. have placed the *N*-methylation of

galanthamine-type alkaloids after phenol-phenol coupling (Eichhorn et al., 1998). Prior to these experiments, it was thought that *N*-methylation occurred prior to phenol-phenol coupling due to radiolabeling experiments of Barton and Kirby (Barton and Kirby, 1962). The ability of N4OMT and CYP96T1 to accept substrates with and without *N*-methylation may explain incorporation of 4'-*O*-methyl-*N*-methylnorbelladine in earlier work.

CYP96T1 generated the enantiomeric mixture (±)-noroxomaritidine when using 4'-*O*-methylnorbelladine as substrate. This could indicate a role of this cytochrome P450 in formation of both the haemanthamine series of compounds and the crinine series. The presence or absence of derivatives from one or the other enantiomer in different Amaryllidaceae species could be a result of degradation pathways, a stereospecific version of this enzyme, or chiral directing of the achiral intermediate by another enzyme *in planta*.

NNR was shown to have norcraugsodine reductase activity, and potentially facilitates the condensation of tyramine and 3,4-dihydroxybenzaldehyde to form norcraugsodine. It was also shown to be multifunctional with the ability to reduce the carbon-carbon double bond on one of the enantiomers of noroxomaritidine. The potential involvement of NNR in all these reactions will be informative to future work on these reactions.

5.2 Application of Minimalist and Extensive Workflows

This thesis demonstrates that a simple workflow consisting of RNA-seq on only a few well-determined tissues and quantification of a limited number of metabolites can be used to discover biosynthetic genes in species and pathways for which previous genetic knowledge is lacking. This is a practical alternative workflow to methods that take a metabolomics approach with substantial expression data (Saito et al., 2008). The extensive approach has better statistical power, but can be cost and labor prohibitive in non-model species without available omic data. The extensive

approach is particularly problematic if the plant tissue being analyzed comes from a rare plant, which is sometimes the case in secondary metabolism research (Augustin et al., 2015). However, the minimalist approach is limited statistically, due to the low sample number, causing small correlations to be undetectable. Also, experimental conditions with strong correlations between genes may be lacking from the dataset. Because of the resulting danger of false negatives when using a minimalist approach, the ability to test for multiple enzymes in a pathway can be vital. If one enzyme in the pathway is not observed to co-express with another under tested conditions a third enzyme in the pathway may be co-expressing and discoveries can be made without 100 % success for every enzymatic activity tested.

The CYP96T1 work demonstrates the ability to identify biosynthetic genes by comparing co-expression across species in the Amaryllidaceae. In a rare species, this could be applied to identify shared biosynthetic genes using data from closely related common species with similar alkaloids, forming the basis for the identification of the unique biosynthetic genes in the rare species. Interspecies comparisons are useful when applied to species with extensive information in relation to a species with limited information. Homology, along with a few expression data points, can be used to link one of these well-studied species with an orphan species of interest. Using known genes along with the unique properties of non-model species, new insights can be obtained by just adding a few more data points to the well-characterized system. The co-expression of transcripts in Amaryllidaceae alkaloid biosynthesis and co-accumulation of alkaloids with these transcripts opens the door for future co-expression studies in Amaryllidaceae alkaloids. It also validates the use of co-expression analysis as a starting point when studying secondary metabolism in non-model species.

5.3 Future Directions for Amaryllidaceae Alkaloid Biosynthesis

The discovery of more biosynthetic genes in the pathways for Amaryllidaceae alkaloids is an area of future interest because of their medicinal potential and importance to Amaryllidaceae biology. Many of the biosynthetic steps leading to compounds like haemanthamine and galanthamine remain undiscovered. These uncharacterized genes likely belong to established biosynthetic gene superfamilies. Given the co-expression observed between the discovered biosynthetic genes, it is likely several of these uncharacterized genes will also co-express. For this reason, the application of co-expression coupled with homology to superfamilies with the same underlying chemistry will likely result in the discovery of additional biosynthetic genes. Steps of particular interest include the reduction of *N*-demethylnarwedine to *N*-demethylgalanthamine, which is likely catalyzed by an AKR or an SDR superfamily member. While this gene was not found to co-express with *N4OMT* in ≥ 4 of the transcriptomes in Chapter 4, it is likely present in the ≥ 3 or the ≥ 2 lists. These genes were observed during the discovery of the *NNR*, but were assigned a lower priority than the genes co-expressing in 4 transcriptomes. The same applies to the reductase responsible for the reduction of noroxomaritidine to normaritidine. These reductases would be of particular biotechnological interest because they are directly downstream of the enzymatically characterized core biosynthetic genes and could prove useful for the production of these compounds in heterologous systems. Also of interest are the *N*-methyltransferase that converts *N*-demethylgalanthamine to galanthamine, the oxidase (probably a cytochrome P450) responsible for the oxide bridge formation in the normaritidine to vittatine conversion, the hydroxylase (possibly a cytochrome P450 or a 2-oxoglutarate-dependent dioxygenase) that converts vittatine to 11-hydroxyvittatine, and the methyltransferase that converts 11-hydroxyvittatine to haemanthamine. A cytochrome P450 that

yields more *N*-demethylnarwedine than CYP96T1 may yet be discovered, given the high content of galanthamine type alkaloids in some Amaryllidaceae.

In addition to biosynthetic genes transcription factors regulating their biosynthesis could be discovered through co-expression analysis, examination of transcription factor binding motifs on the promoters of known genes, or proximity on gene clusters. The discovery of these transcription factors could enable an understanding of the biosynthetic regulation of these compounds and help identify other plant pathways that may share these transcription factors and therefore experience similar regulation. The resulting understanding of this regulation could assist efforts to boost the content of specific Amaryllidaceae alkaloids in members of the Amaryllidaceae and provide insight into other co-regulated pathways that may work with Amaryllidaceae alkaloid biosynthesis to achieve goals such as deterring herbivores.

A potential expansion of the work would be to sequence more members of the Amaryllidaceae. Sequencing projects have already been performed on the closely related bulbs *Allium cepa* (onion), *Allium sativum* (garlic), *Allium fistulosum* (bunching onion), and *Allium tuberosum* (chinese chive) (Kamenetsky et al., 2015; Rajkumar et al., 2015; Tsukazaki et al., 2015). A cross comparison of additional species could make phylogenetic sense of the distribution of Amaryllidaceae alkaloids across the Amaryllidaceae. Searching for the characterized genes in additional species and in the neighboring alliums could yield information regarding their origin. In addition, the sequencing of species with unique metabolites such as *Crinum* spp. would facilitate the discovery of the enzymes that lead to these unique products. These studies on the diversity of Amaryllidaceae alkaloids would provide key information for attempts to tailor Amaryllidaceae alkaloid pathways, in heterologous systems and members of the Amaryllidaceae, for optimal alkaloid production or the absence of these frequently poisonous alkaloids.

The discovery of additional biosynthetic genes may require a more extensive collection of conditions and tissues to use for co-expression analysis. Because some biosynthetic genes differ in regulation (modulated for example through differing transcription factors) depending on tissue, time in development, and stress some genes may require data sets with diverse tissues, time points, and environmental conditions to identify co-expression. To determine how consistently these discovered biosynthetic genes co-express, a finer organ or cell type level than leaf, bulb, and inflorescence could be monitored with qRT-PCR. Effects of environmental conditions and variation over the course of a year could also be monitored. This data could help identify the level of variability between known biosynthetic genes providing a baseline for unknown biosynthetic genes. Problematic tissues, time points, and conditions with little correlation between the known biosynthetic genes could be observed and perhaps excluded in co-expression experiments, or examined more thoroughly to determine the cause of the discrepancy between the expression patterns.

A cell type by cell type analysis of known-gene expression and the alkaloid end products could help determine if the Amaryllidaceae alkaloid biosynthesis is cell type specific or if transport occurs as in the biosynthesis of vindoline in *Catharanthus roseus*. In vindoline biosynthesis different cell types are required for early and late biosynthetic steps (St-Pierre et al., 1999). When transport of intermediates between cell types is required for the biosynthesis of secondary metabolites this other cell type could be next to the original cell or in a completely different organ of the plant. Potential reasons for the transport of intermediates include regulation of the biosynthetic process through prevention of undesired side reactions or a simpler evolutionary route where transport rather than alteration of ancestral biosynthetic gene promoters allow for cell type co-expression. An example of specialized cells for secondary metabolite production is the secretory

cells on the glandular trichomes of *Mentha piperita*. These cells secrete monoterpenoids into a central subcuticular cavity that stores the toxic terpenoids for protection of the leaf from herbivore attack (McCaskill et al., 1992). The need for a specialized cavity to store the toxic monoterpenoids is what drives cell type specific production in *Mentha piperita*. In addition to minimizing toxicity through specialized structures these cells may produce these metabolites to spare other cells from the associated metabolic load.

In addition to measuring transcript level, measuring changes in alkaloid content during the above experiments could provide information on the expression of unknown biosynthetic genes responsible for the elaboration of these alkaloids. For example, if 11-hydroxyvittatine is present but not haemanthamine in one tissue, while haemanthamine and not 11-hydroxvittatine is present in another tissue this would be an indication that the methyltransferase that methylates 11-hydroxyvittatine to haemanthamine is present in the latter tissue but not the former. Such discoveries would help with the detection of the corresponding enzyme by incorporating new information into the co-expression analysis.

The examination of the proteome during any of these experiments could provide the actual levels of the enzymes of interest in different tissues. The correlation of the transcriptome to the proteome can be moderate (Ghazalpour et al., 2011). Discrepancies between transcript levels and active protein levels can be due to protein degradation, protein inactivation by redox signaling, a phosphorylation requirement, or other regulatory modification. These changes can be detected through proteomic analysis (Zhang and Ge, 2011). The ability of proteomic analysis to detect these changes could enable the identification of changes in active protein content and identify candidates correlating with the biosynthesis of Amaryllidaceae alkaloids on the protein level but not on the

transcript level. The proteome would also serve as an independent validation of sequences in the transcriptome that frequently align to peptides.

Another route to discover more biosynthetic genes is to place candidate genes under more biologically relevant conditions for enzyme assays. The temperature and pH used in screening assays are slightly higher than biological 30 °C verses 20-25 °C and the pH is approximately 7.5-8.0 which may not mimic biological conditions if enzymes are compartmentalized in locations such as the vacuole with a reported pH of 5.5 (Mathieu et al., 1989). If the Amaryllidaceae alkaloids are biosynthesized in a complex of enzymes that facilitates biosynthesis through shuttling of the intermediates or a localized increase in intermediates, then some enzymes in the pathway may only function effectively in the context of this complex. It has been observed that when chorismate mutase and 3-deoxy-D-arabino-heptulosonate 7-phosphate synthase, two biosynthetic enzymes in aromatic amino acid biosynthesis, in *Mycobacterium tuberculosis* interact their allosteric properties are extended and enzymatic activity is enhanced (Blackmore et al., 2015). If some Amaryllidaceae alkaloid biosynthetic enzymes require this physical interaction to function, then some enzyme activates may not be identifiable through an enzyme assay approach until a better understanding of this potential complex and its components is achieved.

5.4 Potential Applications to Biotechnology and Amaryllidaceae Basic Science

In addition to interrogating metabolomics, transcriptomic, and genomic datasets for other biosynthetic genes, several experiments could be done using the three genes characterized in this thesis. These genes can be used for the production of noroxomaritidine and some *N*-demethylnarwedine in a heterologous expression system such as yeast or Camelina. The biosynthesis of 3,4-dihydroxybenzaldehyde in the Amaryllidaceae has at least one uncharacterized

enzyme, but a synthetic system for the creation of this compound has already been described using 3-dehydroshikimate dehydratase from *Podospora pauciseta* and an aromatic carboxylic acid reductase from *Nocardia* sp. that could be used in these heterologous systems (Hansen et al., 2009). To create a system for the production of galanthamine in yeast or *in planta* will require 2-3 enzymes in addition to the 3 discovered (if you count CYP96T1) and the 3 required for 3,4-dihydroxybenzaldehdye and tyramine production. CYP96T1 will make some *N*-demethylnarwedine but an enzyme with a higher specificity for this product would be desirable. The reductase for the *N*-demethylnarwedine to *N*-demethylgalanthamine conversion and the *N*-methyltransferase for the *N*-demethylgalanthamine to galanthamine conversion would also be required. A system producing haemanthamine will require the discovery of a reductase, an oxide bridge forming enzyme, and a methyltransferase. A hydroxylase is also required for the conversion of vittatine to 11-hydorxyvittatine in haemanthamine biosynthesis but the 2-oxoglutarate-dependent dioxygenase described in Appendix B would provide this activity if the chirality of the product is appropriate for haemanthamine biosynthesis.

Although many Amaryllidaceae alkaloids can be made chemically, this process can require toxic catalysts and many of the chemical synthesis protocols are not efficient or scalable for commercial production. The ability to produce these compounds with low cost inputs and low toxicity outputs using biological systems including members of the Amaryllidaceae, yeast, or a fast growing plant such as Camelina is appealing. The discovery of the missing enzymes for galanthamine or haemanthamine biosynthesis would enable experiments aimed at improving their production in biological systems. The need to discover biosynthetic genes in valuable secondary metabolite pathways from non-model systems for improved production of these compounds is a general need to achieve the goal of more affordable healthcare. The development of generally

applicable workflows, such as the ones utilized in this thesis featuring technologies such as next generation sequencing, will enable the future discovery of these valuable biosynthetic genes.

Amaryllidaceae alkaloid biosynthetic gene knock-out cultivars with resulting altered alkaloid profiles could prove ideal systems for ecological tests of the effects of Amaryllidaceae alkaloids on plant fitness and species interactions. An Amaryllidaceae plant with a knockout mutation for *N4OMT* or *NNR* could be devoid of Amaryllidaceae alkaloids and a comparison of this plant to wild type under stresses including herbivore challenge would enable a direct testing of the proposed roles of these alkaloids *in planta* as described in section 1.1.2. The knockout of CYP96T1 would potentially result in a deficiency of haemanthamine type alkaloids but not others like galanthamine. The importance of the small quantity of galanthamine type alkaloid phenol-phenol coupling made by CYP96T1 could be determined by the presence or absence of galanthamine derivatives in the CYP96T1 mutant. If galanthamine type alkaloids are not significantly reduced in this mutant, then another enzyme may be primarily responsible for this reaction and the resulting plant may provide a cleaner alkaloid extract when preparing galanthamine commercially. CRISPR/Cas technology enables the creation of these knockout mutants in these genes and their close paralogues in various Amaryllidaceae plants through precise gene editing (Sander and Joung, 2014).

The transcriptomic resources are also useful for CRISPR/Cas guided mutagenesis of other genes not necessarily related to Amaryllidaceae alkaloid biosynthesis, including genes for plant development, floral morphology, and floral color. These traits are of particular interest to the floral industry and using CRISPR/Cas with the transcriptomes in this way could lead to more desirable cultivars. It could also be used for answering basic biology questions regarding bulb development.

Due to the slow turnaround of the tissue culture and transformation protocol previously reported in *Narcissus tazzeta* var. *chinensis*, these projects would be long-term (Lu et al., 2007).

5.5 Summary

In summary, this thesis confirmed the hypothesis that genes involved in the biosynthesis of selected Amaryllidaceae alkaloids are co-regulated and that this leads to correlations between biosynthetic gene expression and corresponding end-product accumulation. Biosynthetic genes that were either co-accumulating with the Amaryllidaceae alkaloid galanthamine or were co-expressing with the biosynthetic gene *N4OMT* were thereby discovered. During the subsequent characterization of these enzymes, hypotheses on how Amaryllidaceae alkaloids are made were tested. The discovery of these enzymes validates a simple workflow for biosynthetic gene discovery in Amaryllidaceae and provides tools in the form of transcriptomes and validated enzymes. These tools can be used for further study of Amaryllidaceae alkaloid biosynthesis and regulation, general molecular biology questions relating to members of the Amaryllidaceae including bulb development, and biotechnology applications. This could lead to better production systems for medicinal compounds such as the Alzheimer's treatment galanthamine and anticancer compound haemanthamine and to a better understanding of this plant family.

References

FASTX-Toolkit website. (Available: http://hannonlab.cshl.edu/fastx_toolkit. Accessed 2014 Jul 2).

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., and Zwart, P.H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Crystallogr. D Biol. Crystallogr. **66,** 213-221.

Aharoni, A., Keizer, L.C., Bouwmeester, H.J., Sun, Z., Alvarez-Huerta, M., Verhoeven, H.A., Blaas, J., van Houwelingen, A.M., De Vos, R.C., van der Voet, H., Jansen, R.C., Guis, M., Mol, J., Davis, R.W., Schena, M., van Tunen, A.J., and O'Connell, A.P. (2000). Identification of the SAAT gene involved in strawberry flavor biogenesis by use of DNA microarrays. Plant Cell **12,** 647-662.

Akashi, T., Sawada, Y., Shimada, N., Sakurai, N., Aoki, T., and Ayabe, S. (2003). cDNA cloning and biochemical characterization of *S*-adenosyl-L-methionine: 2,7,4'-trihydroxyisoflavanone 4'-*O*-methyltransferase, a critical enzyme of the legume isoflavonoid phytoalexin pathway. Plant Cell Physiol. **44,** 103-112.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25,** 3389-3402.

Aramini, J.M., Rossi, P., Anklin, C., Xiao, R., and Montelione, G.T. (2007). Microgram-scale protein structure determination by NMR. Nat. Methods. **4,** 491-493.

Au, K.F., Underwood, J.G., Lee, L., and Wong, W.H. (2012). Improving PacBio long read accuracy by short read alignment. PLoS One **7,** e46679.

Augustin, M.M., Ruzicka, D.R., Shukla, A.K., Augustin, J.M., Starks, C.M., O'Neil-Johnson, M., McKain, M.R., Evans, B.S., Barrett, M.D., Smithson, A., Wong, G.K., Deyholos, M.K., Edger, P.P., Pires, J.C., Leebens-Mack, J.H., Mann, D.A., and Kutchan, T.M. (2015). Elucidating steroid alkaloid biosynthesis in *Veratrum californicum*: production of verazine in Sf9 cells. Plant J. **82,** 991-1003.

Banfi, L., Narisano, E., Riva, R., Stiasni, N., Hiersemann, M., Yamada, T., and Tsubo, T. (2001). Sodium Borohydride. In Encyclopedia of Reagents for Organic Synthesis (John Wiley & Sons, Ltd.

Barik, J., Dajas-Bailador, F., and Wonnacott, S. (2005). Cellular responses to nicotinic receptor activation are decreased after prolonged exposure to galantamine in human neuroblastoma cells. Br. J. Pharmacol. **145,** 1084-1092.

Barton, D.H.R., and Kirby, G.W. (1962). Phenol oxidation and biosynthesis. Part V. The synthesis of galanthamine. J. Chem. Soc.**,** 806-817.

Barton, D.H.R., Kirby, G.W., and Thomas, G.M. (1963). Phenol oxidation and biosynthesis. Part VI. The biogenesis of Amaryllidaceae alkaloids. J. Chem. Soc.**,** 4545-4558.

Barton, D.H.R., Kirby, G.W., Taylor, J.B., and Thomas, G.M. (1961). The biosynthesis of Amaryllidaceae alkaloids. Proc. Chem. Soc.**,** 254-255.

Bastida, J., Berkov, S., Torras, L., Pigni, N.B., Andrade, J.P., Martínez, V., Codina, C., and Vildomat, F. (2011). Chemical and biological aspects of Amaryllidaceae alkaloids. (Kerala, India: Transworld Research Network).

Battersby, A.R., and Binks, R. (1960). Biosynthesis of lycorine. Proc. Chem. Soc.**,** 410-411.

Battersby, A.R., Fales, H.M., and Wildman, W.C. (1961a). Biosynthesis in the Amaryllidaceae. Tyrosine and norbelladine as precursors of haemanthamine. J. Am. Chem. Soc. **83,** 4098-4099.

Battersby, A.R., Bink, R., and Breuer, S.W. (1961b). Biosynthesis in the Amaryllidaceae: incorporation of norbelladine into lycorine and norpluvine. Proc. Chem. Soc.**,** 243.

Battersby, A.R., Binks, R., Breuer, S.W., Fales, H.M., Wildman, W.C., and Highet, R.J. (1964). Alkaloid biosynthesis. Part III.* Amaryllidaceae alkaloids: the biosynthesis of lycorine and its relatives. J. Chem. Soc.**,** 1595-1609.

Belin, P., Le Du, M.H., Fielding, A., Lequin, O., Jacquet, M., Charbonnier, J.B., Lecoq, A., Thai, R., Courçon, M., Masson, C., Dugave, C., Genet, R., Pernodet, J.L., and Gondry, M. (2009). Identification and structural basis of the reaction catalyzed by CYP121, an essential cytochrome P450 in Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U.S.A. **106,** 7426-7431.

Berkov, S., Sidjimova, B., Evstatieva, L., and Popov, S. (2004). Intraspecific variability in the alkaloid metabolism of *Galanthus elwesii*. Phytochemistry **65,** 579-586.

Berkov, S., Martínez-Francés, V., Bastida, J., Codina, C., and Ríos, S. (2014). Evolution of alkaloid biosynthesis in the genus *Narcissus*. Phytochemistry **99,** 95-106.

Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., Horsman, D.E., Connors, J.M., Gascoyne, R.D., Marra, M.A., and Jones, S.J. (2009). *De novo* transcriptome assembly with ABySS. Bioinformatics **25,** 2872-2877.

Blackmore, N.J., Nazmi, A.R., Hutton, R.D., Webby, M.N., Baker, E.N., Jameson, G.B., and Parker, E.J. (2015). Complex formation between two biosynthetic enzymes modifies the allosteric regulatory properties of both: an example of molecular symbiosis. J. Biol. Chem. **290,** 18187-18198.

Boit, H.-G., and Döpke, W. (1961). Alkaloide aus *Haemanthus*-, *Zephranthes*-, *Galanthus*- und *Crinum*-arten. Die Naturwissenschaften **48,** 406-407.

Boycheva, S., Daviet, L., Wolfender, J.L., and Fitzpatrick, T.B. (2014). The rise of operon-like gene clusters in plants. Trends Plant Sci. **19,** 447-459.

Bringmann, G., Wohlfarth, M., Rischer, H., Grüne, M., and Schlauer, J. (2000). A New Biosynthetic Pathway to Alkaloids in Plants: Acetogenic Isoquinolines. Angew. Chem. Int. Ed. Engl. **39,** 1464-1466.

Brock, A., Brandt, W., and Dräger, B. (2008). The functional divergence of short-chain dehydrogenases involved in tropinone reduction. Plant J. **54,** 388-401.

Bruce, I.T., and Kirby, G.W. (1968). Stereochemistry of protonation and hydroxylation in the biosynthesis of norpluviine and lycorine. Chem. Commun. (London)**,** 207-208.

Cabezas, F., Ram, iacute, rez, A., Viladomat, F., Codina, C., and Bastida, J. (2003). Alkaloids from *Eucharis amazonica* (Amaryllidaceae). Chem. Pharm. Bull. **51,** 315-317.

Casida, J.E. (1964). Esterase Inhibitors as Pesticides. Science **146,** 1011-1017.

Castro, A., Moco, S., Coll, J., and Vervoort, J. (2010). LC-MS-SPE-NMR for the isolation and characterization of neo-clerodane diterpenoids from *Teucrium luteum* subsp. *flavovirens* (perpendicular). J. Nat. Prod. **73,** 962-965.

Chae, L., Kim, T., Nilo-Poyanco, R., and Rhee, S.Y. (2014). Genomic signatures of specialized metabolism in plants. Science **344,** 510-513.

Chan, J.L.A. (1973). Biosynthesis of cherylline using doubly-labled norbelladine-type precursors. In Chemistry (Retrospective Theses and Dissertations: Iowa State University), pp. 116.

Chase, M.W., Reveal, J.L., and Michael, F.F. (2009). A subfamilial classification for the expanded asparagalean families Amaryllidaceae, Asparagaceae and Xanthorrhoeaceae. Bot. J. Linn. Soc. **161,** 132-136.

Cheesman, L., Nair, J.J., and Van Staden, J. (2012). Antibacterial activity of crinane alkaloids from *Boophone disticha* (Amaryllidaceae). J. Ethnopharmacol. **140,** 405-408.

Chen, J., and Ziv, M. (2005). The effects of storage condition on starch metabolism and regeneration potentials of twin-scales and inflorescence stem explants of *Narcissus tazetta*. In Vitro Cell. Dev. Biol. Plant **41,** 816-821.

Chevreux, B., Pfisterer, T., Drescher, B., Driesel, A.J., Müller, W.E., Wetter, T., and Suhai, S. (2004). Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res. **14,** 1147-1159.

Chilvers, M.T., and Daft, M.F.J. (1981). Mycorrhizas of the Liliiflorae: II. Mycorrhiza formation and incidence of root hairs in field grown *Narcissus* L., *Tulipa* L., and *Crocus* L. cultivars. New Phytol. **89,** 247-261.

Choi-Rhee, E., and Cronan, J.E. (2005). A nucleosidase required for in *vivo* function of the *S*-adenosyl-L-methionine radical enzyme, biotin synthase. Chem. Biol. **12,** 589-593.

Choi, K.B., Morishige, T., Shitan, N., Yazaki, K., and Sato, F. (2002). Molecular cloning and characterization of coclaurine *N*-methyltransferase from cultured cells of *Coptis japonica*. J. Biol. Chem. **277,** 830-835.

Constantin, M.A., Conrad, J., and Beifuss, U. (2012). Laccase-catalyzed oxidative phenolic coupling of vanillidene derivatives. Green Chem. **14,** 2375-2379.

Cremin, P.A., and Zeng, L. (2002). High-throughput analysis of natural product compound libraries by parallel LC-MS evaporative light scattering detection. Anal. Chem. **74,** 5492-5500.

Davin, L.B., Jourdes, M., Patten, A.M., Kim, K.W., Vassão, D.G., and Lewis, N.G. (2008). Dissection of lignin macromolecular configuration and assembly: comparison to related biochemical processes in allyl/propenyl phenol and lignan biosynthesis. Nat. Prod. Rep. **25,** 1015-1090.

de La Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. In Meth. Enzymol. (Academic Press), pp. 472-494.

Deacon, A.M., Ni, Y.S., Coleman, W.G., and Ealick, S.E. (2000). The crystal structure of ADP-L-glycero-D-mannoheptose 6-epimerase: catalysis with a twist. Structure **8,** 453-462.

Díaz Chávez, M.L., Rolf, M., Gesell, A., and Kutchan, T.M. (2011). Characterization of two methylenedioxy bridge-forming cytochrome P450-dependent enzymes of alkaloid formation in the Mexican prickly poppy *Argemone mexicana*. Arch. Biochem. Biophys. **507,** 186-193.

Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32,** 1792-1797.

Eichhorn, J., Takada, T., Kita, Y., and Zenk, M.H. (1998). Biosynthesis of the amaryllidaceae alkaloid galanthamine. Phytochemistry **49,** 1037-1047.

Eldridge, G.R., Vervoort, H.C., Lee, C.M., Cremin, P.A., Williams, C.T., Hart, S.M., Goering, M.G., O'Neil-Johnson, M., and Zeng, L. (2002). High-throughput method for the production and analysis of large natural product libraries for drug discovery. Anal. Chem. **74,** 3963-3971.

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

Engelhardt, B.E., Jordan, M.I., Repo, S.T., and Brenner, S.E. (2009). Phylogenetic molecular function annotation. J. Phys. Conf. Ser. **180,** 12024.

Fahrendorf, T., and Dixon, R.A. (1993). Stress responses in alfalfa (*Medicago sativa* L.). XVIII: Molecular cloning and expression of the elicitor-inducible cinnamic acid 4-hydroxylase cytochrome P450. Arch. Biochem. Biophys. **305,** 509-515.

Fales, H.M., and Wildman, W.C. (1964). Biological interconversions in the Amaryllidaceae. I. The haemanthamine-haemanthidine-tazettine sequence. J. Am. Chem. Soc. **86,** 294-295.

Feinstein, A.I. (1967). The incorporation of ^{14}C-labeled [beta]-phenylethylamine derivatives and ^{3}H-vittatine into Amaryllidaceae alkaloids. In Chemistry (Retrospective Theses and Dissertations: Iowa State University), pp. 113.

Feinstein, A.I., and Wildman, W.C. (1976). Biosynthetic oxidation and rearrangement of vittatine and its derivatives. J. Org. Chem. **41,** 2447-2450.

Filling, C., Berndt, K.D., Benach, J., Knapp, S., Prozorovski, T., Nordling, E., Ladenstein, R., Jörnvall, H., and Oppermann, U. (2002). Critical residues for structure and catalysis in short-chain dehydrogenases/reductases. J. Biol. Chem. **277,** 25677-25684.

Finnegan, E.J., Peacock, W.J., and Dennis, E.S. (1996). Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. P. Natl. Acad. Sci. U.S.A. **93,** 8449-8454.

Fisher, R.A. (1930). The Genetical Theory of Natural Selection. (Oxford University Press).

Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. Nat. Methods. **7,** 461-465.

Franke, R., Humphreys, J.M., Hemm, M.R., Denault, J.W., Ruegger, M.O., Cusumano, J.C., and Chapple, C. (2002). The Arabidopsis REF8 gene encodes the 3-hydroxylase of phenylpropanoid metabolism. Plant J. **30,** 33-45.

Fratila, R.M., and Velders, A.H. (2011). Small-volume nuclear magnetic resonance spectroscopy. Annu. Rev. Anal. Chem. (Palo Alto Calif) **4,** 227-249.

Frick, S., and Kutchan, T.M. (1999). Molecular cloning and functional expression of *O*-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. Plant J. **17,** 329-339.

Fuganti, C. (1973). Evidence for the intermediacy of 11-hydroxyvittatine in the biosynthesis of narciclasine. Gazz. Chim. Ital.**,** 1255-1258.

Fuganti, C., and Mazza, M. (1971). Late intermediates in the biosynthesis of narciclasine. Chem. Commun.**,** 1388-1389.

Fuganti, C., and Mazza, M. (1972a). Stereochemistry of hydorxylation in the biosynthesis of lycorine in *Clivia miniata* regel. J. Chem. Soc. Chem. Comm.**,** 936-937.

Fuganti, C., and Mazza, M. (1972b). The absolute configuration of narciclasine: A biosynthetic approach. J. C. S. Chem. Comm.**,** 239.

Fuganti, C., Staunton, J., and Battersby, A.R. (1971). The biosynthesis of narciclasine. J. Chem. Soc. Chem. Comm.**,** 1154-1155.

Furnham, N., Sillitoe, I., Holliday, G.L., Cuff, A.L., Laskowski, R.A., Orengo, C.A., and Thornton, J.M. (2012). Exploring the evolution of novel enzyme functions within structurally defined protein superfamilies. PLoS Comput. Biol. **8,** e1002403.

Gabrielsen, B., Monath, T.P., Huggins, J.W., Kefauver, D.F., Pettit, G.R., Groszek, G., Hollingshead, M., Kirsi, J.J., Shannon, W.M., and Schubert, E.M. (1992). Antiviral (RNA) activity of selected Amaryllidaceae isoquinoline constituents and synthesis of related substances. J. Nat. Prod. **55,** 1569-1581.

Gadaga, L.L., Tagwireyi, D., Dzangare, J., and Nhachi, C.F. (2011). Acute oral toxicity and neurobehavioural toxicological effects of hydroethanolic extract of *Boophone disticha* in rats. Hum. Exp. Toxicol. **30,** 972-980.

Gallage, N.J., Hansen, E.H., Kannangara, R., Olsen, C.E., Motawia, M.S., Jørgensen, K., Holme, I., Hebelstrup, K., Grisoni, M., and Møller, B.L. (2014). Vanillin formation from ferulic acid in *Vanilla planifolia* is catalysed by a single enzyme. Nat. Commun. **5,** 4037.

Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P., and Pichersky, E. (2002). Characterization of phenylpropene *O*-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant *O*-methyltransferase family. Plant Cell **14,** 505-519.

Gerardy, R., and Zenk, M.H. (1993). Purification and characterization of salutaridine: NADPH 7-oxidoreductase from *Papaver somniferum*. Phytochemistry **34,** 125-132.

Gesell, A., Rolf, M., Ziegler, J., Díaz Chávez, M.L., Huang, F.C., and Kutchan, T.M. (2009). CYP719B1 is salutaridine synthase, the C-C phenol-coupling enzyme of morphine biosynthesis in opium poppy. J. Biol. Chem. **284,** 24432-24442.

Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Hagopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N., Park, C.C., Wen, P.-Z., Brewer, H., Weitz, K., Camp, D.G., Pan, C., Yordanova, R., Neuhaus, I., Tilford, C., Siemers, N., Gargalovic, P., Eskin, E., Kirchgessner, T., Smith, D.J., Smith, R.D., and Lusis, A.J. (2011). Comparative analysis of proteome and transcriptome variation in mouse. PLoS Genetics **7,** e1001393.

Giddings, L.A., Liscombe, D.K., Hamilton, J.P., Childs, K.L., DellaPenna, D., Buell, C.R., and O'Connor, S.E. (2011). A stereoselective hydroxylation step of alkaloid biosynthesis by a unique cytochrome P450 in *Catharanthus roseus*. J. Biol. Chem. **286,** 16751-16757.

Gnerre, S., Maccallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., Berlin, A.M., Aird, D., Costello, M., Daza, R., Williams, L., Nicol, R., Gnirke, A., Nusbaum, C., Lander, E.S., and Jaffe, D.B. (2011). High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc. Natl. Acad. Sci. U.S.A. **108,** 1513-1518.

Góngora-Castillo, E., Childs, K.L., Fedewa, G., Hamilton, J.P., Liscombe, D.K., Magallanes-Lundback, M., Mandadi, K.K., Nims, E., Runguphan, W., Vaillancourt, B., Varbanova-Herde, M., Dellapenna, D., McKnight, T.D., O'Connor, S., and Buell, C.R. (2012). Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species. PLoS One **7,** e52506.

Gordon, R.H. (2002). The biology of Narcissus. In Narcissus and Daffodil (CRC Press), pp. 5-18.

Greer, S., Wen, M., Bird, D., Wu, X., Samuels, L., Kunst, L., and Jetter, R. (2007). The cytochrome P450 enzyme CYP96A15 is the midchain alkane hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of Arabidopsis. Plant Physiol. **145,** 653-667.

Grobe, N., Zhang, B., Fisinger, U., Kutchan, T.M., Zenk, M.H., and Guengerich, F.P. (2009). Mammalian cytochrome P450 enzymes catalyze the phenol-coupling step in endogenous morphine biosynthesis. J. Biol. Chem. **284,** 24425-24431.

Guterman, I., Shalit, M., Menda, N., Piestun, D., Dafny-Yelin, M., Shalev, G., Bar, E., Davydov, O., Ovadis, M., Emanuel, M., Wang, J., Adam, Z., Pichersky, E., Lewinsohn, E., Zamir, D., Vainstein, A., and Weiss, D. (2002). Rose scent: genomics approach to discovering novel floral fragrance-related genes. Plant Cell **14,** 2325-2338.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., Macmanes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., Leduc, R.D., Friedman, N., and Regev, A. (2013). *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. **8,** 1494-1512.

Hackl, T., Hedrich, R., Schultz, J., and Förster, F. (2014). proovread: large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics **30,** 3004-3011.

Hampp, N., and Zenk, M.H. (1988). Homogeneous strictosidine synthase from cell suspension cultures of *Rauvolfia serpentina*. Phytochemistry **27,** 3811-3815.

Hansen, E.H., Møller, B.L., Kock, G.R., Bünner, C.M., Kristensen, C., Jensen, O.R., Okkels, F.T., Olsen, C.E., Motawia, M.S., and Hansen, J. (2009). *De novo* biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). Appl. Environ. Microbiol. **75,** 2765-2774.

Harken, R.D., Christensen, C.P., and Wildman, W.C. (1976). Interconversions in the pluviine-lycorenine series. J. Org. Chem. **41,** 2450-2454.

Hasemann, C.A., Kurumbail, R.G., Boddupalli, S.S., Peterson, J.A., and Deisenhofer, J. (1995). Structure and function of cytochromes P450: a comparative analysis of three crystal structures. Structure **3,** 41-62.

Hatae, T., Hara, S., Yokoyama, C., Yabuki, T., Inoue, H., Ullrich, V., and Tanabe, T. (1996). Site-directed mutagenesis of human prostacyclin synthase: Alteration of Cys441 of the Cys-pocket, and Glu347 and Arg350 of the EXXR motif. FEBS Lett. **389,** 268-272.

Havelek, R., Seifrtova, M., Kralovec, K., Bruckova, L., Cahlikova, L., Dalecka, M., Vavrova, J., Rezacova, M., Opletal, L., and Bilkova, Z. (2014). The effect of Amaryllidaceae alkaloids haemanthamine and haemanthidine on cell cycle progression and apoptosis in p53-negative human leukemic Jurkat cells. Phytomedicine **21,** 479-490.

Havkin-Frenkel, D., Podstolski, A., and Dixon, R. (2003). Vanillin biosynthetic pathway enzyme from *Vanilla planifolia* (Google Patents).

He, M., Qu, C., Gao, O., Hu, X., and Hong, X. (2015). Biological and pharmacological activities of Amaryllidaceae alkaloids. R.S.C. Adv. **5,** 16562-16574.

Herbert, R.B. (2003). The biosynthesis of plant alkaloids and nitrogenous microbial metabolites. Nat. Prod. Rep. **20,** 494-508.

Higashi, Y., Smith, T.J., Jez, J.M., and Kutchan, T.M. (2010). Crystallization and preliminary X-ray diffraction analysis of salutaridine reductase from the opium poppy *Papaver somniferum*. Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. **66,** 163-166.

Hopley, C., Bristow, T., Lubben, A., Simpson, A., Bull, E., Klagkou, K., Herniman, J., and Langley, J. (2008). Towards a universal product ion mass spectral library - reproducibility of product ion spectra across eleven different mass spectrometers. Rapid Commun. Mass Spectrom. **22,** 1779-1786.

Howes, M.-J.R., and Houghton, P.J. (2003). Plants used in Chinese and Indian traditional medicine for improvement of memory and cognitive function. Pharmacol. Biochem. Be. **75,** 513-527.

Hughes, A., Musher, J., Thomas, S., Beusterien, K., Strunk, B., and Arcona, S. (2004). Gastrointestinal Adverse Events in a General Population Sample of Nursing Home Residents Taking Cholinesterase Inhibitors. Consult. Pharm. **19,** 713-720.

Ibdah, M., Zhang, X.H., Schmidt, J., and Vogt, T. (2003). A novel Mg(2+)-dependent *O*-methyltransferase in the phenylpropanoid metabolism of *Mesembryanthemum crystallinum*. J. Biol. Chem. **278,** 43961-43972.

Ikezawa, N., Iwasa, K., and Sato, F. (2008). Molecular cloning and characterization of CYP80G2, a cytochrome P450 that catalyzes an intramolecular C-C phenol coupling of (*S*)-reticuline in magnoflorine biosynthesis, from cultured *Coptis japonica* cells. J. Biol. Chem. **283,** 8810-8821.

Ikezawa, N., Tanaka, M., Nagayoshi, M., Shinkyo, R., Sakaki, T., Inouye, K., and Sato, F. (2003). Molecular cloning and characterization of CYP719, a methylenedioxy bridge-forming enzyme that belongs to a novel P450 family, from cultured *Coptis japonica* cells. J. Biol. Chem. **278,** 38557-38565.

Irmler, S., Schröder, G., St-Pierre, B., Crouch, N.P., Hotze, M., Schmidt, J., Strack, D., Matern, U., and Schröder, J. (2000). Indole alkaloid biosynthesis in *Catharanthus roseus*: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase. Plant J. **24,** 797-804.

Irwin, R.L., and Smith, H.J. (1960). Cholinesterase inhibition by galanthamine and lycoramine. Biochem. Pharmacol. **3,** 147-148.

Iseli, C., Jongeneel, C.V., and Bucher, P. (1999). ESTScan: a program for detecting, evaluating, and reconstructing potential coding regions in EST sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol.**,** 138-148.

Itkin, M., Heinig, U., Tzfadia, O., Bhide, A.J., Shinde, B., Cardenas, P.D., Bocobza, S.E., Unger, T., Malitsky, S., Finkers, R., Tikunov, Y., Bovy, A., Chikate, Y., Singh, P., Rogachev, I., Beekwilder, J., Giri, A.P., and Aharoni, A. (2013). Biosynthesis of antinutritional alkaloids in solanaceous crops is mediated by clustered genes. Science **341,** 175-179.

Jeffs, P. (1962). The Alkaloids of the Amaryllidaceae. Part X.* Biosynthesis of haemanthamine. Proc. Chem. Soc.**,** 80-81.

Jiang, J., He, X., and Cane, D.E. (2006). Geosmin biosynthesis. *Streptomyces coelicolor* germacradienol/germacrene D synthase converts farnesyl diphosphate to geosmin. J. Am. Chem. Soc. **128,** 8128-8129.

Jiang, Y., Xia, B., Liang, L., Li, X., Xu, S., Peng, F., and Wang, R. (2013). Molecular and analysis of a phenylalanine ammonia-lyase gene (LrPAL2) from *Lycoris radiata*. Mol. Biol. Rep. **40,** 2293-2300.

Jiang, Y., Xia, N., Li, X., Shen, W., Liang, L., Wang, C., Wang, R., Peng, F., and Xia, B. (2011). Molecular cloning and characterization of a phenylalanine ammonia-lyase gene (LrPAL) from *Lycoris radiata*. Mol. Biol. Rep. **38,** 1935-1940.

Jin, Z. (2009). Amaryllidaceae and Sceletium alkaloids. Nat. Prod. Rep. **26,** 363-381.

Jin, Z. (2013). Amaryllidaceae and Sceletium alkaloids. Nat. Prod. Rep. **30,** 849-868.

Johnson, M.T., Carpenter, E.J., Tian, Z., Bruskiewich, R., Burris, J.N., Carrigan, C.T., Chase, M.W., Clarke, N.D., Covshoff, S., Depamphilis, C.W., Edger, P.P., Goh, F., Graham, S., Greiner, S., Hibberd, J.M., Jordon-Thaden, I., Kutchan, T.M., Leebens-Mack, J., Melkonian, M., Miles, N., Myburg, H., Patterson, J., Pires, J.C., Ralph, P., Rolf, M., Sage, R.F., Soltis, D., Soltis, P., Stevenson, D., Stewart, C.N., Surek, B., Thomsen, C.J., Villarreal, J.C., Wu, X., Zhang, Y., Deyholos, M.K., and Wong, G.K. (2012). Evaluating methods for isolating total RNA and predicting the success of sequencing phylogenetically diverse plant transcriptomes. PLoS One **7,** e50226.

Jörnvall, H., Persson, B., Krook, M., Atrian, S., Gonzàlez-Duarte, R., Jeffery, J., and Ghosh, D. (1995). Short-chain dehydrogenases/reductases (SDR). Biochemistry **34,** 6003-6013.

Joshi, C.P., and Chiang, V.L. (1998). Conserved sequence motifs in plant *S*-adenosyl-*L*-methionine-dependent methyltransferases. Plant. Mol. Biol. **37,** 663-674.

Kallberg, Y., Oppermann, U., Jörnvall, H., and Persson, B. (2002). Short-chain dehydrogenases/reductases (SDRs). Eur. J. Biochem. **269,** 4409-4417.

Kalra, S., Puniya, B.L., Kulshreshtha, D., Kumar, S., Kaur, J., Ramachandran, S., and Singh, K. (2013). *De novo* transcriptome sequencing reveals important molecular networks and metabolic pathways of the plant, *Chlorophytum borivilianum*. PLoS One **8,** e83336.

Kamenetsky, R., Faigenboim, A., Shemesh Mayer, E., Ben Michael, T., Gershberg, C., Kimhi, S., Esquira, I., Rohkin Shalom, S., Eshel, D., Rabinowitch, H.D., and Sherman, A. (2015). Integrated transcriptome catalogue and organ-specific profiling of gene expression in fertile garlic (*Allium sativum* L.). BMC Genomics **16,** 12.

Kato, M., Mizuno, K., Crozier, A., Fujimura, T., and Ashihara, H. (2000). Caffeine synthase gene from tea leaves. Nature **406,** 956-957.

Kaufmann, K., Pajoro, A., and Angenent, G.C. (2010). Regulation of transcription in plants: mechanisms controlling developmental switches. Nat. Rev. Genet. **11,** 830-842.

Kavanagh, K.L., Jörnvall, H., Persson, B., and Oppermann, U. (2008). Medium- and short-chain dehydrogenase/reductase gene and protein families: the SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. Cell Mol. Life Sci. **65,** 3895-3906.

Kellner, F., Kim, J., Clavijo, B.J., Hamilton, J.P., Childs, K.L., Vaillancourt, B., Cepela, J., Habermann, M., Steuernagel, B., Clissold, L., McLay, K., Buell, C.R., and O'Connor, S.E. (2015). Genome-guided investigation of plant natural product biosynthesis. Plant J. **82,** 680-692.

Kilgore, M.B., Augustin, M.M., Starks, C.M., O'Neil-Johnson, M., May, G.D., Crow, J.A., and Kutchan, T.M. (2014). Cloning and characterization of a norbelladine 4'-*O*-methyltransferase involved in the biosynthesis of the Alzheimer's drug galanthamine in *Narcissus* sp. *aff. pseudonarcissus*. PLoS One **9,** e103223.

Kim, D.H., Kim, B.G., Lee, Y., Ryu, J.Y., Lim, Y., Hur, H.G., and Ahn, J.H. (2005). Regiospecific methylation of naringenin to ponciretin by soybean *O*-methyltransferase expressed in *Escherichia coli*. J. Biotechnol. **119,** 155-162.

Kirby, G.W., and Tiwari, H.P. (1966). Phenol oxidation and biosynthesis. Part IX.* The biosynthesis of norpluviine and galanthine. J. Chem. Soc. (C)**,** 676-682.

Kopycki, J.G., Stubbs, M.T., Brandt, W., Hagemann, M., Porzel, A., Schmidt, J., Schliemann, W., Zenk, M.H., and Vogt, T. (2008). Functional and structural characterization of a cation-dependent *O*-methyltransferase from the cyanobacterium *Synechocystis* sp. strain PCC 6803. J. Biol. Chem. **283,** 20888-20896.

Kornienko, A., and Evidente, A. (2008). Chemistry, biology, and medicinal potential of narciclasine and its congeners. Chem. Rev. **108,** 1982-2014.

Krauss, M., Singer, H., and Hollender, J. (2010). LC-high resolution MS in environmental analysis: from target screening to the identification of unknowns. Anal. Bioanal. Chem. **397,** 943-951.

Kumari, S., Nie, J., Chen, H.S., Ma, H., Stewart, R., Li, X., Lu, M.Z., Taylor, W.M., and Wei, H. (2012). Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. PLoS One **7,** e50411.

Kutchan, T.M., Bock, A., and Dittrich, H. (1994). Heterologous expression of the plant proteins strictosidine synthase and berberine bridge enzyme in insect cell culture. Phytochemistry **35,** 353-360.

Kwantes, M., Liebsch, D., and Verelst, W. (2012). How MIKC* MADS-box genes originated and evidence for their conserved function throughout the evolution of vascular plant gametophytes. Mol. Biol. Evol. **29,** 293-302.

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics **Chapter:11,** Unit 11.17.

Larsen, P.A., Heilman, A.M., and Yoder, A.D. (2014). The utility of PacBio circular consensus sequencing for characterizing complex gene families in non-model organisms. BMC Genomics **15,** 720.

Latvala, A., Linden, A., Hesse, M., Önür, M.A., Gözler, T., and Kivçak, B. (1995). Alkaloids of *Galanthus elwesii*. Phytochemistry **39,** 1229-1240.

Lee, H., Gurtowski, J., Yoo, S., Marcus, S., McCombie, W.R., and Schatz, M. (2014). Error correction and assembly complexity of single molecule sequencing reads. (bioRxiv).

Lehmann, T., and Pollmann, S. (2009). Gene expression and characterization of a stress-induced tyrosine decarboxylase from *Arabidopsis thaliana*. FEBS Lett. **583,** 1895-1900.

Lester, D.R., Ross, J.J., Davies, P.J., and Reid, J.B. (1997). Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. Plant Cell **9,** 1435-1443.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25,** 1754-1760.

Li, R., Wang, M.Y., and Li, X.B. (2012). Chemical constituents and biological activities of genus *Hosta* (Liliaceae). J. Med. Plant. Res. **6,** 2704-2713.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., Yang, H., and Wang, J. (2010). *De novo* assembly of human genomes with massively parallel short read sequencing. Genome Res. **20,** 265-272.

Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22,** 1658-1659.

Li, Y., Liu, J., Tang, L.J., Shi, Y.W., Ren, W., and Hu, W.X. (2007). Apoptosis induced by lycorine in KM3 cells is associated with the G0/G1 cell cycle arrest. Oncol. Rep. **17,** 377-384.

Liu, J., Hu, W.X., He, L.F., Ye, M., and Li, Y. (2004). Effects of lycorine on HL-60 cells via arresting cell cycle and inducing apoptosis. FEBS Lett. **578,** 245-250.

Liu, R., Loraine, A.E., and Dickerson, J.A. (2014). Comparisons of computational methods for differential alternative splicing detection using RNA-seq in plant systems. BMC Bioinformatics **15,** 364.

Liu, X.S., Jiang, J., Jiao, X.Y., Wu, Y.E., Lin, J.H., and Cai, Y.M. (2009). Lycorine induces apoptosis and down-regulation of Mcl-1 in human leukemia cells. Cancer Lett. **274,** 16-24.

Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods **25,** 402-408.

Loman, N.J., and Watson, M. (2015). Successful test launch for nanopore sequencing. Nat. Methods. **12,** 303-304.

Lottaz, C., Iseli, C., Jongeneel, C.V., and Bucher, P. (2003). Modeling sequencing errors by combining Hidden Markov models. Bioinformatics **19 Suppl 2,** ii103-112.

Louw, C.A.M., Regnier, T.J.C., and Korsten, L. (2002). Medicinal bulbous plants of South Africa and their traditional relevance in the control of infectious diseases. J. Ethnopharmacol. **82,** 147-154.

Lu, G., Zou, Q., Guo, D., Zhuang, X., Yu, X., Xiang, X., and Cao, J. (2007). *Agrobacterium tumefaciens*-mediated transformation of *Narcissus tazzeta* var. *chinensis*. Plant Cell Rep. **26,** 1585-1593.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D.W., Yiu, S.M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., and Lam, T.W. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. Gigascience **1,** 18.

Machocho, A.K., Bastida, J., Codina, C., Viladomat, F., Brun, R., and Chhabra, S.C. (2004). Augustamine type alkaloids from *Crinum kirkii*. Phytochemistry **65,** 3143-3149.

Madoui, M.A., Engelen, S., Cruaud, C., Belser, C., Bertrand, L., Alberti, A., Lemainque, A., Wincker, P., and Aury, J.M. (2015). Genome assembly using Nanopore-guided long and error-free DNA reads. BMC Genomics **16,** 327.

Mahrous, E.A., and Farag, M.A. (2015). Two dimensional NMR spectroscopic approaches for exploring plant metabolome: A review. J. Adv. Res. **6,** 3-15.

Mann, J.D. (1963). Alkaloids and plant metabolism. VI. *O*-Methylation *in vitro* of norbelladine, a precursor of Amaryllidaceae alkaloids. J. Biol. Chem. **238,** 3820-3823.

Mathieu, Y., Guern, J., Kurkdjian, A., Manigault, P., Manigault, J., Zielinska, T., Gillet, B., Beloeil, J.-C., and Lallemand, J.-Y. (1989). Regulation of Vacuolar pH of Plant Cells: I. Isolation and Properties of Vacuoles Suitable for ^{31}P NMR Studies. Plant Physiol. **89,** 19-26.

McCaskill, D., Gershenzon, J., and Croteau, R. (1992). Morphology and monoterpene biosynthetic capabilities of secretory cell clusters isolated from glandular trichomes of peppermint (*Mentha piperita* L.). Planta **187,** 445-454.

McCoy, R.C., Taylor, R.W., Blauwkamp, T.A., Kelley, J.L., Kertesz, M., Pushkarev, D., Petrov, D.A., and Fiston-Lavier, A.S. (2014). Illumina TruSeq synthetic long-reads empower *de novo* assembly and resolve complex, highly-repetitive transposable elements. PLoS One **9,** e106689.

Mellway, R.D., Tran, L.T., Prouse, M.B., Campbell, M.M., and Constabel, C.P. (2009). The wound-, pathogen-, and ultraviolet B-responsive MYB134 gene encodes an R2R3 MYB transcription factor that regulates proanthocyanidin synthesis in poplar. Plant Physiol. **150,** 924-941.

Mercke, P., Kappers, I.F., Verstappen, F.W., Vorst, O., Dicke, M., and Bouwmeester, H.J. (2004). Combined transcript and metabolite analysis reveals genes involved in spider mite induced volatile formation in cucumber plants. Plant Physiol. **135,** 2012-2024.

Mikheyev, A.S., and Tin, M.M. (2014). A first look at the Oxford Nanopore MinION sequencer. Mol. Ecol. Resour. **14,** 1097-1102.

Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006). HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. Acta. Crystallogr. D Biol. Crystallogr. **62,** 859-866.

Mizutani, M., and Sato, F. (2011). Unusual P450 reactions in plant secondary metabolism. Arch. Biochem. Biophys. **507,** 194-203.

Mockler, T.C., Michael, T.P., Priest, H.D., Shen, R., Sullivan, C.M., Givan, S.A., McEntee, C., Kay, S.A., and Chory, J. (2007). The DIURNAL project: DIURNAL and circadian expression profiling, model-based pattern matching, and promoter analysis. Cold Spring Harb. Symp. Quant. Biol. **72,** 353-363.

Moreau, Y., and Tranchevent, L.C. (2012). Computational tools for prioritizing candidate genes: boosting disease gene discovery. Nat. Rev. Genet. **13,** 523-536.

Morishige, T., Dubouzet, E., Choi, K.B., Yazaki, K., and Sato, F. (2002). Molecular cloning of columbamine *O*-methyltransferase from cultured *Coptis japonica* cells. Eur. J. Biochem. **269,** 5659-5667.

Na, X., Hu, Y., Yue, K., Lu, H., Jia, P., Wang, H., Wang, X., and Bi, Y. (2011). Narciclasine modulates polar auxin transport in Arabidopsis roots. J. Plant Physiol. **168,** 1149-1156.

Nelson, D., and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. Plant J. **66,** 194-211.

Nomura, T., and Kutchan, T.M. (2010). Three new *O*-methyltransferases are sufficient for all *O*-methylation reactions of ipecac alkaloid biosynthesis in root culture of *Psychotria ipecacuanha*. J. Biol. Chem. **285,** 7722-7738.

Nützmann, H.W., and Osbourn, A. (2015). Regulation of metabolic gene clusters in *Arabidopsis thaliana*. New Phytol. **205,** 503-510.

Okada, A., Okada, K., Miyamoto, K., Koga, J., Shibuya, N., Nojiri, H., and Yamane, H. (2009). OsTGAP1, a bZIP transcription factor, coordinately regulates the inductive production of diterpenoid phytoalexins in rice. J. Biol. Chem. **284,** 26510-26518.

Oloyede, G.K., Oke, J.M., Raji, Y., and Olugbade, T.A. (2010). Antioxidant and Anticonvulsant Alkaloids in *Crinum ornatum* Bulb Extract. World J. Chem. **5,** 26-31.

Ono, E., Nakai, M., Fukui, Y., Tomimori, N., Fukuchi-Mizutani, M., Saito, M., Satake, H., Tanaka, T., Katsuta, M., Umezawa, T., and Tanaka, Y. (2006). Formation of two methylenedioxy bridges by a Sesamum CYP81Q protein yielding a furofuran lignan, (+)-sesamin. Proc. Natl. Acad. Sci. U.S.A. **103,** 10116-10121.

Ounaroon, A., Decker, G., Schmidt, J., Lottspeich, F., and Kutchan, T.M. (2003). (*R,S*)-Reticuline 7-*O*-methyltransferase and (*R,S*)-norcoclaurine 6-*O*-methyltransferase of *Papaver somniferum* - cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. Plant J. **36,** 808-819.

Pabuççuoglu, V., Richomme, P., Gözler, T., Kivçak, B., Freyer, A.J., and Shamma, M. (1989). Four New Crinine-Type Alkaloids from *Sternbergia Species*. J. Nat. Prod. **52,** 785-791.

Park, J.B. (2014). Synthesis and characterization of norbelladine, a precursor of Amaryllidaceae alkaloid, as an anti-inflammatory/anti-COX compound. Bioorg. Med. Chem. Lett. **24,** 5381-5384.

Penning, T.M. (2015). The aldo-keto reductases (AKRs): Overview. Chem. Biol. Interact. **234,** 236-246.

Persson, B., and Kallberg, Y. (2013). Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). Chem. Biol. Interact. **202,** 111-115.

Pichersky, E., and Lewinsohn, E. (2011). Convergent evolution in plant specialized metabolism. Annu. Rev. Plant. Biol. **62,** 549-566.

Plaitakis, A., and Duvoisin, R.C. (1983). Homer's Moly Identified as *Galanthus nivalis* L.: Physiologic Antidote to Stramonium Poisoning. Clin. Neuropharmacol. **6,** 1-6.

Porté, S., Xavier Ruiz, F., Giménez, J., Molist, I., Alvarez, S., Domínguez, M., Alvarez, R., de Lera, A.R., Parés, X., and Farrés, J. (2013). Aldo-keto reductases in retinoid metabolism: search for substrate specificity and inhibitor selectivity. Chem. Biol. Interact. **202,** 186-194.

Prachayasittikul, S., Buraparuangsang, P., Worachartcheewan, A., Isarankura-Na-Ayudhya, C., Ruchirawat, S., and Prachayasittikul, V. (2008). Antimicrobial and antioxidative activities of bioactive constituents from *Hydnophytum formicarum* Jack. Molecules **13,** 904-921.

Rajkumar, H., Ramagoni, R.K., Anchoju, V.C., Vankudavath, R.N., and Syed, A.U.Z. (2015). *De novo* transcriptome analysis of *Allium cepa* L. (onion) bulb to identify allergens and epitopes. PLoS ONE **10,** e0135387.

Raman, S.B., and Rathinasabapathi, B. (2003). beta-alanine *N*-methyltransferase of *Limonium latifolium*. cDNA cloning and functional expression of a novel *N*-methyltransferase implicated in the synthesis of the osmoprotectant beta-alanine betaine. Plant Physiol. **132,** 1642-1651.

Ravaglia, D., Espley, R.V., Henry-Kirk, R.A., Andreotti, C., Ziosi, V., Hellens, R.P., Costa, G., and Allan, A.C. (2013). Transcriptional regulation of flavonoid biosynthesis in

nectarine (*Prunus persica*) by a set of R2R3 MYB transcription factors. BMC Plant Biology **13,** 68-68.

Rees, A.R. (1969). The initiation and growth of *Narcissus* bulbs. Ann. Bot. **33,** 277-288.

Reinhardt, N., Fischer, J., Coppi, R., Blum, E., Brandt, W., and Dräger, B. (2014). Substrate flexibility and reaction specificity of tropinone reductase-like short-chain dehydrogenases. Bioorg. Chem. **53,** 37-49.

Roberts, M.F. (1971). Polyphenolases in the 1000g fraction of *Papaver somniferum* latex. Phytochemistry **10,** 3021-3027.

Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J., Hoodless, P.A., and Birol, I. (2010). *De novo* assembly and analysis of RNA-seq data. Nat. Methods. **7,** 909-912.

Rønsted, N., Symonds, M.R., Birkholm, T., Christensen, S.B., Meerow, A.W., Molander, M., Mølgaard, P., Petersen, G., Rasmussen, N., van Staden, J., Stafford, G.I., and Jäger, A.K. (2012). Can phylogeny predict chemical diversity and potential medicinal activity of plants? A case study of Amaryllidaceae. BMC Evol. Biol. **12,** 182.

Rossmann, M.G., Lijas, A., Brandén, C.I., and Babaszak, L.J. (1975). Evolutionary and structural relationships among dehydrogenases. In The Enzymes, 3rd edn., Boyer, ed (New York: Academic Press), pp. 61-102.

Saito, K., Hirai, M.Y., and Yonekura-Sakakibara, K. (2008). Decoding genes with coexpression networks and metabolomics - 'majority report by precogs'. Trends. Plant. Sci. **13,** 36-43.

Salmela, L., and Rivals, E. (2014). LoRDEC: accurate and efficient long read error correction. Bioinformatics **30,** 3506-3514.

Samanani, N., and Facchini, P.J. (2002). Purification and characterization of norcoclaurine synthase. The first committed enzyme in benzylisoquinoline alkaloid biosynthesis in plants. J. Biol. Chem. **277,** 33878-33883.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Biotech. **32,** 347-355.

Schlauer, J., Rückert, M., Wiesen, B., Herderich, M., Assi, L.A., Haller, R.D., Bär, S., Fröhlich, K.U., and Bringmann, G. (1998). Characterization of enzymes from Ancistrocladus (Ancistrocladaceae) and Triphyophyllum (Dioncophyllaceae) catalyzing oxidative coupling of naphthylisoquinoline alkaloids to michellamines. Arch. Biochem. Biophys. **350,** 87-94.

Schoch, G., Goepfert, S., Morant, M., Hehn, A., Meyer, D., Ullmann, P., and Werck-Reichhart, D. (2001). CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. J. Biol. Chem. **276,** 36566-36574.

Schröder, G., Wehinger, E., Lukacin, R., Wellmann, F., Seefelder, W., Schwab, W., and Schröder, J. (2004). Flavonoid methylation: a novel 4'-*O*-methyltransferase from *Catharanthus roseus*, and evidence that partially methylated flavanones are substrates of four different flavonoid dioxygenases. Phytochemistry **65,** 1085-1094.

Schulz, M.H., Zerbino, D.R., Vingron, M., and Birney, E. (2012). Oases: robust *de novo* RNA-seq assembly across the dynamic range of expression levels. Bioinformatics **28,** 1086-1092.

Schweigert, N., Zehnder, A.J., and Eggen, R.I. (2001). Chemical properties of catechols and their molecular modes of toxic action in cells, from microorganisms to mammals. Environ. Microbiol. **3,** 81-91.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J., and Birol, I. (2009). ABySS: a parallel assembler for short read sequence data. Genome Res. **19,** 1117-1123.

Spannagl, M., Martis, M.M., Pfeifer, M., Nussbaumer, T., and Mayer, K.F. (2013). Analysing complex Triticeae genomes - concepts and strategies. Plant Methods **9,** 35.

St-Pierre, B., Vazquez-Flota, F.A., and De Luca V. (1999). Multicellular compartmentation of *catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. Plant Cell **11,** 887-900.

Stavrinides, A., Tatsis, E.C., Foureau, E., Caputi, L., Kellner, F., Courdavault, V., and O'Connor, S.E. (2015). Unlocking the diversity of alkaloids in *Catharanthus roseus*: nuclear localization suggests metabolic channeling in secondary metabolism. Chem. Biol. **22,** 336-341.

Suhadolnik, R.J., Fischer, A.G., and Zulalian, J. (1962). The biogenic origin of the C_6-C_1 unit of lycorine. J. Am. Chem. Soc. **84,** 4348-4349.

Suhadolnik, R.J., Fischer, A.G., and Zulalian, J. (1963a). Biogenesis of the Amaryllidaceae alkaloids. II. Studies with whole plants, floral primordia and cell free extracts. Biochem. Biophys. Res. Commun. **11,** 208-212.

Suhadolnik, R.J., Fischer, A.G., and Zulalian, J. (1963b). Biogenesis of the Amaryllidaceae alkaloids. Part III. Phenylalanine and protocatechuic aldehyde as C_6-C_1 precursors of haemanthamine and lycorine. Pro. Chem. Soc.**,** 132.

Sumner, L.W., Lei, Z., Nikolau, B.J., and Saito, K. (2015). Modern plant metabolomics: advanced natural product gene discoveries, improved technologies, and future prospects. Nat. Prod. Rep. **32,** 212-229.

Sun, Y., Luo, H., Li, Y., Sun, C., Song, J., Niu, Y., Zhu, Y., Dong, L., Lv, A., Tramontano, E., and Chen, S. (2011). Pyrosequencing of the *Camptotheca acuminata* transcriptome reveals putative genes involved in camptothecin biosynthesis and transport. BMC Genomics **12,** 533.

Syed, K., and Mashele, S.S. (2014). Comparative analysis of P450 signature motifs EXXR and CXG in the large and diverse kingdom of fungi: identification of evolutionarily conserved amino acid patterns characteristic of P450 family. PLoS One **9,** e95616.

Takeshita, N., Fujiwara, H., Mimura, H., Fitchen, J.H., Yamada, Y., and Sato, F. (1995). Molecular cloning and characterization of *S*-adenosyl-L-methionine: scoulerine-9-*O*-methyltransferase from cultured cells of *Coptis japonica*. Plant Cell Physiol. **36,** 29-36.

Takos, A.M., and Rook, F. (2012). Why biosynthetic genes for chemical defense compounds cluster. Trends Plant Sci. **17,** 383-388.

Takos, A.M., and Rook, F. (2013). Towards a molecular understanding of the biosynthesis of Amaryllidaceae alkaloids in support of their expanding medical use. Int. J. Mol. Sci. **14,** 11713-11741.

Takos, A.M., Knudsen, C., Lai, D., Kannangara, R., Mikkelsen, L., Motawia, M.S., Olsen, C.E., Sato, S., Tabata, S., Jørgensen, K., Møller, B.L., and Rook, F. (2011). Genomic clustering of cyanogenic glucoside biosynthetic genes aids their identification in *Lotus japonicus* and suggests the repeated evolution of this chemical defence pathway. Plant J. **68,** 273-286.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28,** 2731-2739.

Tanaka, Y., Matsuoka, M., Yamanoto, N., Ohashi, Y., Kano-Murakami, Y., and Ozeki, Y. (1989). Structure and characterization of a cDNA clone for phenylalanine ammonia-lyase from cut-injured roots of sweet potato. Plant Physiol. **90,** 1403-1407.

Teutsch, H.G., Hasenfratz, M.P., Lesot, A., Stoltz, C., Garnier, J.M., Jeltsch, J.M., Durst, F., and Werck-Reichhart, D. (1993). Isolation and sequence of a cDNA encoding the Jerusalem artichoke cinnamate 4-hydroxylase, a major plant cytochrome P450 involved in the general phenylpropanoid pathway. Proc. Natl. Acad. Sci. U.S.A. **90,** 4102-4106.

Tian, S., Zi, W., and Ma, D. (2012). Potentially biomimetic total synthesis and relative stereochemical assignment of (±)-gracilamine. Angew. Chem. Int. Ed. **51,** 10141-10144.

Tsukazaki, H., Yaguchi, S., Sato, S., Hirakawa, H., Katayose, Y., Kanamori, H., Kurita, K., Itoh, T., Kumagai, M., Mizuno, S., Hamada, M., Fukuoka, H., Yamashita, K.-i., McCallum, J., Shigyo, M., and Wako, T. (2015). Development of transcriptome shotgun assembly-derived markers in bunching onion (*Allium fistulosum*). Mol. Breeding **35,** 1-11.

Unver, N., Kaya, G.I., Werner, C., Verpoorte, R., and Gözler, B. (2003). Galanthindole: a new indole alkaloid from *Galanthus plicatus* ssp. *byzantinus*. Planta Med. **69,** 869-871.

Ünver, N., and Kaya, G.I. (2005). An unusual pentacyclic dinitrogenous alkaloid from *Galanthus gracilis*. Turk. J. Chem. **29,** 547-553.

Ünver, N., Gözler, T., Walch, N., Gözler, B., and Hesse, M. (1999). Two novel dinitrogenous alkaloids from *Galanthus plicatus* subsp. *byzantinus* (Amaryllidaceae). Phytochemistry **50,** 1255-1261.

Uyeo, S., and Kobayashi, S. (1953). Lycoris alkaloids. XXIV. Isolation and characterization of lycoremine. Pharm. Bull. **1,** 139-142.

Van Goietsenoven, G., Andolfi, A., Lallemand, B., Cimmino, A., Lamoral-Theys, D., Gras, T., Abou-Donia, A., Dubois, J., Lefranc, F., Mathieu, V., Kornienko, A., Kiss, R., and Evidente, A. (2010). Amaryllidaceae alkaloids belonging to different structural subgroups display activity against apoptosis-resistant cancer cells. J. Nat. Prod. **73,** 1223-1227.

Velten, R., Erdelen, C., Gehling, M., Göhrt, A., Gondol, D., Lenz, J., Lockhoff, O., Wachendorff, U., and Wendisch, D. (1998). Cripowellin A and B, a novel type of Amaryllidaceae alkaloid from *Crinum powellii*. Tet. Lett. **39,** 1737-1740.

Wang, R., Xu, S., Jiang, Y., Jiang, J., Li, X., Liang, L., He, J., Peng, F., and Xia, B. (2013). *De novo* sequence assembly and characterization of Lycoris aurea transcriptome using GS FLX titanium platform of 454 pyrosequencing. PLoS One **8,** e60449.

Wang, Y.H., Gao, S., Yang, F.M., Sun, Q.Y., Wang, J.S., Liu, H.Y., Li, C.S., Di, Y.T., Li, S.L., He, H.P., and Hao, X.J. (2007). Structure elucidation and biomimetic synthesis of hostasinine A, a new benzylphenethylamine alkaloid from *Hosta plantaginea*. Org. Lett. **9,** 5279-5281.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. **10,** 57-63.

Wei, H., Persson, S., Mehta, T., Srinivasasainagendra, V., Chen, L., Page, G.P., Somerville, C., and Loraine, A. (2006). Transcriptional coordination of the metabolic network in Arabidopsis. Plant Physiol. **142,** 762-774.

Wetzl, D., Berrera, M., Sandon, N., Fishlock, D., Ebeling, M., Müller, M., Hanlon, S.P., Wirz, B., and Iding, H. (2015). Expanding the imine reductase toolbox by exploring the bacterial protein sequence space. ChemBioChem **16,** 1749-1756.

Widhalm, J.R., and Dudareva, N. (2015). A familiar ring to it: biosynthesis of plant benzoic acids. Mol. Plant **8,** 83-97.

Widiez, T., Hartman, T.G., Dudai, N., Yan, Q., Lawton, M., Havkin-Frenkel, D., and Belanger, F.C. (2011). Functional characterization of two new members of the caffeoyl

CoA *O*-methyltransferase-like gene family from *Vanilla planifolia* reveals a new class of plastid-localized *O*-methyltransferases. Plant Molecular Biology **76,** 475-488.

Wilcock, G., Howe, I., Coles, H., Lilienfeld, S., Truyen, L., Zhu, Y., Bullock, R., Kershaw, P., and Group, G.-G.-S. (2003). A long-term comparison of galantamine and donepezil in the treatment of Alzheimer's disease. Drugs Aging **20,** 777-789.

Wilcock, G.K., Lilienfeld, S., and Gaens, E. (2000). Efficacy and safety of galantamine in patients with mild to moderate Alzheimer's disease: multicentre randomised controlled trial. Galantamine International-1 Study Group. BMJ **321,** 1445-1449.

Wildman, W.C., and Heimer, N.E. (1967). Alkaloid biosynthesis and interconversions. The conversion of caranine to lycorine. J. Am. Chem. Soc. **89,** 5265-5269.

Wildman, W.C., and Bailey, D.T. (1967). Pretazettine. J. Am. Chem. Soc. **89,** 5514-5515.

Wildman, W.C., and Bailey, D.T. (1969). Amaryllidaceae interconversions. Partial synthesis of [2]benzopyran[3,4c]indoles. J. Am. Chem. Soc. **91,** 150-157.

Wildman, W.C., Fales, H.M., and Battersby, A.R. (1962a). Biosynthesis in the Amaryllidaceae. The incorporation of 3-C^{14}-tyrosine in *Sprekelia formosissima*. J. Am. Chem. Soc. **84,** 681-682.

Wildman, W.C., Battersby, A.R., and Breuer, S.W. (1962b). Biosynthesis in the Amaryllidaceae. Incorporation of 3-C^{14}-tyrosine and phenylalanine in *Nerine bowdenii* W. Wats. J. Am. Chem. Soc. **84,** 4599-4600.

Wildman, W.C., Fales, H.M., Highet, R.J., Breuer, S.W., and Battersby, A.R. (1962c). Biosynthesis in the Amaryllidaceae: Evidence for intact incorporation of norbelladine into lycorine, crinamine, and belladine. Pro. Chem. Soc.**,** 180-181.

Wils, C.R., Brandt, W., Manke, K., and Vogt, T. (2013). A single amino acid determines position specificity of an *Arabidopsis thaliana* CCoAOMT-like *O*-methyltransferase. FEBS Lett. **587,** 683-689.

Yamashita, A., Kato, H., Wakatsuki, S., Tomizaki, T., Nakatsu, T., Nakajima, K., Hashimoto, T., Yamada, Y., and Oda, J. (1999). Structure of tropinone reductase-II complexed with NADP+ and pseudotropine at 1.9 A resolution: implication for stereospecific substrate binding and catalysis. Biochemistry **38,** 7630-7637.

Yamazaki, M., Mochida, K., Asano, T., Nakabayashi, R., Chiba, M., Udomson, N., Yamazaki, Y., Goodenowe, D.B., Sankawa, U., Yoshida, T., Toyoda, A., Totoki, Y., Sakaki, Y., Góngora-Castillo, E., Buell, C.R., Sakurai, T., and Saito, K. (2013). Coupling deep transcriptome analysis with untargeted metabolic profiling in *Ophiorrhiza pumila* to further the understanding of the biosynthesis of the anti-cancer alkaloid camptothecin and anthraquinones. Plant Cell Physiol. **54,** 686-696.

Yeo, Y.S., Nybo, S.E., Chittiboyina, A.G., Weerasooriya, A.D., Wang, Y.H., Góngora-Castillo, E., Vaillancourt, B., Buell, C.R., DellaPenna, D., Celiz, M.D., Jones, A.D., Wurtele, E.S., Ransom, N., Dudareva, N., Shaaban, K.A., Tibrewal, N., Chandra, S., Smillie, T., Khan, I.A., Coates, R.M., Watt, D.S., and Chappell, J. (2013). Functional identification of valerena-1,10-diene synthase, a terpene synthase catalyzing a unique chemical cascade in the biosynthesis of biologically active sesquiterpenes in *Valeriana officinalis*. J. Biol. Chem. **288,** 3163-3173.

Zerbino, D.R., and Birney, E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18,** 821-829.

Zhang, H., and Ge, Y. (2011). Comprehensive analysis of protein modifications by top-down mass spectrometry. Circ. Cardiovasc. Genet. **4,** 711-711.

Zonneveld, B. (2010). The involvement of *Narcissus hispanicus* Gouan in the origin of *Narcissus bujei* and of cultivated trumpet daffodils (Amaryllidaceae). Anales del Jardín Botánico de Madrid **67,** 29-39.

Appendix A: The Identification and Confirmation of the *Narcissus* sp. *aff. pseudonarcissus* Tyrosine Decarboxylase

A.1 Introduction, Results, and Discussion

Tyrosine decarboxylase (TyrDC) is the enzyme responsible for the conversion of tyrosine to tyramine in an early stage of the Amaryllidaceae alkaloid biosynthetic pathway. This enzyme was previously discovered in other plants (Lehmann and Pollmann, 2009). To confirm the presence of a similar enzyme in *Narcissus*, homologues to a previously characterized *TyrDC* from *Arabidopsis thaliana* were identified and tested for enzymatic activity. A *TyrDC* homologue was found to encode an enzyme with tyrosine decarboxylase as shown in Figure A.1 and is deposited in GenBank under the Accession Number KT378599.

A.2 Methods

A.2.1 PCR and Cloning

Nested PCR reactions were used to amplify the *TyrDC* as in as for *N4OMT* in Chapter 2 except *TyrDC* specific primers were used for outer and inner PCR, the outer PCR consisted of 30 cycles with a T_m of 52 ºC, and the inner PCR consisted of 30 cycles with a T_m of 54 ºC. Digestion with NotI and BamHI with subsequent ligation into pET28a and transformation into DH5a for plasmid amplification are also as in Chapter 2. The resulting plasmid was purified, sequenced, and transformed into BL21 Rosetta cells as in Chapter 2.

A.2.2 Protein Purification, Enzyme Assays, and HPLC

The protein was purified by inducing the transformed *E. coli* overnight at 16 ºC with 1 mM IPTG and subsequent His-tag purification as in Chapter 2. The protein concentration was determined by the Bradford assay and the purity evaluated with SDS-PAGE as in Chapter 2. The enzyme assays are based on the work of Lehmann and Pollmann with 100 mM potassium phosphate pH 7.5, 5 µg pure TyrDC, 2 mM pyridoxal 5'-phosphate, and 200 µM tyrosine at 30 ºC for 2 hr (Lehmann and Pollmann, 2009). The assays were extracted by adjusting the pH to 10 with the addition of two volumes 1 M $NaCO_2$ and extracting with two volumes ethyl acetate two times as in Chapter 2. Samples were analyzed using the same HPLC system and program as described in Chapter 2.

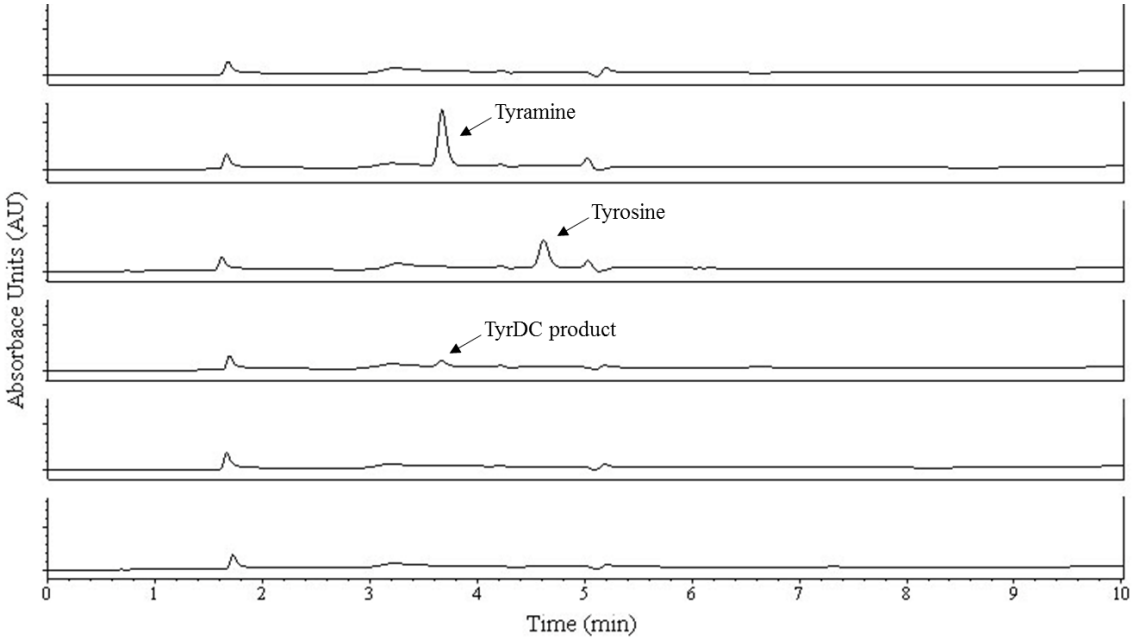

Figure A.1. Production of tyramine by tyrosine decarboxylase monitored at 288 nm on HPLC. Top to bottom; blank, tyramine, tyrosine, functioning assay with TyrDC, no substrate control, and no enzyme control.

A.3 References

Lehmann, T., and Pollmann, S. (2009). Gene expression and characterization of a stress-induced tyrosine decarboxylase from *Arabidopsis thaliana*. FEBS Lett. **583,** 1895-1900.

Appendix B: The Identification of a 2-Oxoglutarate Dependent Dioxygenase Capable of Hydroxylating Vittatine into One of the Two Diastereomeric Forms of 11-Hydroxvittatine.

B.1 Introduction, Results, and Discussion

The biosynthesis of the haemanthamine-type alkaloids involves the hydroxylation of vittatine to 11-hydroxyvittatien as discussed in detail in Chapter 1.1.5. During the research described in this thesis, a 2-oxoglutarate dioxygenase was found to co-express with *N4OMT* in all the *Galanthus* spp. assemblies described in Chapter 3. This cDNA was expressed in Sf9 cells and the recombinant enzyme tested for the ability to hydroxylate vittatine. It was found to hydroxylate vittatine with a fragmentation pattern and retention time identical to one of the two peaks produced by the diastereomeric mixture present in the reference compound (Figure B.1A-C), however, the absolute configuration of the product is unknown. The enzyme clearly only makes one diastereomer (Figure B.1A). The cDNA sequence of vittatine 11-hydroxylase has been deposited in GenBank under the accession number KT985905.

B.2 Methods

B.2.1 Reagents

Reagents were as described in Chapters 2, 3, and 4. Vittatine and 11-hydroxyvittatine were synthesized previously. Reagents unique to this study were sodium ascorbate and α-ketoglutaric acid potassium salt (2-oxoglutarate) from Sigma-Aldrich and ferrous sulfate from Fisher Scientific.

B.2.2 Candidate Gene Selection

As a part of the search for oxidoreductases in Chapter 4, *Galanthus*-20120814-25202 (vittatine 11-hydroxylase), a homologue of the *Pisum sativum* 2-oxoglutarate-dependent dioxygenase (O24648), was found to co-express with *N4OMT* in both the ABySS and MIRA and the Trinity based assemblies for *Galanthus* sp. and *Galanthus elwesii* (Lester et al., 1997). Its co-expression with four of the five available transcriptomes prompted further study.

B.2.3 PCR

Vittatine 11-hydroxylase had a complete open reading frame in the *Galanthus* sp. ABySS and MIRA assembly. Nested PCR was used to generate this ORF, followed by cloning and expression in insect cells as in Chapter 3 (see Table B.1 for primers). When sequenced, several assembly errors were identified with several substitutions and an insertion of nucleotides at both the 3'- and

5’ end. These changes corrected several frame shifts predicted by the assembly that would have resulted in a truncated protein.

B.2.4 Enzyme assays

Enzyme assays were incubated at 30 ºC for 4 hr and consisted of 100 µM Tris pH 7.5, 500 µM ferrous sulfate, 500 µM sodium ascorbate, 500 µM 2-oxoglutarate, 35 % Sf9 cell suspension, and 10 µM vittatine. Assays were extracted with EtOAc, resuspended, and run on the same LC-MS/MS hardware set-up as in Chapter 3. Specific parameters for vittatine monitoring were m/z 272.1, declustering potential (70), and collision energy (35). The product 11-hydroxyvittatine was monitored at m/z 288.1, declustering potential (70), and collision energy (35).

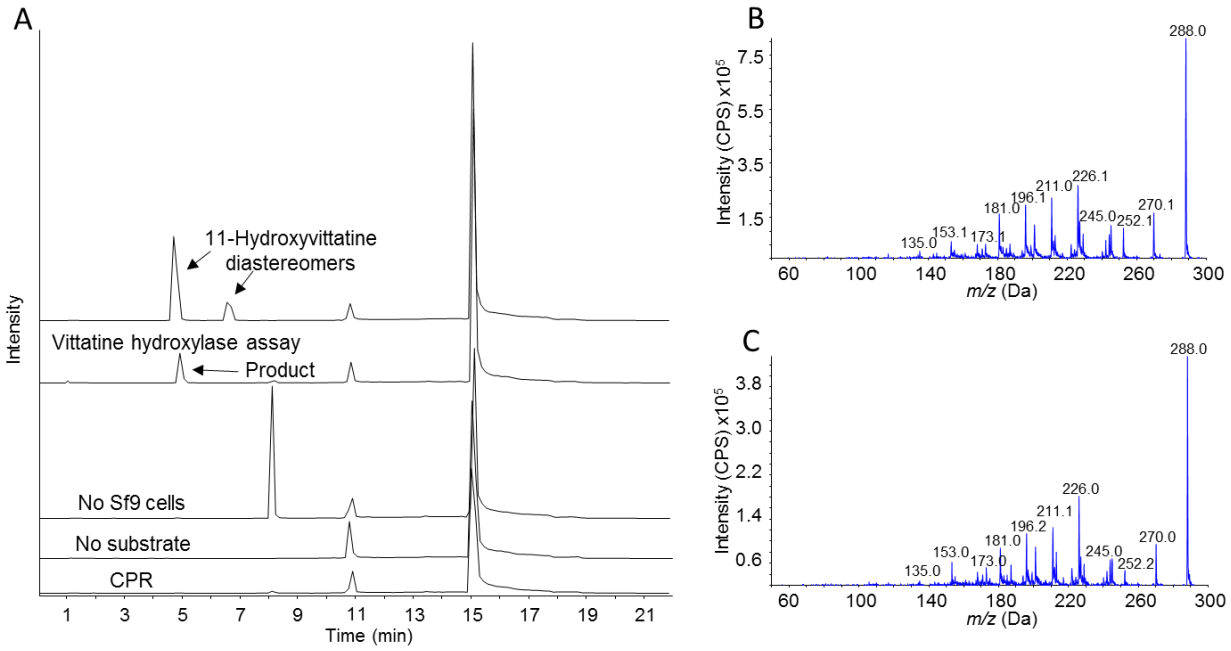

Figure B.1. Assays for the vittatine 11-hydroxylase. (A) Top to bottom 11-hydroxyvittatine, assay with vittatine 11-hydroxylase expressing Sf9 cells, assay with no Sf9 cells, assay with no vittatine substrate, and Sf9 cells expressing CPR instead of vittatine 11-hydroxylase. (B) Fragmentation pattern of 11-hydroxyvittatine. (C) Fragmentation pattern of vittatine 11-hydroxylase product.

Table B.1. Primers used

Primer name	Primer sequence
Ga25202_FW_o	TCAAGCAAAGACACCCATTCT
Ga25202_FW_i	aattGGATCCATGGGTTCCGATGTCAAATCA
Ga25202_RV_o	CTTGTTCCAAACACATTACAACA
Ga25202_RV_i	aattGCGGCCGCTTACACGATTCGTTTCTGCCA

B.3 References

Lester, D.R., Ross, J.J., Davies, P.J., and Reid, J.B. (1997). Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. Plant Cell **9,** 1435-1443.

Appendix C: Supplementary Material for Chapter 2

C.1 Figures and Tables

Table C.1. Methyltransferases used in BLAST search

Accession number	Substrate specificity	Reference
AAQ01669.1	(*R*,*S*)-Norcoclaurine, (*R*)-norprotosinomenine, (*S*)-norprotosinomenine, (*R*,*S*)-isoorientaline,	(Ounaroon et al., 2003)
AAQ01670.1		unpublished
AAQ01668.1	Guaiacol, isovanillic acid, (*R*)-reticuline, (*S*)-reticuline, (*R*,*S*)-orientaline, (*R*)-protosinomenine, (*R*,*S*)-laudanidine	(Ounaroon et al., 2003)
BAI79244.1	(1*S*)-*N*-Deacetylisoipecoside, (1*R*)-*N*-deacetylipecoside, (13a*R*)-demethylalangiside, (11bS)-7′-*O*-demethylcephaeline, (13a*S*)-redipecamine, (1*R*,*S*)-isococlaurine, (1*R*,*S*)-norcoclaurine, (1*R*,*S*)-isoorientaline, oripavine	(Nomura and Kutchan, 2010)
BAI79245.1	(13a*S*)-3-*O*-Methylredipecamine, (1*S*)-coclaurine, (1*R*,*S*)-*N*-methylcoclaurine, (1*R*,*S*)-4′-*O*-methylcoclaurine, (1*R*,*S*)-6-O-methyllaudanosoline, (1*R*,*S*)-nororientaline, (1*S*)-norreticuline, (1*S*)-reticuline, (13a*S*)-coreximine	(Nomura and Kutchan, 2010)
BAI79243.1	(1*S*)-*N*-Deacetylisoipecoside, (1*S*)-7-*O*-methyl-*N*-deacetylisoipecoside, (11b*S*)-cephaeline, (1*R*,*S*)-isococlaurine, (1*R*,*S*)-norcoclaurine, (1*S*)-4′-*O*-methyllaudanosoline, (1*R*,*S*)-nororientaline, (1*R*,*S*)-isoorientaline, (1*S*)-norprotosinomenine, (1*R*)-norprotosinomenine, (1*R*,*S*)-protosinomenine	(Nomura and Kutchan, 2010)
BAA06192.1	(*R*,*S*)-Scoulerine	(Takeshita et al., 1995)
AAD29843.1	See reference	(Frick and Kutchan, 1999)
AAD29841.1	See reference	(Frick and Kutchan, 1999)
AAD29845.1	See reference	(Frick and Kutchan, 1999)
AAD29842.1	See reference	(Frick and Kutchan, 1999)
AAD29844.1	See reference	(Frick and Kutchan, 1999)
BAC22084.1	Columbamine, tetrahydrocolumbamine, (*S*)-scoulerine, 2,3,9,10-tetrahydroxyprotoberberine	
ACV50428.1	Homology with caffeoyl-CoA *O*-methyltransferase described in (Day et al., 2009)	(Eswaran et al., 2010)
AAN61072.1	Quercetin, 7-*O*-methylquercetin, quercetin-3-*O*-glucoside, quercetagetin, 3-*O*-methylquercetagetin 6-*O*-methylquercetagetin, 6-hydroxykaempferol, myricetin, luteolin, caffeoyl-CoA	(Ibdah et al., 2003)
AAR02420.1	Eriodictyol, homoeriodictyol, kaempferol, quercetin, isorhamnetin, chrysoeriol	(Schröder et al., 2004)

Table C.2. Methyltransferases used in phylogeny

Accession number	Short name	Species	Substrate specificity	Reference
AAQ01669.1	*Ps*N6OMT	*Papaver somniferum*	(*R*,*S*)-Norcoclaurine, (*R*)-norprotosinomenine, (*S*)-norprotosinomenine, (*R*,*S*)-isoorientaline,	(Ounaroon et al., 2003)
AAQ01670.1	*Ps*COMT	*Papaver somniferum*		Unpublished
AAQ01668.1	*Ps*R7OMT	*Papaver somniferum*	Guaiacol, isovanillic acid, (*R*)-reticuline, (*S*)-reticuline, (*R*,*S*)-orientaline, (*R*)-protosinomenine, (*R*,*S*)-laudanidine	(Ounaroon et al., 2003)
BAI79244.1	*Pi*OMT2	*Psychotria ipecacuanha*	(1*S*)-*N*-Deacetylisoipecoside, (1*R*)-*N*-deacetylipecoside, (13a*R*)-demethylalangiside, (11b*S*)-7′-*O*-demethylcephaeline, (13a*S*)-redipecamine, (1*R*,*S*)-isococlaurine, (1*R*,*S*)-norcoclaurine, (1*R*,*S*)-isoorientaline, oripavine	(Nomura and Kutchan, 2010)
BAI79243.1	*Pi*OMT1	*Psychotria ipecacuanha*	(1*S*)-*N*-Deacetylisoipecoside, (1*S*)-7-*O*-methyl-*N*-deacetylisoipecoside, (11b*S*)-cephaeline, (1*R*,*S*)-Isococlaurine, (1*R*,*S*)-norcoclaurine, (1*S*) 4′*O*-methyllaudanosoline, (1*R*,*S*)-nororientaline, (1*R*,*S*)-isoorientaline, (1*S*)-norprotosinomenine, (1*R*)-norprotosinomenine, (1*R*,*S*)-protosinomenine	(Nomura and Kutchan, 2010)
BAA06192.1	*Cj*S9OMT	*Coptis japonica*	(*R*,*S*)-Scoulerine	(Takeshita et al., 1995)
AAD29843.1	*Tt*COMT3	*Thalictrum tuberosum*	See reference	(Frick and Kutchan, 1999)
AAD29841.1	*Tt*COMT1	*Thalictrum tuberosum*	See reference	(Frick and Kutchan, 1999)
AAD29845.1	*Tt*COMT5	*Thalictrum tuberosum*	See reference	(Frick and Kutchan, 1999)
AAD29842.1	*Tt*COMT2	*Thalictrum tuberosum*	See reference	(Frick and Kutchan, 1999)
AAD29844.1	*Tt*COMT4	*Thalictrum tuberosum*	See reference	(Frick and Kutchan, 1999)
BAC22084.1	*Cj*COMT	*Coptis japonica*	Columbamine, tetrahydrocolumbamine, (*S*)-scoulerine, 2,3,9,10-tetrahydroxyprotoberberine	(Morishige et al., 2002)
ACV50428.1	*Jc*CCoAOMT	*Jatropha curcas*	Homology with caffeoyl-CoA *O*-methyltransferase described in (Day et al., 2009)	(Eswaran et al., 2010)
AAR02420.1	*Cr*F4OMT	*Catharanthus roseus*	Eriodictyol, homoeriodictyol, kaempferol, quercetin, isorhamnetin, chrysoeriol	(Schröder et al., 2004)
Q9C5D7.1	*At*CCoAOMT	*Arabidopsis thaliana*	Not determined	(Ibrahim et al., 1998)
C7AE94.1	*Vv*AOMT	*Vitis vinifera*	Cyanidin 3-glucoside, delphinidin 3-glucoside, quercetin 3-glucoside, cyanidin, quercetin, myricetin, pelargonidin 3-glucoside, catechin, epicatechin	(Hugueney et al., 2009)
ADZ76153.1	*Vp*OMT4	*Vanilla planifolia*	Tricetin, 5-hydroxyferulic acid ethyl ester, 5-hydroxyferulic acid, myricetin, 3,4-	(Widiez et al., 2011)

			dihydroxybenzaldehyde, quercetin, 5-hydroxyconiferaldehyde, caffeoyl CoA, caffeic acid ethyl ester, caffeoylaldehyde, caffeic acid	
ADZ76154.1	*Vp*OMT5	*Vanilla planifolia*	Tricetin, 5-hydroxyferulic acid ethyl ester, 5-hydroxyferulic acid, myricetin, 3,4-dihydroxybenzaldehyde, quercetin, 5-hydroxyconiferaldehyde, caffeoyl CoA, caffeic acid ethyl ester, caffeoylaldehyde, caffeic acid	(Widiez et al., 2011)
Q84KK6	*Ge*I4OMT	*Glycyrrhiza echinata*	2,7,4'-Trihydroxyisoflavanone, medicarpin	(Akashi et al., 2003)
C6TAY1	*Gm*F4OMT	*Glycine max*	Apigenin, daidzein, genistein, quercetin, naringenin	(Kim et al., 2005)
AAY89237.1	*Lu*CCoA3OMT	*Linum usitatissimum*		(Apuya et al., 2008)
3C3Y|A	*Mc*PFOMT	*Mesembryanthemum crystallinum*	Quercetin, quercetagetin, caffeic acid, CoA, caffeoyl glucose	(Kopycki et al., 2008a)
62361_DF6	*Np*N4OMT1	*Narcissus* sp. *aff. pseudonarcissus*	Norbelladine, *N*-methylnorbelladine, dopamine	this study
BAB71802.1	*Cj*CNMT	*Coptis japonica*	(*R*)-Coclaurine, (*S*)-coclaurine, (*R*,*S*)-norreticuline, (*R*,*S*)-norlaudanosoline, (*R*,*S*)-6-*O*-methylnorlaudanosoline, 6,7-dimethoxyl-1,2,3,4-tetrahydroisoquinoline, 1-methyl-6,7-dihydroxy-1,2,3,4-tetrahydroisoquinolinne	(Choi et al., 2002)
BAB12278.1	*Cs*CNMT	*Camellia sinensis*	7-Methylxanthine, 3-methylxanthine, 1-methylxanthine, theobromine, theophylline, paraxanthine	(Kato et al., 2000)
Q93WU3	*Ob*CV4OMT	*Ocimum basilicum*	Chavicol, phenol, eugenol, t-isoeugenol, t-anol	(Gang et al., 2002)
Q8WZ04	*Hs*COMT	*Homo sapiens*	A catechol	
3CBG|A	*SynOMT*	*Cyanobacterium Synechocystis* sp. strain PCC 6803	Hydroxyferulic acid, caffeic acid, caffeoyl-CoA, caffeoylglucose, 3,4,5-trihydorxycinnamic acid, tricetin, 3,4-dihydroxybenzoic acid	(Kopycki et al., 2008b)

Table C.3. Primers used in RACE, cloning, and colony PCR

Primer name	Primer sequence
62361_5'_RACE_outer	TCCACCTCATCTTCCGGACGAA
62361_5'_RACE_inner	ACTTCCGTTCCAGAGCGTGTT
62361_3'_RACE_outer	AGAAGACCTGTACGACCATGCAT
62361_3'_RACE_inner	ACGAGCGATTAGTGAAGCTCGTCA
62361_forward_outer	CTTCACTTGTGTCAAGTTCAAT
62361_reverse_outer	CCRATAGATAGCATGCAGAATCT
62361_forward_inner	[a,b]aattCATATGGGTGCTAGCATAGATGATT
62361_reverse_inner	[a,b]aattGCGGCCGCTCAATAAAGACGTCGGCAAATAGT
qRT-PCR_forward_62361	ATTGGTGTGTACACCGGCTATT
qRT-PCR_reverse_62361	TTCCATCTTCCGGTAAAGCCAAA
qRT-PCR_probe_62361	CTCTGCTCACAACTGC
T7 sequencing	TAATACGACTCACTATA
T7 terminator	GCTAGTTATTGCTCAGCGG

Table C.4. Parameters used for LC/MS/MS analysis

Compound	Predicted molecular ion *m/z* [M+H]	CE value (V)	DP value (V)	Injection volume (µl)
Galanthamine	288.1	35	70	10
Norbelladine*	260.1	15	60	10
4'-*O*-Methylnorbelladine*	274.1	35	60	10
N-Methylnorbelladine	274.1	20	60	10
4'-*O*-Methyl-*N*-methylnorbelladine*	288.2	20	60	10
Dopamine*	154.1	20	70	20
3'-*O*-Methyldopamine	168.1	20	70	20
Methylated dopamine product	168.1	20	70	20
Papaverine	340.2	52	70	10

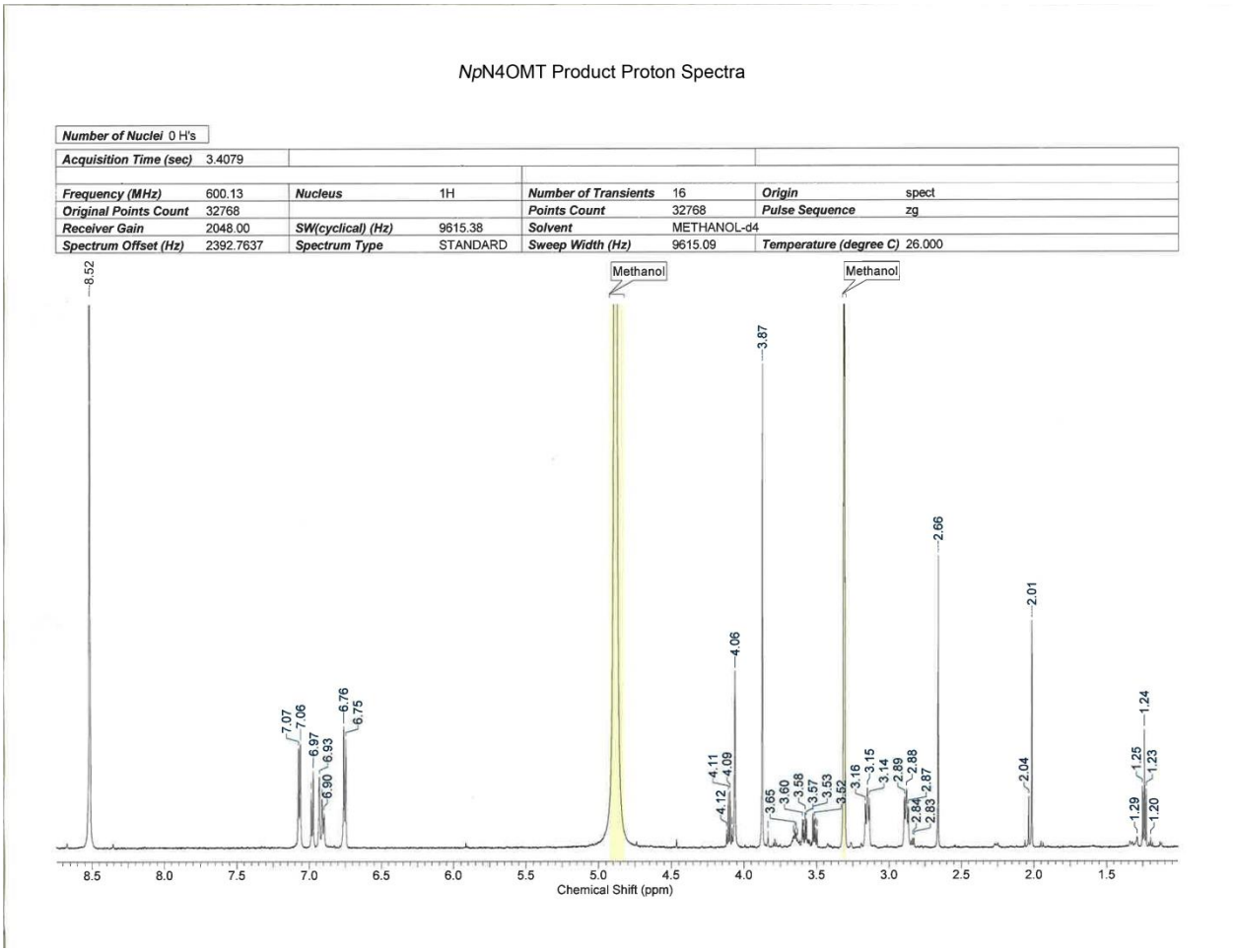

Figure C.1. *Np*N4OMT1 product 4'-*O*-methylnorbelladine proton NMR spectra with peak assignments.

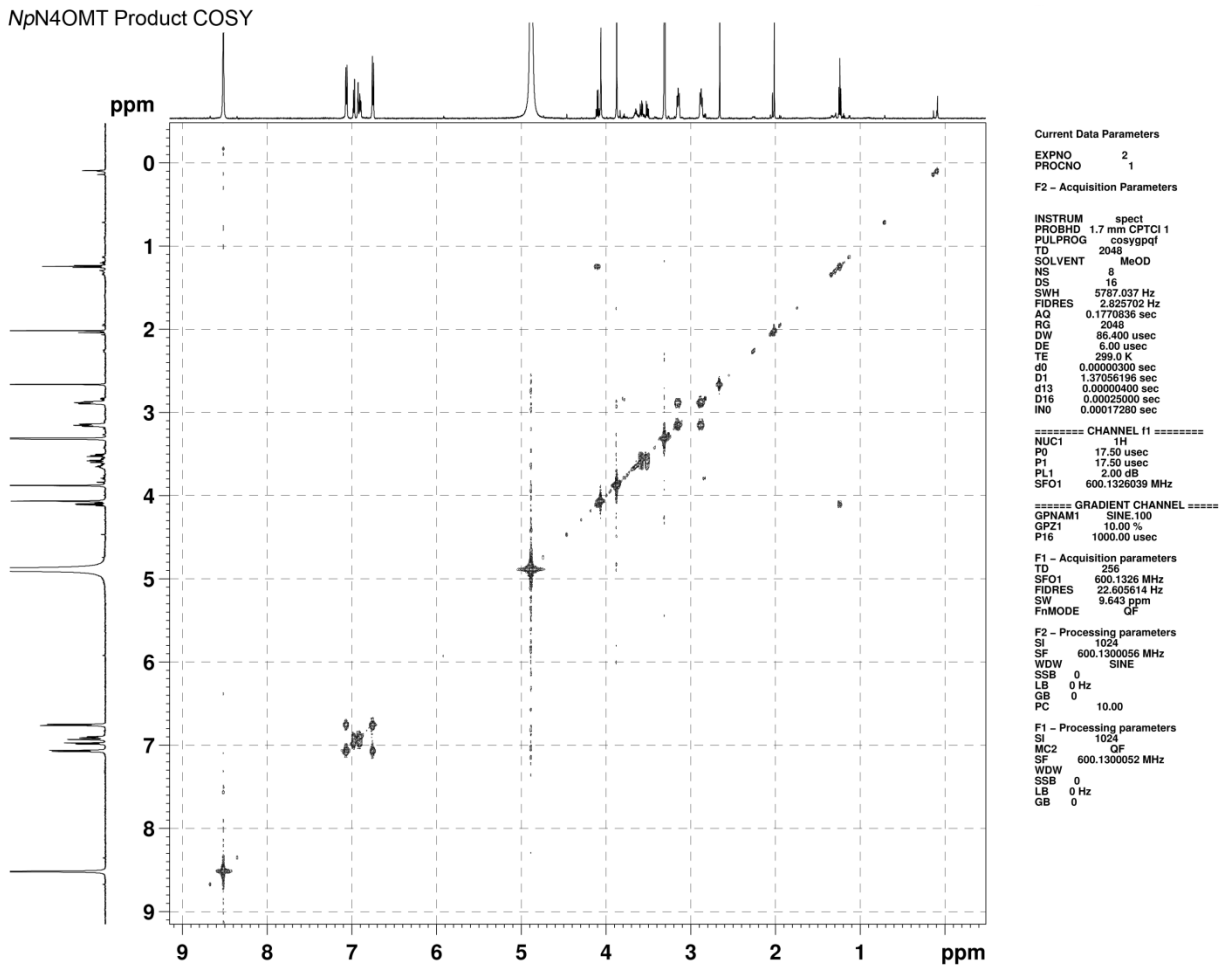

Figure C.2. *Np*N4OMT1 product 4'-*O*-methylnorbelladine COSY spectra.

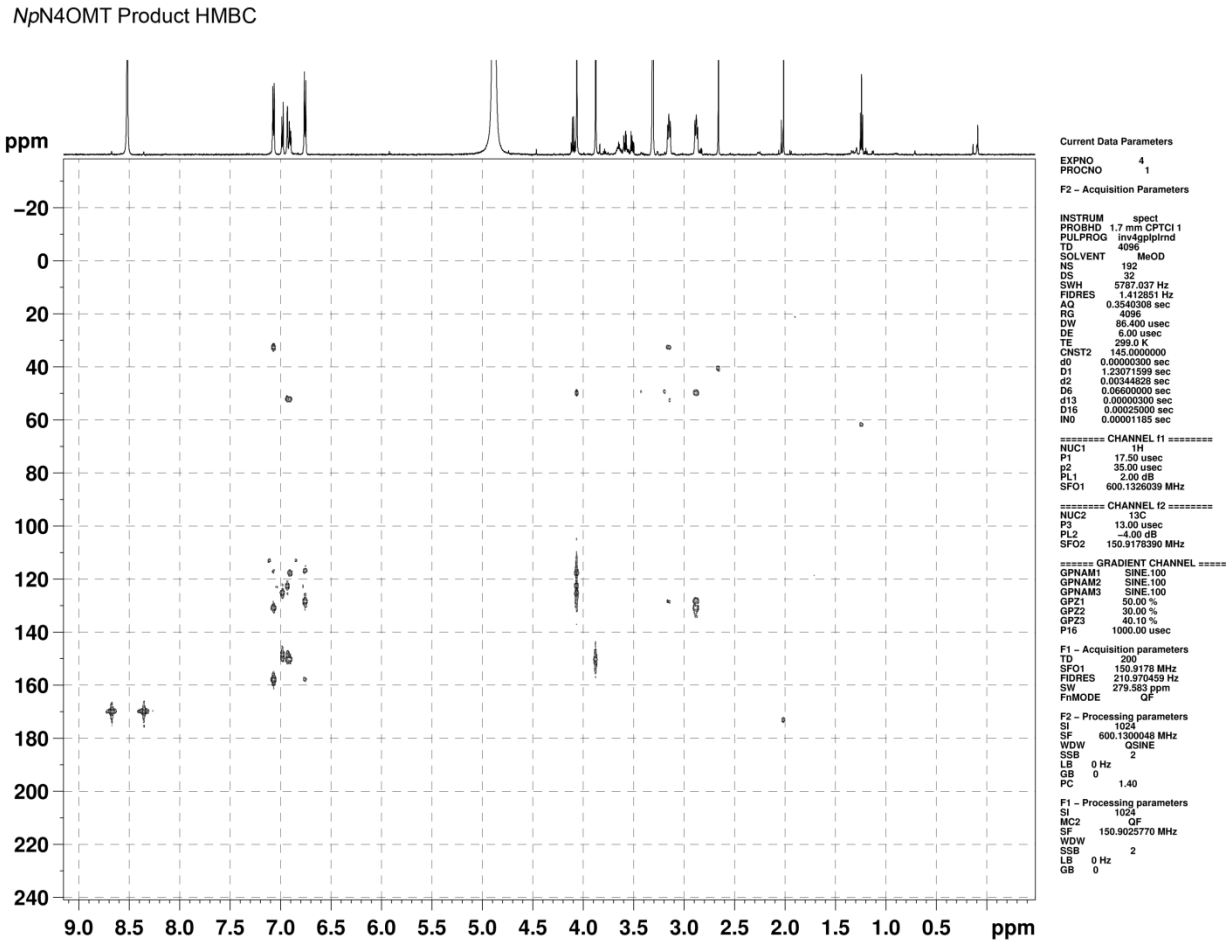

Figure C.3. *Np*N4OMT1 product 4'-*O*-methylnorbelladine HMBC spectra.

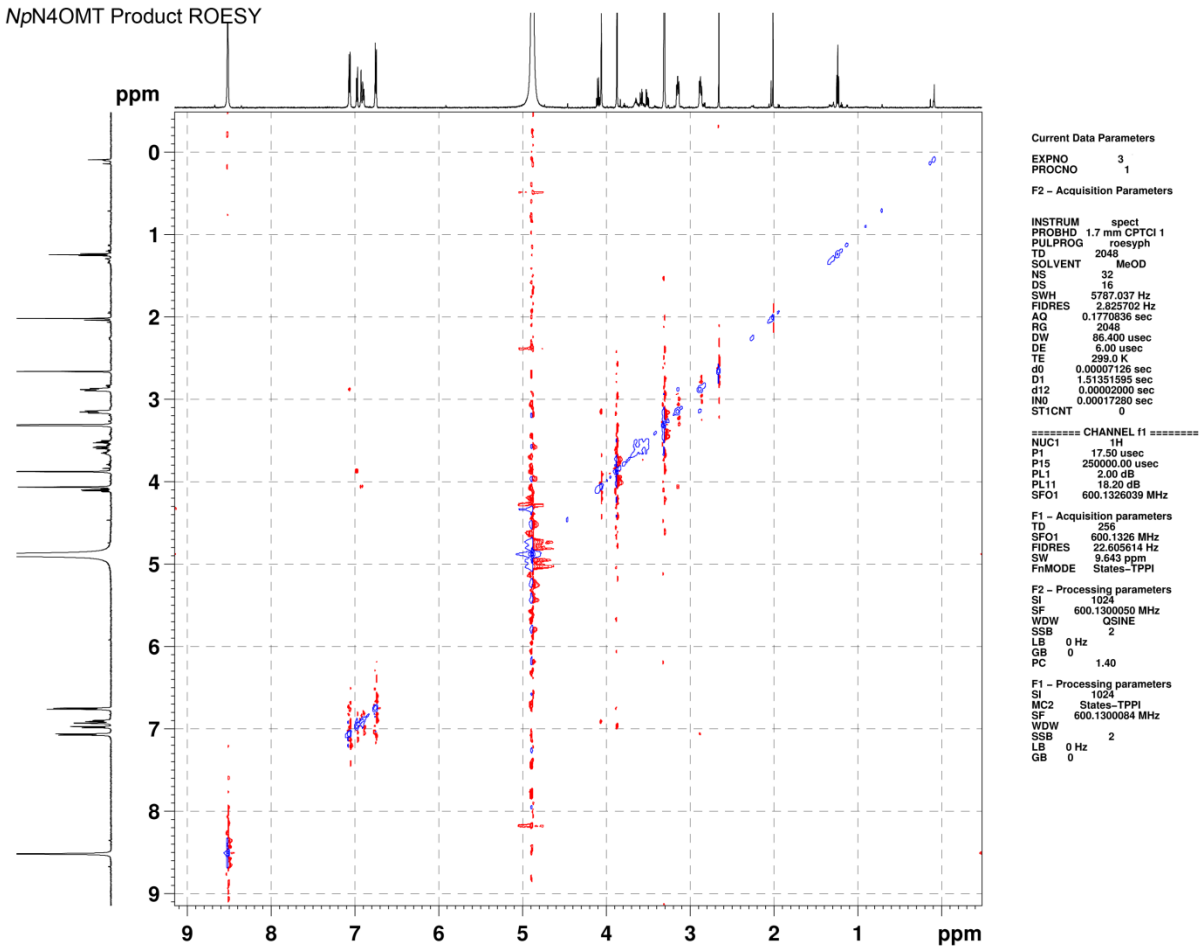

Figure C.4. *Np*N4OMT1 product 4'-*O*-methylnorbelladine ROESY spectra.

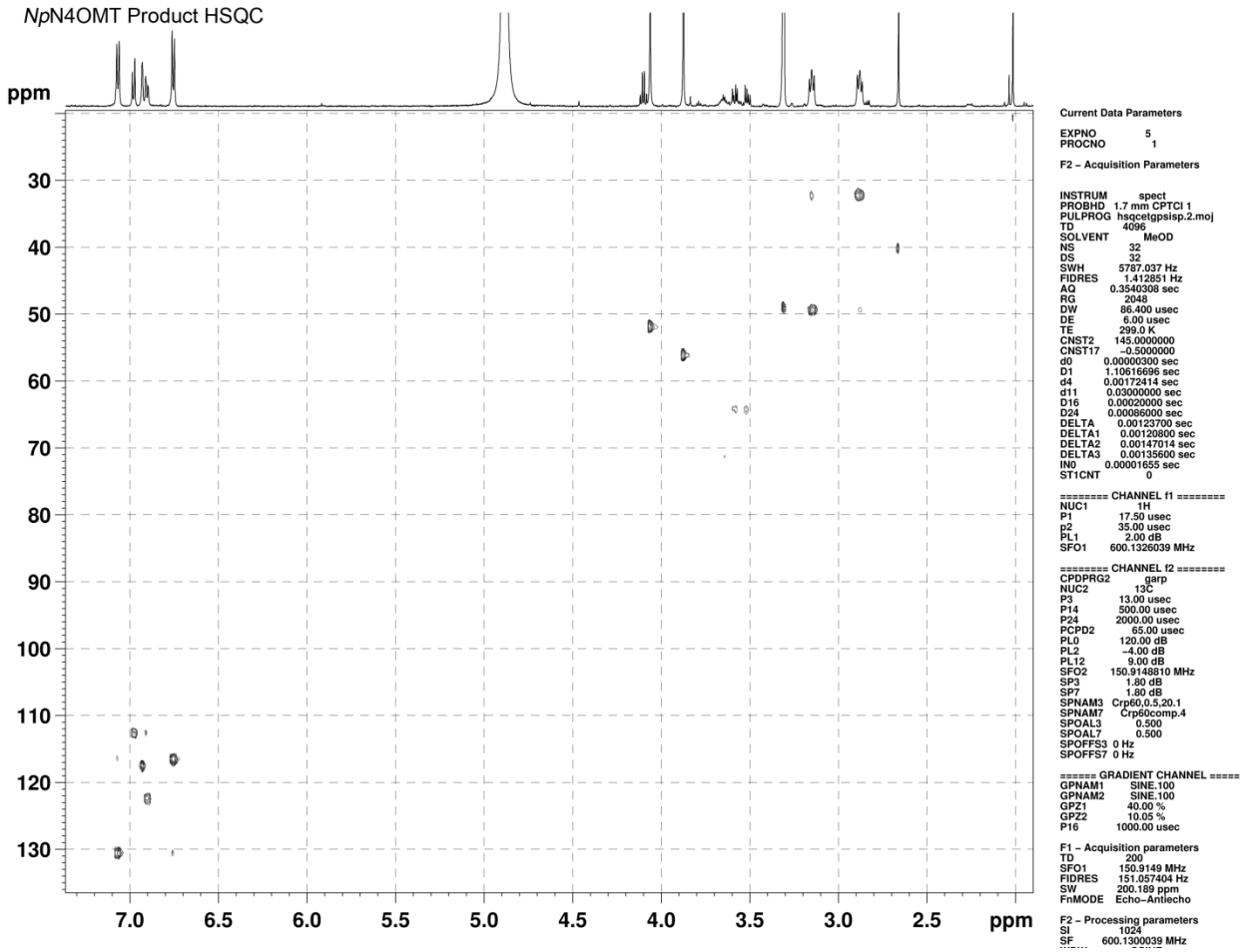

Figure C.5. *Np*N4OMT1 product 4'-*O*-methylnorbelladine HSQC spectra.

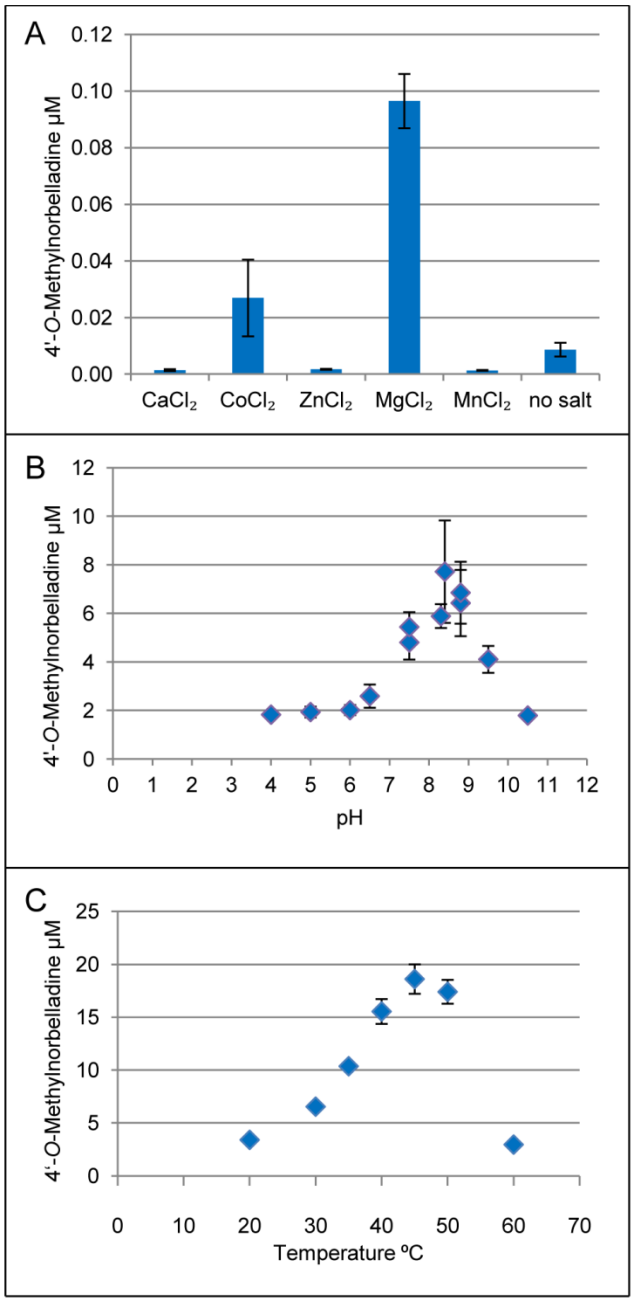

Figure C.6. Effect of divalent cations, temperature, and pH on *Np*N4OMT1 enzyme activity. (A) Divalent cations tested with 5 min assays with 5 µM of cation Ca^{2+}, Co^{2+}, Zn^{2+}, Mg^{2+} or Mn^{2+}. (B) pH optimum 15 min assays with 5 µM Mg^{2+}. (C) Temperature optimum 15 min assays with 5 µM Mg^{2+}. Divalent cation and pH testing reactions are 100 µl reactions at 37^0C. The divalent cation test contained 4 µM norbelladine, while pH and temperature optimum tests contained 100 µM norbelladine in the assay mix.

C.2 References

Akashi, T., Sawada, Y., Shimada, N., Sakurai, N., Aoki, T., and Ayabe, S. (2003). cDNA cloning and biochemical characterization of *S*-adenosyl-L-methionine: 2,7,4'-trihydroxyisoflavanone 4'-*O*-methyltransferase, a critical enzyme of the legume isoflavonoid phytoalexin pathway. Plant Cell Physiol. **44,** 103-112.

Apuya, N., Bobzin, S., Okamuro, J., and Zhang, K. (2008). Modulating lignin in plants, I.N.C. Ceres, ed (WO).

Choi, K.B., Morishige, T., Shitan, N., Yazaki, K., and Sato, F. (2002). Molecular cloning and characterization of coclaurine *N*-methyltransferase from cultured cells of *Coptis japonica*. J. Biol. Chem. **277,** 830-835.

Day, A., Neutelings, G., Nolin, F., Grec, S., Habrant, A., Crônier, D., Maher, B., Rolando, C., David, H., Chabbert, B., and Hawkins, S. (2009). Caffeoyl coenzyme A *O*-methyltransferase down-regulation is associated with modifications in lignin and cell-wall architecture in flax secondary xylem. Plant Physiol. Biochem. **47,** 9-19.

Eswaran, N., Parameswaran, S., Sathram, B., Anantharaman, B., Kumar G, R.K., and Tangirala, S.J. (2010). Yeast functional screen to identify genetic determinants capable of conferring abiotic stress tolerance in *Jatropha curcas*. BMC Biotechnol. **10,** 23.

Frick, S., and Kutchan, T.M. (1999). Molecular cloning and functional expression of *O*-methyltransferases common to isoquinoline alkaloid and phenylpropanoid biosynthesis. Plant J. **17,** 329-339.

Gang, D.R., Lavid, N., Zubieta, C., Chen, F., Beuerle, T., Lewinsohn, E., Noel, J.P., and Pichersky, E. (2002). Characterization of phenylpropene *O*-methyltransferases from sweet basil: facile change of substrate specificity and convergent evolution within a plant *O*-methyltransferase family. Plant Cell **14,** 505-519.

Hugueney, P., Provenzano, S., Verriès, C., Ferrandino, A., Meudec, E., Batelli, G., Merdinoglu, D., Cheynier, V., Schubert, A., and Ageorges, A. (2009). A novel cation-dependent *O*-methyltransferase involved in anthocyanin methylation in grapevine. Plant Physiol. **150,** 2057-2070.

Ibdah, M., Zhang, X.H., Schmidt, J., and Vogt, T. (2003). A novel Mg(2+)-dependent *O*-methyltransferase in the phenylpropanoid metabolism of *Mesembryanthemum crystallinum*. J. Biol. Chem. **278,** 43961-43972.

Ibrahim, R.K., Bruneau, A., and Bantignies, B. (1998). Plant *O*-methyltransferases: molecular analysis, common signature and classification. Plant. Mol. Biol. **36,** 1-10.

Kato, M., Mizuno, K., Crozier, A., Fujimura, T., and Ashihara, H. (2000). Caffeine synthase gene from tea leaves. Nature **406,** 956-957.

Kim, D.H., Kim, B.G., Lee, Y., Ryu, J.Y., Lim, Y., Hur, H.G., and Ahn, J.H. (2005). Regiospecific methylation of naringenin to ponciretin by soybean *O*-methyltransferase expressed in *Escherichia coli*. J. Biotechnol. **119,** 155-162.

Kopycki, J.G., Rauh, D., Chumanevich, A.A., Neumann, P., Vogt, T., and Stubbs, M.T. (2008a). Biochemical and structural analysis of substrate promiscuity in plant Mg^{2+}-dependent *O*-methyltransferases. J. Mol. Biol. **378,** 154-164.

Kopycki, J.G., Stubbs, M.T., Brandt, W., Hagemann, M., Porzel, A., Schmidt, J., Schliemann, W., Zenk, M.H., and Vogt, T. (2008b). Functional and structural characterization of a cation-dependent *O*-methyltransferase from the cyanobacterium *Synechocystis* sp. strain PCC 6803. J. Biol. Chem. **283,** 20888-20896.

Morishige, T., Dubouzet, E., Choi, K.B., Yazaki, K., and Sato, F. (2002). Molecular cloning of columbamine *O*-methyltransferase from cultured *Coptis japonica* cells. Eur. J. Biochem. **269,** 5659-5667.

Nomura, T., and Kutchan, T.M. (2010). Three new *O*-methyltransferases are sufficient for all *O*-methylation reactions of ipecac alkaloid biosynthesis in root culture of *Psychotria ipecacuanha*. J. Biol. Chem. **285,** 7722-7738.

Ounaroon, A., Decker, G., Schmidt, J., Lottspeich, F., and Kutchan, T.M. (2003). (*R,S*)-Reticuline 7-*O*-methyltransferase and (*R,S*)-norcoclaurine 6-*O*-methyltransferase of *Papaver somniferum* - cDNA cloning and characterization of methyl transfer enzymes of alkaloid biosynthesis in opium poppy. Plant J. **36,** 808-819.

Schröder, G., Wehinger, E., Lukacin, R., Wellmann, F., Seefelder, W., Schwab, W., and Schröder, J. (2004). Flavonoid methylation: a novel 4'-*O*-methyltransferase from *Catharanthus roseus*, and evidence that partially methylated flavanones are substrates of four different flavonoid dioxygenases. Phytochemistry **65,** 1085-1094.

Takeshita, N., Fujiwara, H., Mimura, H., Fitchen, J.H., Yamada, Y., and Sato, F. (1995). Molecular cloning and characterization of *S*-adenosyl-L-methionine: scoulerine-9-*O*-methyltransferase from cultured cells of *Coptis japonica*. Plant Cell Physiol. **36,** 29-36.

Widiez, T., Hartman, T.G., Dudai, N., Yan, Q., Lawton, M., Havkin-Frenkel, D., and Belanger, F.C. (2011). Functional characterization of two new members of the caffeoyl CoA *O*-methyltransferase-like gene family from *Vanilla planifolia* reveals a new class of plastid-localized *O*-methyltransferases. Plant Molecular Biology **76,** 475-488.

Appendix D: Supplementary Material for Chapter 4

D.1 Figures and Tables

Table D. 1. Oxidoreductases used in BLASTP search

Accession number	Short name	Species	Reference
Q9SE94		Maize	(Roje et al., 1999)
AAW60421.1		Gluconobacter oxydans 621H	(Chen et al., 2010; Schweiger et al., 2010)
Q9SQ70	NADPH-dependent codeinone reductase	Papaver somniferum	(Unterlinner et al., 1999)
ABO93462.1	SalR	P. somniferum, Papaver bracteatum	(Ziegler et al., 2006; Geissler et al., 2007)
P51102	Dihydroflavonol-4-reductase	*Arabidopsis thaliana*	(Saito et al., 2013)
Q40316	Vestitone reductase	*Medicago sativa*	(Guo and Paiva, 1995)
3H7U	Akr4c9	*Arabidopsis thaliana*	(Simpson et al., 2009)
AAD37373.1	Quinone oxidoreductase	*Arabidopsis thaliana*	(Sparla et al., 1999)
B5WWZ8	FAO1	*Lotus japonicus*	(Zhao et al., 2008)
O23240	D-2-Hydroxyglutarate dehydrogenase	*Arabidopsis thaliana*	(Engqvist et al., 2009)
O24562	Probable cinnamyl alcohol dehydrogenase	*Zea mays*	(Halpin et al., 1998)
O24648	2-Oxoglutarate-dependent dioxygenase	*Pisum sativum*	(Lester et al., 1997)
O48741	Protochlorophyllide reductase C, chloroplastic	*Arabidopsis thaliana*	(Oosawa et al., 2000)
O49213	GDP-L-fucose synthase 1	*Arabidopsis thaliana*	(Bonin and Reiter, 2000)
O49482	Cinnamyl alcohol dehydrogenase 5	*Arabidopsis thaliana*	(Sibout et al., 2005)
O80944	Aldo-keto reductase family 4 member C8	*Arabidopsis thaliana*	(Simpson et al., 2009)
O81852	Bifunctional aspartokinase/homoserine dehydrogenase 2	*Arabidopsis thaliana*	(Paris et al., 2002)
P06525	Alcohol dehydrogenase class-P	*Arabidopsis thaliana*	(Chang and Meyerowitz, 1986)
P16972	Ferredoxin-2, chloroplastic	*Arabidopsis thaliana*	(Somers et al., 1990)
Q05431	Ascorbate peroxidase 1, cytosolic	*Arabidopsis thaliana*	(Davletova et al., 2005)
Q0PGJ6	Aldo-keto reductase family 4 member C9	*Arabidopsis thaliana*	(Simpson et al., 2009)
Q39219	Ubiquinol oxidase 1a, mitochondrial	*Arabidopsis thaliana*	(Berthold, 1998)

Q39659	Glyoxysomal fatty acid beta-oxidation multifunctional protein MFP-a	*Cucumis sativus*	(Preisig-Müller et al., 1994)
Q43727	Glucose-6-phosphate 1-dehydrogenase 1, chloroplastic	*Arabidopsis thaliana*	
Q56Y42	Pyridoxal reductase, chloroplastic	*Arabidopsis thaliana*	(Herrero et al., 2011)
Q56YU0	Aldehyde dehydrogenase family 2 member C4	*Arabidopsis thaliana*	
Q5C9I9	(-)-Isopiperitenol/(-)-carveol dehydrogenase, mitochondrial	*Mentha piperita*	(Ringer et al., 2005)
Q5N800	Probable chlorophyll(ide) b reductase NYC1 chloroplastic	*Oryza sativa* subsp. Japonica	(Sato et al., 2009)
Q6ZHS4	Cinnamyl alcohol dehydrogenase 2	*Oryza sativa* subsp. japonica	
Q8GWA1	Internal alternative NAD(P)H-ubiquinone oxidoreductase A1, mitochondrial	*Arabidopsis thaliana*	(Michalecka et al., 2003)
Q8GXR9	Alternative NAD(P)H-ubiquinone oxidoreductase C1, chloroplastic/mitochondrial	*Arabidopsis thaliana*	
Q8L799	Inositol oxygenase 1	*Arabidopsis thaliana*	
Q8L9C4	Very-long-chain 3-oxoacyl-CoA reductase 1	*Arabidopsis thaliana*	(Beaudoin et al., 2009)
Q8LEU3	Chlorophyll(ide) b reductase NOL, chloroplastic	*Arabidopsis thaliana*	(Horie et al., 2009)
Q949Q0	Glycerol-3-phosphate dehydrogenase [NAD(+)] 2, chloroplastic	*Arabidopsis thaliana*	(Nandi et al., 2004)
Q94AX4	lactate dehydrogenase [cytochrome], mitochondrial	*Arabidopsis thaliana*	
Q9C9W5	Glycerate dehydrogenase HPR, peroxisomal	*Arabidopsis thaliana*	
Q9CA90	Glyoxylate/hydroxypyruvate reductase A HPR2	*Arabidopsis thaliana*	
Q9FJ95	Sorbitol dehydrogenase	*Arabidopsis thaliana*	
Q9LKA3	Malate dehydrogenase 2, mitochondrial	*Arabidopsis thaliana*	
Q9LW56	Long-chain-alcohol oxidase	*Arabidopsis thaliana*	
Q9SA18	Bifunctional aspartokinase/homoserine dehydrogenase 1, chloroplastic	*Arabidopsis thaliana*	
Q9SJ10	Cinnamyl alcohol dehydrogenase 3	*Arabidopsis thaliana*	
Q9SQT8	Bifunctional 3-dehydroquinate dehydratase/shikimate dehydrogenase, chloroplastic	*Arabidopsis thaliana*	
Q9ST62	External alternative NADPH-ubiquinone oxidoreductase B1, mitochondrial	*Solanum tuberosum*	
Q9ST63	Internal alternative NADPH-ubiquinone oxidoreductase A1, mitochondrial	*Solanum tuberosum*	

Q9SU56	Galactono-1,4-lactone dehydrogenase, mitochondrial	*Arabidopsis thaliana*	
Q9SZB3	Putative uncharacterized protein	*Arabidopsis thaliana*	
Q9ZP06	Malate dehydrogenase 1, mitochondrial	*Arabidopsis thaliana*	
Q9ZPI5	Peroxisomal fatty acid beta-oxidation multifunctional protein MFP2	*Arabidopsis thaliana*	

*the end list contains all sequences in this table plus all plant AKRs listed on the http://www.med.upenn.edu/akr/members.shtml web page and all sequences minus GRMZM2G086773 from (Moummou et al., 2012) Table #3 were included.

Table D.2. Imine oxidoreductases used in BLASTP search

Accession number	Short name	Species	Reference
AED92078.1	Pyrroline-5-carboxylate reductase	*Arabidopsis thaliana*	(Funck et al., 2012)
Q05762	Bifunctional dihydrofolate reductase-thymidylate synthase 1	*Arabidopsis thaliana*	(Lazar et al., 1993)
F4JVY8	Lysine-ketoglutarate reductase/saccharopine dehydrogenase bifunctional enzyme	*Arabidopsis thaliana*	(Stepansky et al., 2005)
O22213	Cytokinin dehydrogenase 1	*Arabidopsis thaliana*	(Bilyeu et al., 2001)
Q4ADV8	Cytokinin dehydrogenase 2	*Oryza sativa* subsp. japonica	
Q6NKX1	Proline dehydrogenase 2, mitochondrial	*Arabidopsis thaliana*	(Funck et al., 2010)
Q8H191	Probable polyamine oxidase 4	*Arabidopsis thaliana*	(Tavladoraki et al., 2006)
Q9FNA2	Polyamine oxidase 1	*Arabidopsis thaliana*	
Q9FUJ3	Cytokinin dehydrogenase 2	*Arabidopsis thaliana*	
Q9LYT1	Polyamine oxidase 3	*Arabidopsis thaliana*	(Tavladoraki et al., 2006)
Q9SJA7	Probable sarcosine oxidase	*Arabidopsis thaliana*	
Q9SKX5	Probable polyamine oxidase 2	*Arabidopsis thaliana*	

Table D.3. Primers used

Primer name	Primer sequence
NNR forward outer	GGAAAGCCTTCAGAGGAGATT
NNR forward inner	aattGGATCCATGTCGTTGGAGAAGAGATGGT
NNR reverse outer	AGATAGCACCGTGGAGAT
NNR reverse inner	aattGCGGCCGCTCAACCATTTATGGTCCGTCCT

Table D.4. Product MS/MS parameters

Substrate	Product	Parameters (parent ion *m/z*)(CE)(DP)
3,4-Dihydroxybenzaldehyde and tyramine	Norbelladine	(260.0)(15)(50)
Noroxomaritidine	Oxomaritinamine	(274.3)(35)(70)
Vanillin and tyramine	3'-*O*-Methylnorbelladine	(274.3)(20)(60)
Isovanillin and tyramine	4'-*O*-Methylnorbelladine	(274.3)(20)(60)
Piperonal and tyramine	4-(2-((1,3-Benzodioxol-5-ylmethyl)amino)ethyl)phenol	(272.1)(20)(60)
8-*O*-,9-*O*-Didemethyloxomaritidine	9-*O*-Demethyloxomaritinamine	(260.1)(35)(70)

Collision energy=CE
Declustering potential=DP

Table D.5. NNR crystal structure parameters

Crystal	NNR•$NADP^+$	NNR•$NADP^+$•Piperonal	NNR•$NADP^+$•Tyramine
Space group	P 42 2 2	P 31 2 1	P 21 21 21
Cell dimensions	*a* = *b* = 60.377, *c* = 136.142Å	*a* = *b* = 73.386, *c* = 167.892Å	*a* = 61.8, *b* = 86.964, *c* = 186.115 Å
Data collection			
Wavelength (Å)	0.98	0.98	0.98
Resolution range (Å) (highest shell resolution)	45.17 - 1.73 (1.792 - 1.73)	35.85-1.501 (1.554-1.501)	39.39 - 1.814 (1.879 - 1.814)
Reflections (total/unique)	27,098/2,606	84,467/8,243	90,854/8,266
Completeness (highest shell)	99.73% (98.60%)	99.86% (98.67%)	99.17% (91.62%)
〈*I*/σ〉 (highest shell)	21.96 (2.39)	17.55 (2.30)	23.73 (2.30)
Model and refinement			
*R*cryst/*R*free	18.1/21.5%	14.9/16.0%	18.3/21.3%
No. of protein atoms	1870	3971	7,650
No. of water molecules	136	644	443
No. of ligand atoms	48	107	174
Root mean square deviation, bond lengths (Å)	0.013	0.007	0.008
Root mean square deviation, bond angles (°)	1.21	1.18	1.13
Average *B*-factor (Å2), protein, solvent	37.9, 37.7, 41.8	20.2,18.2,33.1	41.2, 41.2, 42.4
Stereochemistry, most favored, allowed, outliers			

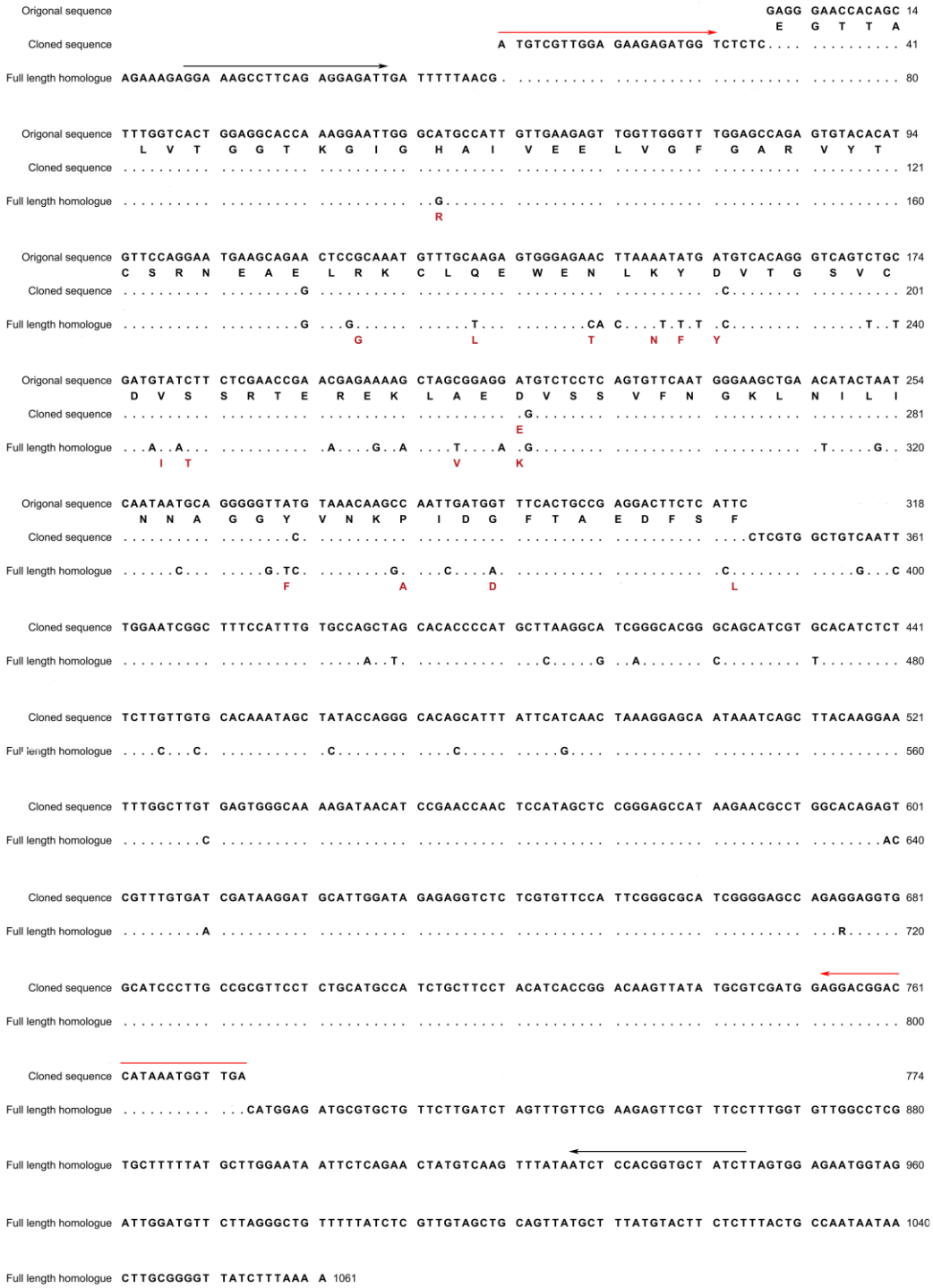

Figure D.1. Alignment, generated with CLC main workbench version 6.9.1, of the original sequence (medp_9narc_20101112|58880), the full length homologue (medp_9narc_20101112|12438), and the cloned sequence (*NNR*). Dots indicate identical base pairs, identical amino acids are not shown, altered amino acids are shown in red below the corresponding sequence, black arrows are over outer primer sequences, and red arrows are over inner primer sequences.

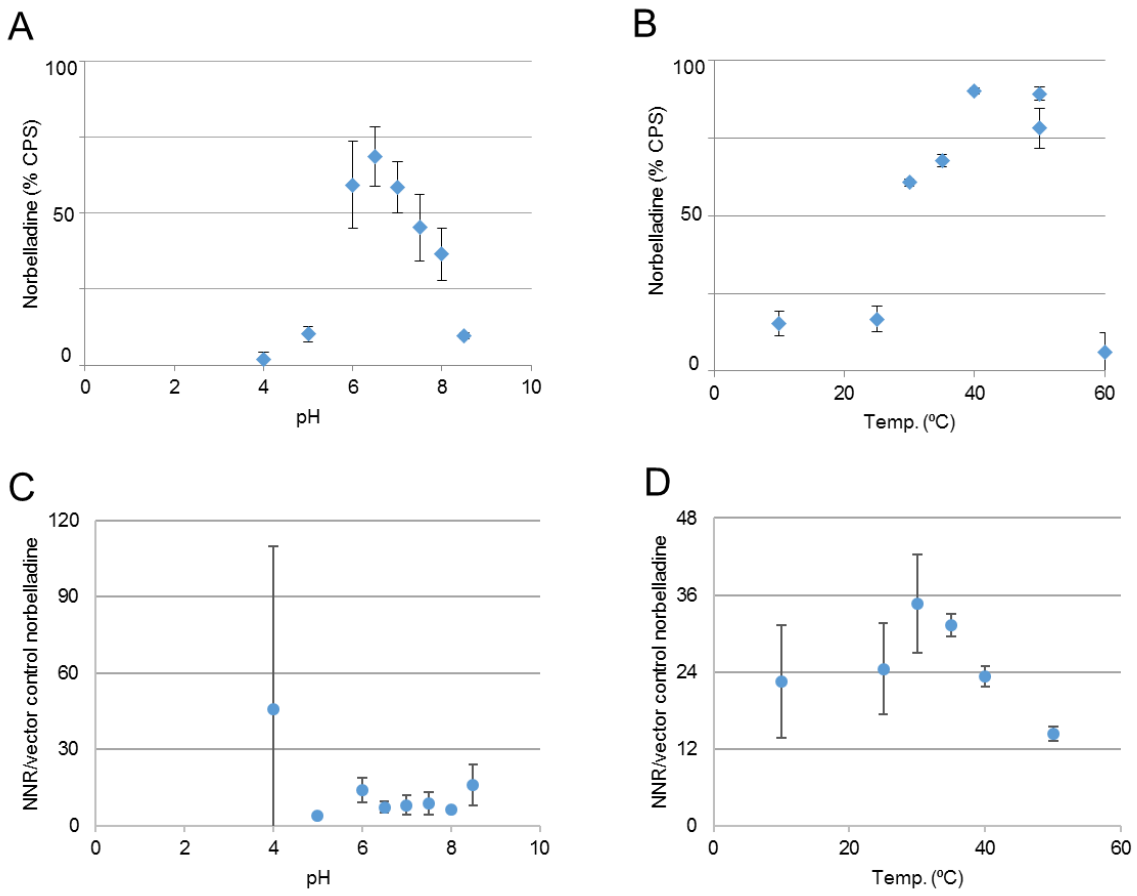

Figure D.2. pH and temperature optima for NNR using MRM based quantification of norbelladine. (A) pH optimum for NNR. (B) Temperature optimum for NNR. (C) NNR product:vector control background ratio with pH on the x-axis. (D) NNR product:vector control background ratio with temperature on the x-axis. Abbreviations counts per second (CPS) and noroxomaritidine/norcraugsodine reductase (NNR).

D.2 References

Beaudoin, F., Wu, X., Li, F., Haslam, R.P., Markham, J.E., Zheng, H., Napier, J.A., and Kunst, L. (2009). Functional characterization of the Arabidopsis beta-ketoacyl-coenzyme A reductase candidates of the fatty acid elongase. Plant Physiol. **150,** 1174-1191.

Berthold, D.A. (1998). Isolation of mutants of the *Arabidopsis thaliana* alternative oxidase (ubiquinol:oxygen oxidoreductase) resistant to salicylhydroxamic acid. Biochim. Biophys. Acta. **1364,** 73-83.

Bilyeu, K.D., Cole, J.L., Laskey, J.G., Riekhof, W.R., Esparza, T.J., Kramer, M.D., and Morris, R.O. (2001). Molecular and biochemical characterization of a cytokinin oxidase from maize. Plant Physiol. **125,** 378-386.

Bonin, C.P., and Reiter, W.D. (2000). A bifunctional epimerase-reductase acts downstream of the MUR1 gene product and completes the *de novo* synthesis of GDP-L-fucose in Arabidopsis. Plant J. **21,** 445-454.

Chang, C., and Meyerowitz, E.M. (1986). Molecular cloning and DNA sequence of the *Arabidopsis thaliana* alcohol dehydrogenase gene. Proc. Natl. Acad. Sci. U.S.A. **83,** 1408-1412.

Chen, M., Lin, J., Ma, Y., and Wei, D. (2010). Characterization of a novel NADPH-dependent oxidoreductase from Gluconobacter oxydans. Mol. Biotechnol. **46,** 176-181.

Davletova, S., Rizhsky, L., Liang, H., Shengqiang, Z., Oliver, D.J., Coutu, J., Shulaev, V., Schlauch, K., and Mittler, R. (2005). Cytosolic ascorbate peroxidase 1 is a central component of the reactive oxygen gene network of Arabidopsis. Plant Cell **17,** 268-281.

Engqvist, M., Drincovich, M.F., Flügge, U.I., and Maurino, V.G. (2009). Two D-2-hydroxy-acid dehydrogenases in *Arabidopsis thaliana* with catalytic capacities to participate in the last reactions of the methylglyoxal and beta-oxidation pathways. J. Biol. Chem. **284,** 25026-25037.

Funck, D., Eckard, S., and Müller, G. (2010). Non-redundant functions of two proline dehydrogenase isoforms in Arabidopsis. BMC Plant Biol. **10,** 70.

Funck, D., Winter, G., Baumgarten, L., and Forlani, G. (2012). Requirement of proline synthesis during Arabidopsis reproductive development. BMC Plant Biol. **12,** 191.

Geissler, R., Brandt, W., and Ziegler, J. (2007). Molecular modeling and site-directed mutagenesis reveal the benzylisoquinoline binding site of the short-chain dehydrogenase/reductase salutaridine reductase. Plant Physiol. **143,** 1493-1503.

Guo, L., and Paiva, N.L. (1995). Molecular cloning and expression of alfalfa (*Medicago sativa* L.) vestitone reductase, the penultimate enzyme in medicarpin biosynthesis. Arch. Biochem. Biophys. **320,** 353-360.

Halpin, C., Holt, K., Chojecki, J., Oliver, D., Chabbert, B., Monties, B., Edwards, K., Barakate, A., and Foxon, G.A. (1998). Brown-midrib maize (bm1)--a mutation affecting the cinnamyl alcohol dehydrogenase gene. Plant J. **14,** 545-553.

Herrero, S., González, E., Gillikin, J.W., Vélëz, H., and Daub, M.E. (2011). Identification and characterization of a pyridoxal reductase involved in the vitamin B6 salvage pathway in Arabidopsis. Plant Molecular Biology **76,** 157-169.

Horie, Y., Ito, H., Kusaba, M., Tanaka, R., and Tanaka, A. (2009). Participation of chlorophyll b reductase in the initial step of the degradation of light-harvesting chlorophyll a/b-protein complexes in Arabidopsis. J. Biol. Chem. **284,** 17449-17456.

Lazar, G., Zhang, H., and Goodman, H.M. (1993). The origin of the bifunctional dihydrofolate reductase-thymidylate synthase isogenes of *Arabidopsis thaliana*. Plant J. **3,** 657-668.

Lester, D.R., Ross, J.J., Davies, P.J., and Reid, J.B. (1997). Mendel's stem length gene (Le) encodes a gibberellin 3 beta-hydroxylase. Plant Cell **9,** 1435-1443.

Michalecka, A.M., Svensson, A.S., Johansson, F.I., Agius, S.C., Johanson, U., Brennicke, A., Binder, S., and Rasmusson, A.G. (2003). Arabidopsis genes encoding mitochondrial type II NAD(P)H dehydrogenases have different evolutionary origin and show distinct responses to light. Plant Physiol. **133,** 642-652.

Moummou, H., Kallberg, Y., Tonfack, L.B., Persson, B., and van der Rest, B. (2012). The plant short-chain dehydrogenase (SDR) superfamily: genome-wide inventory and diversification patterns. BMC Plant Biol. **12,** 219.

Nandi, A., Welti, R., and Shah, J. (2004). The *Arabidopsis thaliana* dihydroxyacetone phosphate reductase gene SUPPRESSSOR OF FATTY ACID DESATURASE DEFICIENCY1 is required for glycerolipid metabolism and for the activation of systemic acquired resistance. Plant Cell **16,** 465-477.

Oosawa, N., Masuda, T., Awai, K., Fusada, N., Shimada, H., Ohta, H., and Takamiya, K. (2000). Identification and light-induced expression of a novel gene of NADPH-protochlorophyllide oxidoreductase isoform in *Arabidopsis thaliana*. FEBS Lett. **474,** 133-136.

Paris, S., Wessel, P.M., and Dumas, R. (2002). Overproduction, purification, and characterization of recombinant bifunctional threonine-sensitive aspartate kinase-homoserine dehydrogenase from *Arabidopsis thaliana*. Protein Expr. Purif. **24,** 105-110.

Preisig-Müller, R., Gühnemann-Schäfer, K., and Kindl, H. (1994). Domains of the tetrafunctional protein acting in glyoxysomal fatty acid beta-oxidation. Demonstration of epimerase and isomerase activities on a peptide lacking hydratase activity. J. Biol. Chem. **269,** 20475-20481.

Ringer, K.L., Davis, E.M., and Croteau, R. (2005). Monoterpene metabolism. Cloning, expression, and characterization of (-)-isopiperitenol/(-)-carveol dehydrogenase of peppermint and spearmint. Plant Physiol. **137,** 863-872.

Roje, S., Wang, H., McNeil, S.D., Raymond, R.K., Appling, D.R., Shachar-Hill, Y., Bohnert, H.J., and Hanson, A.D. (1999). Isolation, characterization, and functional expression of cDNAs encoding NADH-dependent methylenetetrahydrofolate reductase from higher plants. J. Biol. Chem. **274,** 36089-36096.

Saito, K., Yonekura-Sakakibara, K., Nakabayashi, R., Higashi, Y., Yamazaki, M., Tohge, T., and Fernie, A.R. (2013). The flavonoid biosynthetic pathway in Arabidopsis: structural and genetic diversity. Plant Physiol. Biochem. **72,** 21-34.

Sato, Y., Morita, R., Katsuma, S., Nishimura, M., Tanaka, A., and Kusaba, M. (2009). Two short-chain dehydrogenase/reductases, NON-YELLOW COLORING 1 and NYC1-LIKE, are required for chlorophyll b and light-harvesting complex II degradation during senescence in rice. Plant J. **57,** 120-131.

Schweiger, P., Gross, H., and Deppenmeier, U. (2010). Characterization of two aldo-keto reductases from Gluconobacter oxydans 621H capable of regio- and stereoselective alpha-ketocarbonyl reduction. Appl. Microbiol. Biotechnol. **87,** 1415-1426.

Sibout, R., Eudes, A., Mouille, G., Pollet, B., Lapierre, C., Jouanin, L., and Séguin, A. (2005). CINNAMYL ALCOHOL DEHYDROGENASE-C and -D are the primary genes involved in lignin biosynthesis in the floral stem of Arabidopsis. Plant Cell **17,** 2059-2076.

Simpson, P.J., Tantitadapitak, C., Reed, A.M., Mather, O.C., Bunce, C.M., White, S.A., and Ride, J.P. (2009). Characterization of two novel aldo-keto reductases from Arabidopsis: expression patterns, broad substrate specificity, and an open active-site structure suggest a role in toxicant metabolism following stress. J. Mol. Biol. **392,** 465-480.

Somers, D.E., Caspar, T., and Quail, P.H. (1990). Isolation and characterization of a ferredoxin gene from *Arabidopsis thaliana*. Plant Physiol. **93,** 572-577.

Sparla, F., Tedeschi, G., Pupillo, P., and Trost, P. (1999). Cloning and heterologous expression of NAD(P)H:quinone reductase of *Arabidopsis thaliana*, a functional homologue of animal DT-diaphorase. FEBS Lett. **463,** 382-386.

Stepansky, A., Yao, Y., Tang, G., and Galili, G. (2005). Regulation of lysine catabolism in Arabidopsis through concertedly regulated synthesis of the two distinct gene products of the composite AtLKR/SDH locus. J. Exp. Bot. **56,** 525-536.

Tavladoraki, P., Rossi, M.N., Saccuti, G., Perez-Amador, M.A., Polticelli, F., Angelini, R., and Federico, R. (2006). Heterologous expression and biochemical characterization of a polyamine oxidase from Arabidopsis involved in polyamine back conversion. Plant Physiol. **141,** 1519-1532.

Unterlinner, B., Lenz, R., and Kutchan, T.M. (1999). Molecular cloning and functional expression of codeinone reductase: the penultimate enzyme in morphine biosynthesis in the opium poppy *Papaver somniferum*. Plant J. **18,** 465-475.

Zhao, S., Lin, Z., Ma, W., Luo, D., and Cheng, Q. (2008). Cloning and characterization of long-chain fatty alcohol oxidase LjFAO1 in *Lotus japonicus*. Biotechnol. Prog. **24,** 773-779.

Ziegler, J., Voigtländer, S., Schmidt, J., Kramell, R., Miersch, O., Ammer, C., Gesell, A., and Kutchan, T.M. (2006). Comparative transcript and alkaloid profiling in Papaver species identifies a short chain dehydrogenase/reductase involved in morphine biosynthesis. Plant J. **48,** 177-192.