People of mediocre ability sometimes achieve outstanding success because they don't

know when to quit. Most men succeed because they are determined to.

-George Allen, Sr.,

The γ-tocopherol-like family of *N*-methyltransferases: A taxonomically clustered gene

family encoding enzymes responsible for *N*-methylation of monoterpene indole alkaloids

Dylan Edward Ryan Levac, B.Sc.

Biotechnology

Submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Faculty of Biological Sciences, Brock University

St. Catharines, Ontario

©2013

ABSTRACT

The plant family Apocynaceae accumulates thousands of monoterpene indole alkaloids (MIAs) which originate, biosynthetically, from the common secoiridoid intermediate, strictosidine, that is formed from the condensation of tryptophan and secologanin molecules. MIAs demonstrate remarkable structural diversity and have pharmaceutically valuable biological activities. For example; a subunit of the potent anti-neoplastic molecules vincristine and vinblastine is the aspidosperma alkaloid, vindoline. Vindoline accumulates to trace levels under natural conditions. Research programs have determined that there is significant developmental and light regulation involved in the biosynthesis of this MIA. Furthermore, the biosynthetic pathway leading to vindoline is split among at least five independent cell types. Little is known of how intermediates are shuttled between these cell types.

The late stage events in vindoline biosynthesis involve six enzymatic steps from tabersonine. The fourth biochemical step, in this pathway, is an indole *N*-methylation performed by a recently identified *N*-methyltransfearse (NMT). For almost twenty years the gene encoding this NMT had eluded discovery; however, in 2010 Liscombe *et al*. reported the identification of a γ-tocopherol *C*-methyltransferase homologue capable of indole *N*-methylating 2,3-dihydrotabersonine and Virus Induced Gene Silencing (VIGS) suppression of the messenger has since proven its involvement in vindoline biosynthesis.

Recent large scale sequencing initiatives, performed on non-model medicinal plant transcriptomes, has permitted identification of candidate genes, presumably involved, in MIA biosynthesis never seen before in plant specialized metabolism research. Probing the transcriptome assemblies of *Catharanthus roseus* (L.)G.Don,

Vinca minor L., *Rauwolfia serpentine* (L.)Benth *ex* Kurz, *Tabernaemontana elegans*, and *Amsonia hubrichtii*, with the nucleotide sequence of the *N*-methyltransferase involved in vindoline biosynthesis, revealed eight new homologous methyltransferases. This thesis describes the identification, molecular cloning, recombinant expression and biochemical characterization of two picrinine NMTs, one from *V. minor* and one from *R. serpentina*, a perivine NMT from *C. roseus*, and an ajmaline NMT from *R. serpentina*. While these TLMTs were expressed and functional *in planta*, they were active at relatively low levels and their *N*-methylated alkaloid products were not apparent our from alkaloid isolates of the plants. It appears that, for the most part, these TLMTs, participate in apparently silent biochemical pathways, awaiting the appropriate developmental and environmental cues for activity.

ACKNOWLEDGEMENTS

First and foremost, I wish to express my appreciation to those who have contributed directly the planning and execution of this work. I am extremely grateful to Professor Vincenzo De Luca. His guidance and helpful discussions, especially during the difficult stages, has been critical to my success. Dr. De Luca has guided me through a maturation process, both as a professional scientist and as a person. I am indebted for his understanding, compassion and support during tough times, especially a difficult 2012. I would like to also thank the rest of my thesis committee: Prof. Charles Després and Prof. Costa Metallinos for their encouragement, and insightful comments. I begin my preparations for every experiment, or every presentation, with considering a very specific question; 'what criticism will Charles ask?' This extra pressure has given me an additional push to develop as a scientist and public speaker.

This thesis would not have been possible had it not been for the efforts of two key individuals. I would like to thank above all Dr. Jun Murata. Although Jun left before my Ph.D. began, he has been central to my professional development, molding me into the scientist I am today. Had it not been for his seemingly endless desire to help and mentor me I do not believe I would be here to defend this thesis. I would also like to thank Dr. Fang Yu. Fang has been key to my further maturation as a scientist. Where Jun would imagine elegant approaches to solving problems, Fang has taught me brute force, persistence, and a seemingly endless well of optimism. I apply these opposing philosophies not only my research, but also in my personal life. I cannot thank you both enough.

more reasonable and shorter time frame. One cannot be an official student for forever and I am sure you are sighing in relief that I am finally moving to the work force.

Finally, I would like to acknowledge the most important person in my life – my fiancée, Brandi. She has been a constant source of strength and support. I would have never completed this program without Brandi and I look forward to spending the rest of our lives together.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

16OMT – 16-Hydroxytabersonine 16-*O*-Methyltransfearse

AAE – Acetylajmaline Esterase

Ac – Cyanogenic glucoside dominant genotype

AdoMet – S-Adenosyl-L-Methionine

AjNMT – Ajmaline *N*-Methyltransferase

BCIP – 5-Bromo-4-Chloro-3'-Indoylphosphate

BLAST – Basic Local Alignment Search Tool

CAS - 1% Cerium (IV) Ammonium Sulfate in 85% phosphoric acid spray

cDNA – Complementary DNA

CHI – Chalcone Isomerase

CPR – Ferrihemoprotein Reductase

Ct – Threshold Cycle

CYP – Cytochrome P450 Monoxygenase

D4H – Desacetoxyvindoline 4-Hydroxylase

DAT – Deacetylvindoline Acetyltransferase

DhtNMT – 2,3-Dihydro 3-Hydroxytabersonine *N*-Methyltransferase

DNA – Deoxyribonucleic Acid

ESI – Electrospray Ionization

EST – Expressed Sequence Tag

F5H – Ferulate – 5 – Hydroxylase

FDA – Food and Drug Administration

G10H – Geraniol 10- Hydroxylase

HTS – High Throughput Screening

IPAP – Internal Phloem-Associated Parenchyma

LAMT – Loganic Acid Methyltransferase

Li – β-glucosidase dominant genotype

matk – Maturase K

MEGA – Molecular Evolutionary Genetics Analysis

MEP - 2-*C*-Methyl-D-Erythritol 4-Phosphate

MIA – Monoterpenoid indole alkaloid

MP – Maximum Parsimony

Mr – Relative Molecular Mass

MS – Mass Spectrometry

MuSCLE – Multiple Sequence Comparison by Log-Expectation

NBT – Nitro Blue Tetrazolium Chloride

NME – New Molecular Entities

NMI – *N*-Methylated Molecule Index

NMR - Nuclear Magnetic Resonance

NMT – *N*-methyltransferase

NSERC - Natural Sciences and Engineering Research Council of Canada

OMT – *O*-Methyltransfearse

ORF – Open Reading Frame

PAGE – Polyacrylamide Gel Electrophoresis

PCR – Polymerase Chain Reaction

PDA – Photo Diode Array

PeNMT – Perivine *N*-Methyltransferase

PiNMT – Picrinine *N*-Methyltransfease

PNAE – Polyneuridine Aldehyde Esterase

PR – Perakine Reductase

PSORT – Protein Subcellular Localization Prediction

PVDF - Polyvinylidene Difluoride

RACE – Rapid Amplification of cDNA Ends

Rf – Retention Factor

RG – Raucaffricine β-glucosidase

RNA – Ribonucleic Acid

RPM – Revolutions Per Minute

Rt – Retention Time

RT – Reverse Transcription

RT-PCR – Reverse Transcription Polymerase Chain Reaction

SDS – Sodium Dodecyl Sulfate

SG – Stricotidine β-glucosidase

SGD – Strictosidine β-Glucosidase

S-Lignin –

SLS – Secologanin Synthase

SQD – Single Quadruple Detector

STR – Strictosidine Synthase

Syringyl monolignols

T16H – Tabersonine 16-Hydroxylase

TDC – Tryptophan Decarboxylase

TLC – Thin-Layer Chromatography

TLMT – Tocopherol-Like Methyltransferase

ToL – Tree of Life

UPLC - Ultra Performance Liquid Chromatography

VIGS – Virus Induced Gene Silencing

VS – Vinorine Synthase

xg – Times the force of gravity

γTMT - γ-Tocopherol Methyltransferase

CHAPTER 1 – INTRODUCTION

The plant family Apocynaceae is composed of almost 500 species of plants that cumulatively accumulate thousands of monoterpene indole alkaloids (MIAs) which demonstrate remarkable structural diversity and have pharmaceutically valuable biological activities (Potgeiter, K., and Albert, V., 2001). There has been over 50 years of intense biochemical and more recently, molecular investigation to elucidate the sequence of biochemical pathways, their molecular regulatory mechanisms and the cell specific compartmentalization involved in MIA biosynthesis (Facchini, P., De Luca, V., 2008). These investigations have depended on the non-model *Catharanthus roseus* whole plants and plant cell cultures, as well as *Rauvolfia serpentina* plant cell culture systems for pathway elucidation. Furthermore, up until recently the tool box available for these investigations was quite limited. Much of our molecular and biochemical understanding of MIA biosynthesis is derived from traditional protein purification and peptide sequencing, as well as homology based molecular cloning strategies (Facchini, P., De Luca, V., 2008).

The biosynthesis of vinblastine and vincristine, two potent anti-cancer drugs, requires the coupling of the MIAs, catharanthine and vindoline (Facchini, P., De Luca, V., 2008). Early investigations involving *C. roseus* plant cell cultures determined that while catharanthine accumulates readily under these artificial growth conditions, vindoline does not (Moreno, P., *et al.* 1995). This early observation spurred the study of how vindoline is biosynthesized in whole plants. These investigations have discovered that the conversion of tabersonine to vindoline is developmentally regulated, occurring in young leaf tissues and to be activated by light (Aerts, R., De Luca, V., 1992). Whole

plant studies also showed that MIA biosyntheis is divided between at least five different leaf cell types (internal phloem associated parenchyma, epidermis, mesophyll, idioblast, laticifer) (St-Pierre, B., *et al.* 1998; Oudin, A., *et al.* 2007; Levac, D., *et al.* 2008).

The conversion of tabersonine into vindoline requires six enzymatic steps that include two hydroxylations, an *O*-methylation, an *N*-methylation and an *O*-acetylation which have been characterized at the biochemical and molecular level (Facchini, P., De Luca, V., 2008). At present, only the fourth to last step in vindoline biosynthesis, involving a hydratase, remains to be identified. Recent efforts in several laboratories have used Virus Induced Gene Silencing (VIGS), to transiently suppress candidate hydratases in *C. roseus* and the gene involved should soon be available (Liscombe, D., O'Connor, S., 2011; Glenn, W., *et al.* 2012; De Luca, V., *et al*. 2012).

For almost twenty years the gene encoding the *N*-methyltransferase (NMT) involved in vindoline biosynthesis has eluded discovery. Traditional protein purification approaches, which had been successful in identifying the 16-hydroxytabersonine-*O*-methyltransferase involved in vindoline biosynthesis (Levac, D., *et al*. 2008), were not effective since this NMT appeared to be associated with chloroplast thylakoid membranes (Dether, M., De Luca, V., 1993). However, in 2010 Liscombe *et al*. reported the identification of a γ-tocopherol *C*-methyltransferase homologue capable of indole *N*-methylating 2,3-dihydrotabersonine, an artificial substrate accepted by the natural enzyme purified from *C. roseus* chloroplasts (Liscombe, D., *et al.* 2010; Dethier, M., De Luca, V., 1993). VIGS suppression of the corresponding messenger has since proven its involvement in vindoline biosynthesis (Liscombe, D., O'Connor, S., 2011). Moreover, discovery of this methyltransferase has enabled the identification of an entirely new

family of enzymes; the Tocopherol-Like Methyltransferase family (TLMT) (Liscombe, D., *et al.* 2010).

1.1 THESIS OUTLINE

The studies described in this thesis occurred during a very exciting period for biochemists and molecular biologists involved in elucidating plant specialized biochemical pathways. The research presented in this thesis describes how new methodologies of high throughput sequencing and comparative bioinformatics are dramatically increasing the rate of candidate gene discovery involved in biosynthetic pathway steps. Where, in the past, not only was substrate availability limiting for functional characterization of enzymes, so were candidate genes to screen. This paradigm has shifted however. Today, we are inundated with candidate genes that require new, creative triage strategies to prune them to manageable levels for functional analysis, or alternatively, that require development of high throughput screening methods to speed up their functional characterization. This thesis describes some philosophies relating to each of those principles.

Chapter 2 is a literature review which examines plant specialized metabolic pathways. This review focuses on the biosynthetic costs associated with specialized metabolism and documents the prevailing theories on how and why plants re-allocate resources towards or away from specialized metabolism in ecological time frames, as well as how plant populations evolve to balance metabolic costs with biotic and abiotic environmental stresses. This review concludes by discussing why most naturally occurring small molecules have no apparent function and how silent metabolic pathways

could be a strategy that plant populations employ to maximize chemical diversity over time.

Chapter 3 is a manuscript that has been accepted with revisions. This study shows that most or all of the Monoterpenoid indole alkaloids (MIA) of Eurasian *Vinca minor*, African *Tabernaemontana elegans* and American *Amsonia hubrichtii* accumulate in leaf wax exudates, while the rest of the leaf is almost devoid of alkaloids. Biochemical studies with *Vinca minor* show that tryptopan decarboxylase (TDC) activity and protein accumulation occur preferentially in epidermis-enriched leaf extracts compared to whole leaves. *In situ* hybridization studies localize TDC and strictosidine synthase (STR) to the upper and lower epidermis of *V. minor* and *T. elegans*. These observations suggest that biosynthesis of MIAs in epidermal, or closely related cells and their secretion to plant surfaces may have evolved as a common feature in MIA-producing members of the Apocynaceae.

Chapter 4 is a manuscript that is ready for submission. This manuscript discusses how *N*-methylated MIAs are taxonomically clustered in Apocynaceae and appear to be clustered in plants that express TLMTs. This co-clustering has been used to suggest the possible utility of similar methods to identify genes involved in interesting chemistries that could be exploited for biotechnological purposes, such as generating improved therapeutics through combinatorial biosynthesis. This manuscript reports the identification and functional characterization of two TLMTs; a perivine NMT from *C. roseus*, and an ajmaline NMT from *R. serpentina*.

Chapter 5 is a manuscript that is ready for submission. This study describes the biochemical and molecular characterization of two members of the TLMT family of enzymes from *Vinca minor* (VmPiNMT) and *Rauvolfia serpentina* (RsPiNMT). Recombinant VmPINMT and RSPiNMT were shown to have high substrate specificity and affinity for picrinine, converting it to *N*-methylpicrinine (ervincine). Both enzymes also accepted 21-cyclolochnericine for *N*-methylation suggesting that the cyclic ether ring system, a component of the hemiaminal functional group of both molecules, is important for substrate recognition and enzyme methylation. Developmental studies with *V. minor* and *R. serpentina* showed that *RsPiNMT* and *VmPiNMT* gene expression and biochemical activities were highest in younger leaf tissues, as also found with developmental studies of MIA biosynthesis in *Catharanthus roseus* leaves. Whereas the results suggest that MIA production is an early event in the developmental program of these plants, only trace levels of picrinine accumulate in *R. serpentina* and *V. minor*, whereas ervincine has only been detected in *Vinca erecta*. These results suggest that under certain environmental conditions both species of plants may be able to produce and accumulate these *N*-methylated MIAs.

Chapter 6 discusses the keys to my success, potential application of the methods and the philosophies that lead to their development, as well as the potential application of discoveries documented in this thesis. This chapter also discusses some potential future projects that could be fruitful avenues of research, discovering new mechanisms for putativeprotein-membrane association, as well as elucidating the crystalstructure of this new protein family, the TLMTs.

1.2 BIBLIOGRAPHY

Aerts, R., De Luca, V., (1992) Phytochrome is involved in the light-regulation of vindoline biosynthesis in *Catharanthus*, *Plant Physiology*, **100**, 1029-1032

De Luca, V., Salim, V., Masada-Atsumi, S., Yu, F., (2012) Mining the biodiversity of plants: A revolution in the making, *Science*, **336**, 1658-1661

Dethier, M., De Luca, V., (1993) Partial purification of an N-methyltransferase involved in vindoline biosynthesis in Catharanthus roseus, *Phytochemistry*, **32**, 673-678.

Facchini, P., De Luca, V., (2008) Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants, *The Plant Journal*, **54**, 763-784.

Glenn, W., Runguphan, W., O'Connor, S., (2012) Recent progress in the metabolic engineering of alkaloids in plant systems, *Current Opinion in Biotechnology*, In Press

Levac, D., Murata, J., Kim, W., De Luca, V., (2008) Application of carborundum abrasion technique for investigating the leaf epidermis: molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-*O*-methyltransferase, *The Plant Journal*, **53**, 225-236

Liscombe, D., O'Connor, S., (2011) A virus induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*, *Phytochemistry*, **72**, 1969-1977.

Moreno, P., van der Heijden, R., Verpoorte, R., (1995) Cell and tissue cultures of *Catharanthus roseus*: A literature survey**,** *Plant Cell, Tissue and Organ Culture*, **42**, 1-25

Oudin, A., Mahroug, S., Courdavault, V., Hervouet, N., Zelwer, C., Rodriguez-Concepcion, St-Pierre, B., Burlat, V., (2007) Spatial distribution and hormonal regulation of gene productions from methyl erythriotol phosphate and monoterpene-secoiridoid pathways in *Catharanthus roseus*, *Plant Biolecular Biology*, **65**, 13-30

Potgierter, K., Albert, V., (2001) Phylogenetic relationships within Apocynaceae s.1. Based on trnL Intron and trnL-F spacer sequences and propagule characters, *Annals of the Missouri Botanical Garden*, **88**, 523-549

CHAPTER 2 – LITERATURE REVIEW –SPECIALIZED METABOLISM AND SILENT BIOCHEMICAL PATHWAYS

2.1 LIFE NECESSITATES SPECIALIZED ADAPTATIONS

Living organisms require diverse adaptations that address specific survival and competitive challenges in order for them to reach the reproductive stage of life. Examples of structural adaptations in animals include the evolution of avian wings in birds (Dial, K., 2003), canine teeth in carnivores (Valkenburgh, B., 1987), opposable thumbs and enhanced palm musculature in primates (Ambrose, S., 2001), and enlarged brains in *Homo sapiens* (Kappelman, J., 1996). More subtle adaptations, such as animal and insect social behaviour (Hamilton, W., 1964), or endurance running in bipedal apes (Bramble, D., Lieberman, D., 2004), are just as important for the survival and reproduction of these organisms and yet these adaptations are almost entirely overlooked by the general, non-scientific population. However, scientists have shown keen interest into the role(s) played by such adaptations in survival of such organisms. If subtle adaptations are deeply fascinating to the scientific community, then plants present the most remarkable research opportunities of all.

As immobile organisms, plants cannot chase down prey, or move to escape herbivores. Instead they have evolved elegant strategies to survive and reproduce while maintaining fixed positions. For example, carnivorous plants have adapted to grow in soils that are poor in nutrients, by trapping insect prey and harnessing these nutrients by

digesting them. *Dionaea muscipula*, the venus fly trap, captures its prey by employing a mechanism that is analogous to trip-wire activated snares (Forterre, Y., *et al.,* 2005), while *Nepenthes albomarginata*, attracts termites by emitting volatile scents (Giusto, B., *et al.* 2008) and presenting white hairs that resemble food to termites (Merbach, M., *et al.* 2002). While reaching for the hairs, termites fall into a viscous liquid trap within the pitcher and are unable to climb out. At the base of the pitcher the termites are digested, supplying nourishment to the plant. While carnivorous plants are generally considered exceptions to the photo-autotrophic rule of the plant kingdom, emissions of volatile attractants by the *Nepenthes* genus is very much a subtle adaptation in line with how other plants access nutrition and cope with natural stresses.

2.2 EVOLUTIONARY EXPANSION OF CHEMICAL STRUCTURES FOR DIVERSE BIOLOGICAL FUNCTION

Ernst Stahl, considered to be one of the fathers of plant chemical ecology, was one of the first to test hypotheses about how plants deploy chemical defences to combat herbivory (Hartmann, T., 2008). Remarkably, these early experiments were entirely forgotten by early 20th century biochemists and organic chemists who were more involved in investigating the molecular structure and function of compounds rather than their raison d'être (Fraenkel, G., 1969). Fraenkle and those that followed him continued chemical ecological studies that have led to detailed understanding of the mechanisms involved for maintaining the chemical diversity of plants.

The volatile attractant bouquet emitted by *Nepenthes* is composed of fatty acid derivatives, benzene-containing molecules, monoterpenes and sesquiterpenes (Giusto, B., *et al.* 2010). In a generalized sense these molecules are referred to as specialized metabolites, or phytochemicals. Historically, they were referred to as secondary metabolites, due to the belief that they were products of aberrant plant metabolism and not essential to the biology of the organism. We now know that small molecules are central to biological processes and plants synthesize specialized metabolites to perform many core biological functions. For example: communicating with other plants (Farmer, E., Ryan, C., 1990); attracting pollinators for fertilization and reproduction (Kessler, D., Baldwin, I., 2006); defending against herbivory and pathogens (Pare´, P., Tumlinson, J., 1999; Wink, M., 1988); as is accessing and maintaining control of natural resources (Hierro, J., Callaway, R., 2003). One could argue that plants are strikingly similar to other advanced organisms, except they have evolved different strategies to survive and reproduce.

The biodiversity of plants, estimated at approximately 250,000 species, is matched by over 100,000 unique specialized metabolite structures that have been identified (Verporte, R., 1998). Between 25,000 to 40,000 of these structures are terpenes (McCaskill, D., Croteau, R., 1997), 12,000 are alkaloids (Facchini, P., 2001), and 8,000 are phenylpropanoids (Croteau, R., *et al*. 2000). In addition to these three main specialized metabolite classes, some plant taxa synthesize hundreds of glucosinolates (Fahey, J., *et al.* 2001), cyanogenic glucosides (Vetter, J., 2000) and other structures. If generating elaborate small molecule structural diversity has been the primary strategy of plants to colonize hostile biotic and abiotic niches, compete with

and/or communicate with other plants, attract pollinators, and to deter herbivores and pathogens, it should be possible to document the chemical escalation involved in these adaptations over evolutionary time.

Remarkably, among the few documented studies of such chemical escalation, it is worth considering the 'squirting' genus *Bursera. Bursera* is composed of approximately 100 species native to the southern United States, tropical Central America and Peru. This genus uses both mechanical and chemical defenses against chrysomelid beetles, a typical insect herbivore of *Bursera* (Bacerra, J., Venable, D., 1990). Molecular phylogentics of 70 *Bursera* species were used to correlate the expansion of chemical defenses within this genus along with species diversification (Becerra, J., *et al*. 2009); striking positive-correlations were obtained among *Bursera* lacking the mechanical 'squirt gun' defense. Among the *Bullockia* subgenus of *Bursera*, a collection of plants which specialize exclusively in chemical defenses against beetles, there was a very strong positive correlation between expansion of specialized metabolite chemistry and species diversification. This positive correlation was still apparent within mechanically defended *Bursera*, but the relationship was less significant (Becerra, J., *et al*. 2009). The expansion of chemical diversity during speciation suggests that there must be ecological pressures favouring chemical diversity in plants. This is particularly evident in plant species that are exclusively dependent on chemical defences, such as in the sub-genus *Bullockia*. In contrast, development and maintenance of chemical diversity appears to come at a cost. This is evidenced by the reduced chemical diversity (one to two terpenes) in most members of the Bursera genus that can also defend themselves through mechanical means (Becerra, J., *et al*. 2001).

2.3 THE COST OF SPECIALIZED METABOLISM

For the last 30 years ecologists have been applying established economic principles in efforts to explain plant life-history models (Perrin, N., 1992; Perrin, N., 1993) and resource allocation models (Lerdeau, M., 1992). While there has been some success with economic models explaining how plants dedicate resources to growth and resource storage (Monteith, J., Elston, J., 1983; Bloom, A., 1986), evaluating cost optimization of plant life-history models is remarkably difficult under natural conditions (Chapin, F., 1989), and while testing plant resource allocation models isn't trivial, it may prove less difficult (Gershenzon, J., 1994).

Friedrich von Wieser suggested that the cost of anything should be viewed as an opportunity cost, whereby the opportunity cost is the difference between mutually exclusive objectives. Plants, like other organisms, dedicate resources for functions that are considered primary, essential to life (e.g. growth, differentiation and reproduction). However, since they also dedicate resources for chemical defenses, plants must establish a delicate balance between resource use for primary and specialized metabolism and scientists have attempted to explain this balance through resource availability models (e.g. carbon-nitrogen balance, growth-differentiation balance, and resource availability hypotheses) (Gershenzon, J., 1994).

It is helpful to understand the overall cost of different specialized metabolite classes and the factors that contribute to their overall value, before we can appreciate the balance of resource allocation. The appropriate currency to measure costs depends on what specifically is being measured (Lerdau, M., 1993). The common currency which is

used to report costs associated with specialized metabolite synthesis is grams of glucose consumed per gram of metabolite synthesized (g glucose/g metabolite) (Gershenzon, J., 1994). The biochemical cost of selected plant primary and specialized small molecules, has been calculated using primary metabolite substrates, cofactors and co-substrate costs (Gershenzon, J., 1994). It should be mentioned that these values are speculative as our understanding of the biochemical pathways yielding specialized metabolites is mostly incomplete. Nevertheless it is helpful to appreciate that terpenes (3.2 g glucose/g) and alkaloids (3.2 g glucose/g) are the most costly compared to phenolics (2.11 g glucose/g), which cost significantly less (Gershenzon, J., 1994). The value of this appreciation will become apparent when discussing mobile vs. fixed chemical defenses.

2.4 BALANCING ALLOCATION COSTS

Plants interact with their environment through small molecules. For example; they communicate with other plants through volatilized plant hormones (Farmer, E., Ryan, C., 1991; Karban, R., *et al.* 2000), they communicate with animals (Schaefer, H., *et al.* 2004) or insects (Baldwin, I., *et al*. 2001) through pigmentation patterns, scent production and other mechanisms. Plants also develop resource sharing relationships with fungi (Behie, S., *et al.* 2012) and bacteria (Burns, R., Hardy, R., 1975), both of which are established through small molecule signals. The small molecules that mediate these interactions are synthesized through biochemical pathways that are, in many cases, not completely understood. However, what seems to be clear is that flux through these biochemical pathways is not always constant. In fact, plants demonstrate remarkable

metabolic plasticity. For example, temporal stresses, such as forest leaf canopy closure, cause shade tolerant *Trillium grandiflorum* to decrease its photosynthetic and respiratory rates (Taylor, R., Pearcy, R., 1976). To an extent plants also appear to be metabolically primed through evolution; for example, pervasive abiotic stress, like cold environmental conditions, has resulted in the evolution of reduced photosynthetic and respiratory rates in arctic shrubs when compared to non-arctic shrubs (Billings, W., Mooney, H., 1968). The degree, nature and quickness with which plants can respond to environmental stresses, both on ecological and evolutionary time scales, appears to be a function of resource availability. Insofar as the chemical responses produced, plants that grow in resource rich environments, appears to synthesize so called mobile defences, like alkaloids, cyanogenic glucosides, or glucosinolates, where as plants that grow in resource poor environments, like arctic shrubs, maintain permanent chemical defenses, like condensed tannins, polyphenols, and terpenes (Billings, W., Mooney, H., 1968; Grime, J., 1979; Coley, P., Bryant, F., Chapin, F., 1985; Gershenzon, J., 1984, 1994).

The tradeoffs made by plants between expending resources for growth or for specialized chemical production are called allocation costs (Herms, D., Mattson, W., 1992). Three theories attempt to predict how allocation costs are balanced in plants; the carbon-nitrogen balance theory (Brant, J., *et al.* 1983, Loomis, W., 1932; Herms, D., Mattson, W., 1992), the growth-differentiation balance theory (Loomis, W., 1932; Herms, D., Mattson, W., 1992), and the resource availability theory (Coley, P., *et al.* 1985; Gershenzon, J., 1994). Each theory provides different approaches to explain how specialized metabolite abundance and diversity are affected by abiotic and biotic stress

and how these adaptations enable plants to best survive the diversity of stresses that they experience under natural conditions.

2.4.1 THE CARBON-NITROGEN BALANCE THEORY

The carbon-nitrogen balance theory (Brant, J., et al. 1983), also referred to as the carbohydrate-nitrogen balance theory (Loomis, W., 1932), was developed to predict how access to light and/or the nutritional status of a plant, affects the composition and concentration of specialized metabolites in plants (Herms, D., Mattson, W., 1992). The most fundamental premise of this theory is that access to nutrients is more important than light quantity, or quality, for plant growth and reproduction. This theory suggests that nutritional deficits have a greater adverse affect on the plant than reduced photosynthetic rate (Herms, D., Mattson, W., 1992). Plants growing under environmental conditions that provide sufficient light for normal photosynthesis, but exposed to nitrogen limitation, appear to reprioritize their fixed carbon resources to enhance the biosynthesis and accumulation of carbon-based specialized metabolites, such as terpenes or phenolics (Gershenzon, J., 1994). A review of the empirical literature would, for the most part, seem to support this prediction; low nutrient conditions increased caffeoylquinic acids in *Helianthus annus* L. and *Nicotiana tabacum* L. (Del Moral, R., 1972; Armstrong, G., et al. 1970), anthocyanins in *Brassica oleracea* L., *Hordeum vulgare* L., *Sorghum bicolor* (L.) Moench, and *Vitis vinifera* L.,(Nozzolillo, C., 1978), polyphenols in cell suspension cultures of *Catharanthus roseus* (Knobloch, K., Berlin, J., 1980), and total phenolic compounds in *Helianthus annus* plants (Hall, A., *et al.* 1982), and *Nicotiana tabacum* cell suspension cultures (Knobloch, K., Berlin, J., 1981). Conversely, under environmental conditions where photosynthesis is limited, but nutrients are abundant, the C-N balance

theory predicts that the plant will grow at a normal, or accelerated, rate until growth is limited by other parameters not considered by the carbon-nitrogen balance theory. At this point, excess nitrogen will be repurposed for biosynthesis of nitrogen-based specialized metabolites, like alkaloids (Gershenzon, J., 1984). However, a review of the scientific literature suggests that this correlation may not always hold. For example; levels of amino-acid derived alkaloids, like morphine in *Papaver somniferum* L., nicotine in *N. tabacum*, and gramine in *Hordeum vulgare* L., increase when nitrogen availability is also increased (Nowwacki, E., et al. 1976). In contrast, alkaloids of mixed origin derived from amino acid and terpene components, like *C. roseus* MIAs remained unchanged when nitrogen was not rate limiting (Nowwacki, E., et al. 1976). Furthermore, nitrogen- and sulfur- containing glucosinolates found in *Brassica* rose over 120%, relative to baseline levels, in plants grown under nitrogen limiting conditions (Gershenzon, J., 1984). These studies suggest that the parameters governing the synthesis and accumulation of specialized metabolites are not only limited by the levels of carbon and nutrient availability (such as nitrogen), but other unaccounted for factors are also involved.

2.4.2 THE GROWTH-DEVELOPMENT BALANCE THEORY

The growth and development balance hypothesis broadens the carbon-nitrogen balance hypothesis by also considering environmental effects on both photosynthetic and sink tissues. This hypothesis predicts that environmental factors which perturb growth in general, while having limited, or no effect on photosynthetic rates, will also affect repurposing of photosynthates and other resources normally allocated to growth and

development, towards specialized metabolisms (Loomis, W., 1932; Herms, D., Mattson, W., 1992, Gershenzon, J., 1994). Clearly, this could take into account more factors that affect plant growth beyond the availability of carbon and nitrogen. For example; conditions such as drought, or high salinity, could reduce plant growth in favour of increased specialized metabolite synthesis that would not be predicted by the carbon-nitrogen balance hypothesis. Drought appears to enhance the accumulation of cyanogenic glucosides in *Sorghum bicolour* (Nelson, C., 1953), glucosinolates in cabbage and watercress (Freeman, G, Mossadeghi, N., 1971), and alkaloids in *Papaver somniferum* (Hofman, P., Menary, R., 1979) when compared to properly hydrated plants. In fact, *Cinchona ledgeriana* has been reported to yield no quinine when grown during the rainy season (Gershenzon, J., 1984).

Together, the carbon-nitrogen and growth-development balance theories explain the plastic response of plants to abiotic stresses on ecological time frames. While these theories explain plant responses to temporal nutrient or water deficits, or low lighting conditions, these theories do not attempt to explain how, or why, plants have evolved to be primed for these pervasive ecological conditions. Furthermore, these theories appear to ignore the fact that specialized metabolites and their specific molecular classes are distributed throughout the plant kingdom in taxonomically clustered patterns.

2.4.3 THE RESOURCE AVAILABILITY THEORY

Herbivory is a major, pervasive stress that plants must respond to if they hope to grow and reach reproductive stages of life. Approximately 10 percent of the biomass

produced by plants each year is consumed by herbivores (Coley, P., *et al.* 1985). The resource availability theory suggests that plants have evolved strategies to balance the value of biomass, lost to herbivores with the relative value of available resources (Coley, P., *et al.* 1985). For example, if ecological conditions place a premium value on the biomass together with its protection and the relative value of fixed or mobile chemical defences is less than this premium value, these conditinos will accelerate the evolution of plant populations with improved defence capabilities is accelerated. Examples of ecological conditions that put a cost premium on biomass protection are resource poor environments, such as the shaded understory of forests, or arid and arctic climates where access to light, adequate temperatures, nutrients and water availability is scarce. These sorts of environmental conditions significantly limit plant growth rates and as a consequence plant populations that evolve under these conditions also tend to accumulate more fixed chemical defences like condensed tannins, terpenes, and polyphenols when compared to plant populations that evolve in resource rich environments (Gershenzon, J., 1984, 1992). An example of heightened chemical defence in resource poor environments is the combined accumulation of fixed chemical defences with increased lignification in conifer leaves that dramatically reduces their palatability and decreases biomass loss to herbivores (Coley, P., *et al.* 1985).

In comparison, plant populations that colonize resource-rich environments tend to have resource allocation balances that place less value on biomass protection. Instead of devoting large resources to establish fixed chemical defences, similar to the situation of conifers, these plant populations are devoted to growth. Such plants respond to herbivores by mobilizing chemical defences though *de novo* biosynthesis of specialized biologically

active metabolites (Goley, P., *et al.* 1985) that may in some cases be recovered and repurposed through appropriate degradation pathways (Gershenzon, J., 1984, 1992). In addition, plant populations occupying resource-rich environments also appear to be capable of sacrificing foliage in lieu of maintaining high levels of fixed chemical defences if regeneration of foliage costs less than the synthesis and storage of the fixed defences (Gershenzon, J., 1992).

2.4.4 EVOLVED COST BALANCE; CYANOGENIC GLUCOSIDES

Plants that evolve unique and highly successful strategies to respond to environmental stresses are more likely to be maintained in plant populations over successive generations (Jones, D., 1972). The selective pressures that maintain successful genetic traits also maintain successful resource allocation balances, as discussed in the last section. So far we have considered resource balancing strategies insofar as the carbon-nitrogen and growth-differentiation theories, which account for responses in ecological time scales, as well as the resource allocation theory which accounts for plant defence strategies developed over evolutionary time scales. These theories, however, only predict binary plant responses to environmental stresses; whether they will synthesize carbon- or nitrogen-based specialized metabolites, if they will invest in fixed chemical defences or rely on metabolic plasticity and the *de novo* synthesis of specialized metabolites. Where ecological stresses vary gradually and persist over evolutionary time frames, it would be expected that there will be graduated phenotypic

variability within plant populations and among individuals within a population, in order to respond to these pervasive, graduated stresses.

Cyanogenic glycosides are a class of specialized metabolite that are synthesized from amino acids which break down to prussic acid and release hydrogen cyanide under the correct environmental or ecological conditions, or that may be triggered by herbivores (Harborne, J., 1988). Hydrogen cyanide binds to cytochrome c oxidase, thereby interfering with the electron transport chain and preventing aerobic production of ATP within cells (Jones, M., *et al.* 1984). Furthermore, ingestion of sufficient hydrogen cyanide can lead to death (Jones, D., 1972). The herbivore-deterring potential of cyanogenic glucosides in plants is undeniable (Jones, D., 1972), yet if hydrogen cyanide is liberated within plant cells, these specialized metabolites can also be autotoxic. As a result, plants that synthesize cyanogenic glucosides also spatially separate the cyanogenic glucoside molecules from the β-glucosidase enzyme, which is responsible for generating the labile cyanogenic aglycone that decomposes to liberate hydrogen cyanide (Wink, M., 1993; Jones, D., 1972). With this compartmentation, any mechanism that disrupts the spatial separation of cyanogenic glucoside and the β-glucosidase will result in hydrogen cyanide liberation, for example; mechanical disruption of cells by herbivores as well as freezing temperatures (Dayday, H., 1965).

Clover populations in Britain and Europe appear to modulate their cyanogenic phenotype in response to the competing ecological pressures of herbivory and cold climates (Jones, D., 1972). The cyanogenic phenotype is a function of two genotypes, the production of cyanogenic glucoside (*Ac*) and the accumulation of the β-glucosidase (*Li*). While rapid production of hydrogen cyanide occurs when cyanogenic glucosides

and β-glucosidase are mixed, some hydrogen cyanide liberation can occur at low rates in the absence of the β-glucosidase (Harborne, J., 1988; Jones, D., 1972). Remarkably, *Trifolium repens* L. populations, growing from sea level to 580 m above sea level, tend to be highly cyanogenic (*AcLi* genotypes), whereas those growing between 580 and 1070 m above sea level do not produce cyanogenic glucosides (*acLi* or *Acli* genotypes). Clover populations growing 1400 m, or more, above sea level make neither cyanogenic glucoside nor β-glucosidase (*acli* genotypes) (Dayday, H., 1954a).

Slugs and snails, which consume clover leaf material, do not inhabit cold climate ecosystems (Harborne, J., 1988). The graduated cyanogenic glucoside phenotypic variation that has been observed with *Trifolium repens* populations through elevation and over geographic latitude in Europe and in Britain (Dayday, H., 1954b), appears to be correlated with the prevalence of mollusc herbivores. This classic example perfectly exemplifies the type of phytochemical calculus performed by plants to balance resource allocation towards, or away, from chemical defences.

2.5 ABERRANT METABOLISM VERSES THE CHEMICAL ARMS RACE

The chemical strategies plants use to interact with their surroundings come at a cost to the organism (Gershenzon, J., 1994) and it is clear that plant populations have evolved to elegantly balance resource allocation in response to different ecological pressures (Harborne, J., 1988; Jones, D., 1972; Dayday, H., 1954a; Dayday, H., 1954b). This balance has been achieved through natural selection for individuals, within populations, that demonstrate cost balance phenotypes that are better suited for the

prevailing ecological conditions during that discrete plant generation. Therefore if specialized metabolites are synthesied at a cost, and plant populations balance resource cost allocation, then eliminating specialized metabolites and their biosynthetic pathways, which lead to products lacking biological purpose, would be expected to be favoured. This, however, does not appear to be the case in Nature.

Many of these metabolites have powerful biological activities that enhance their value as pharmaceuticals, herbicides and insecticide (Jones, C., Firn, R., 2003); however the roles of most specialized metabolites remain difficult to identify.   Large scale screening programs to identify commercially relevant biologically active plant extracts, or individual active ingredients have had remarkably low hit rates. For example; with respect to activity against cancerous cell lines, the National Cancer Institute found that only 0.07% of plants screened for anti-neoplastic activities contained compounds with high activity. Furthermore, it was estimated that only 4.3% of all plants contain small molecules with any anti-cancer potential at all (Suffness, M., Douros, J., 1982). Other programs attempting to identify plant extracts with insecticidal activity have reported similar, low hit rates (Jones, C., Firn, R., 1991). So while it remains to be established if the apparently low success rates of large scale screening programs can be used to extrapolate activity rates of specialized metabolites in Nature, the low hit rates of these screening programs have forced scientists to revisit their preconceptions about the natural function of specialized metabolites.

Chemists have argued that specialized metabolites are by-products of aberrant primary metabolism; they are not synthesized for any discrete biological function (Jones,

D., 1974) and any perceived biological function of these specialized metabolites is mere happenstance (Jones, D., 1974; Jennewein, S., *et al*. 2003).

Chemical ecologists have proposed the theory of the chemical arms race where specialized metabolites are synthesized by plants to perform important biological functions in inter- and intra-organism communication and in plant defence. The biosynthetic pathways yielding these important, functional metabolites have evolved over millions of years of selection (Després, L., *et al.* 2007; Dawkins, R., Krebs, J., 1979). The very fact that these small molecules and the biochemical pathways that generate them, are maintained in a plant population is argued as proof, in and of itself, that specialized metabolites have real biological activities.

Demonstrating function for some specialized metabolites is not proof that all specialized metabolites have relevant biological activity under natural conditions. Furthermore, the remarkably low hit rates of screening programs might be used to suggest that perhaps biologically inactive specialized metabolites are maintained in plant populations, even at an apparent resource cost disadvantage to the plant, in order to have access to a reservoir of chemical diversity that can be selected from. The apparent disparity that exists between the metabolic cost disadvantage of synthesizing inactive metabolites and the maintenance of these metabolites in plant populations is reconciled if the principles of the screening hypothesis are true (Jones, C., Firn, R., 1991).

2.6 THE SCREENING HYPOTHESIS

If potent biological activity is an inherently rare property of plant-based small molecules (Firn, R., Jones, C., 2003), then it follows that mechanisms favouring inexpensive diversification of chemical structures are important to ensure continued adaptation of plants within an ever changing environment (Jones, C., Firn, R., 1991). As a pure probabilistic argument, plants that have more diversified specialized metabolites may be selected for over individuals with more restricted chemical profiles because it is easier to generate further diversity from an already diversified background.   The mechanisms that promote diversified structure elaboration of small molecules in plants have been reviewed elsewhere in detail (Jones, C., Firn, R., 1991, 2003) and remarkably these mechanisms are quite similar to those generally used by plants to minimize resource costs in specialized metabolic pathways (Gershenzon, J., 1994); enzymes that accept multiple substrates [cytochrome P450 monoxygenases (Chapple, C., 1998; Gonzalez, F., 1990) or that generate multiple products from a single substrate [terpene synthases (Gershenzon, J., 1994)], enzymes that conduct multiple sequential reactions [polyketide synthases (Ferrer, J., *et al*. 1999), certain methyltransferases (Zubieta, C., *et al.* 2001)], and certain dioxygenases (Matsuda, J., *et al.* 1991)] significantly reduce resource costs because less genetic material, relative to that required to encode single function enzymes performing all of the above activities, still generate versatile intermediates. Furthermore, since genetic mutations are random, they are more likely to knock out linear biochemical pathways, rather than to extend them (Firn, R., Jones, C., 2003). Therefore, to minimize the potential loss of established chemical diversity it is expected that plants develop biochemical pathways that are matrices, instead of linear

biochemical pathways, whereby small molecules can be biosynthesized through many possible reaction orders (Firn, R., Jones, C., 2003). Clear examples of biochemical matrices exist for phenylpropanoid biosynthesis (Humphreys, M., Chapple, C., 2002; Dixon, R., *et al.* 2002), as well as for benzylisoquinoline alkaloid biosynthesis (Desgagne-Penix, I., Facchini, P., 2011). Finally, chemical diversification can also be achieved at minimal cost to the plant by biosynthetic coupling of specialized metabolite classes, such as seen in the formation of MIAs by coupling iridoid and tryptamine moieties within the Apocynaceae, Loganiaceae, and Rubiaceae plant families (Facchini, P., De Luca, V., 2008).

2.7 SILENT METABOLIC PATHWAYS IN PLANTS

Over a century of research to elucidate the chemistry and biosynthesis of special plant metabolites has expanded our understanding of their role(s) from aberrant by-products or strange nutrient storage molecules to potent environmental effectors. Some argue that our apparent inability to elucidate the discrete function of each molecule is due to insufficient study (Pichersky, E., *et al*. 2006). However, the remarkably low hit rates of plant small screening programs make it is difficult to argue that all specialized metabolites have individual functions, rather than the chemical diversity itself being functional.

Plant metabolomes are much larger than what would be expected based on their genetic complement (Schwab, W., 2003). Moreover, the few reports available that document the diversification of specialized molecule metabolomes through evolution

suggest that chemical diversification is itself beneficial (Becerra, J., et al. 2001; Becerra, J., et al. 2009). So, how do plants generate this amazing chemical diversity, with limited genetic resources and at an apparent metabolic cost disadvantage? A number of the biochemical properties, such as enzymes accepting multiple substrates, enzymes that generate multiple products, or enzymes that conduct multiple sequential reactions, supported by the screening hypothesis, would facilitate this chemical diversification with limited genetic material (Firn, R., Jones, C., 2003). Also, recovering special metabolites through degradation pathways would mitigate the cost of their biosynthesis (Gershenzon, J., 1994). Silent metabolic pathways, could also enhance metabolome diversification under the correct environmental, developmental, or mutational circumstances (Lewinsohn, E., Gijzen, M., 2008).

Silent metabolic pathways, as their names suggest, are biochemically inactive because one or multiple enzymes involved in the pathways have been repurposed or inactivated through mutation. Alternatively, the enzyme, or suite of enzymes involved in the pathway have an altered spatial and/or temporal expression profile relative to what would be necessary for appropriate substrate supply resulting in a latent metabolic pathway that awaits the correct circumstances for reactivation.

Weng *et al.* recently published the discovery of an entirely new class of specialized molecule from *Arabidopsis*, named arabinopyrones. These molecules are derived through the redirection of *p*-coumaryl alcohol, originally used for H-lignin biosynthesis, to produce caffealdehyde via a unique hydroxylase (Weng, J., *et al*. 2012). Caffealdehyde undergoes ring opening, mediated by an enzyme encoded by *AtLigB* and a number of presently uncharacterized biochemical steps to finally yield arabinopyrones

(Weng, J., *et al.* 2012). Remarkably, the cytochrome P540 monoxygenase responsible for this step (*CYP84A4*), is a paralogue of the ferulate-5-hydroxylase involved in sinapaldehyde and sinapyl alcohol biosynthesis. Moreover, CYP84A4's substrate specificity, as well as its spatial and temporal expression profile, correlates directly with the accumulation of arabinopyrones in *Arabidopsis*. The evolutionary events that led to the appearance of CYP84A4 in *Arabidopsis* may have also coincided with the spatial and temporal requirements for expression of the remainder of arabinopyrone pathway and led to the activation of a previously silent metabolic pathway and the accumulation of these specialized metabolites (Weng, J., *et al*. 2012).

A better known example of silent metabolic pathways involves monoterpene biosynthesis in Scotch spearmint (*Mentha x gracilis*) which accumulates monoterpenes that are derived through initial 3-hydroxylation of (-)-limonene to form (-)-trans-isopiperitenol. Screening of a γ-irradiated mutant population of Scotch spearmint plants led to the identification of accession 643, which accumulates dihydrocarvones and dihydrocarverols (Croteau, R., *et al.* 1991) characteristic of peppermint (*Mentha* x *piperita*). Biochemical analysis of this mutant revealed that (-)-limonene 6-hydroxylase which is responsible for the formation (-)-transcarveol (Croteau, R., *et al.* 1991) in the parental line had been replaced with a 3-hydroxylase responsible for the formation of (-)-transpiperitol. This reaction product is converted to peppermint menthol isomers by a (-)-transpiperitol specific cytochrome P450 epoxidase. Although this enzyme was also present in the parental line, it was not apparent since appropriate substrates to form reaction products is not available without the 3-hydroxylase activity. In summary, Spearmint contains the entire pathway for biosynthesis of peppermint monoterpenes,

except for the 3-hydroxylase that 'silenced' the accumulation of dihydrocarvones and dihydrocarverols.

While it remains to be resolved where and when tomato (*Solanum lycopersicon v. esculentum*) domestication occurred (Razdan, M., Mattoo, A., 2008) in South America, this would not have been possible without human involvement. The selection of traits such as tomato appearance (color and size), taste and resistance to environmental stress took place over hundreds of years and continued selection in modern times has modified the chemistry of tomato cultivars (Razdan, M., Mattoo, A., 2008; Butelli, E., *et al.* 2008). Present day commercial tomato cultivars (*Solanum lycoperisicum*) are the result of generations of cultivation and selection for more visually attractive and shiny tomatoes. While farmers were not knowingly altering the phytochemical profile of their tomato crops, they did so inadvertently and by progressively selecting against tomatoes that accumulate high levels of flavonoids that produce a visually dull phenotype within their tomato populations (Razdan, M., Mattoo, A., 2008; Butelli, E., *et al.* 2008). Molecular and biochemical analysis showed that all modern tomatoes are low in their flavonoid content since they do not express chalcone isomerase (CHI) but they do express the rest of the pathway for flavonoid biosynthesis. High flavonoid tomatoes could easily be produced by genetic engineering expression of CHI to reactivate the 'silent' flavonol biosynthesis pathway in commercial tomatoes (Muir, S., *et al.* 2001). This example shows once again how suppressing particular genes through selection (human in this case) has led to creation of a silent pathway that could be reactivated. Since different flavonoids, and specifically flavonols, have potent cardiovascular protective properties (Hertog, M., *et al*. 1993), genetic engineering of tomato to overexpress CHI was patented

in 2007 (Bovy, A., *et al*. 2003; US6608246). So clearly the identification of silent

metabolic pathways is also an attractive biotechnological target.

2.8 BIOPROSPECTING FOR NOVEL TOCOPHEROL-LIKE *N*-METHYLTRANSFERASES.

Many specialized metabolites accumulate to levels, in Nature, that preclude their

testing against biological targets. For example, the MIA, conolidine, only accumulates to

trace levels (0.0004 % of dry weight) in Tabernamontana species which made it very

difficult to determine if this molecule has any interesting biological activity. The

development of a successful synthesis of this MIA (Tarselli, M., *et al.* 2011), allowed the

performance of bioassays that showed its effectiveness as a potent non-opioid analgesic

with the same pain killing properties of morphine. This raises the question of how many

more trace level MIAs from plants might have important undetermined biological

activities. In addition to chemical synthesis, recent advances in biochemical engineering

(Yeast - Da Silva, N., Srikrishnan, S., 2012; *E. coli* - Yi, D., *et al.* 2000) have permitted

large scale production of valuable small molecules and intermediates by transferring

plant special metabolite pathways into microbial hosts (Polyketides - Weissman, K.,

Leadlay, P., 2005; Alkaloids, Terpenes – Facchini, P., *et al.* 2012; Glenn, W., *et al.*

2012.). These advances in pathway engineering will provide remarkable opportunities for

producing rare plant special metabolites in quantities sufficient for biological testing by

expressing candidate genes that are required for their production.

With the recent explosion of medicinal plant transcriptome sequencing we are entering an age where evolutionary relationships among plants are more important than ever to those engaged in elucidating plant biochemical pathways. If specialized metabolites are not aberrant products (Gershenzon, J., 1994) then natural selection may play a role in defining their chemistry (Pichersky, E., Gang, D., 2000) by influencing the genes responsible for their biosynthesis as illustrated by their taxonomical clustering of biochemical pathways (Wink, M., 2003). Taxonomical clustering of syringyl monolignol (S-lignin) biosynthesis pathways have been reported in angiosperm and lycophyte plant lineages because the ferulate-5-hydroxylases that commit carbon to syringyl monolignol biosynthesis are phylogenetically clustered (Weng, J., *et al*. 2008; Weng, J., *et al*. 2010).

Caution should be used when looking for homologous transcripts among plants that accumulate structurally similar molecules, as the genes encoding the biochemical events leading to these molecules may not be derived from the same ancestral progenitor. The larger the evolutionary distance between plant species the greater this risk exists. Independent evolution of biosynthetic pathways has been documented in tropane alkaloid biosynthesis (*Solanaceae* Tropanone reductase II - Nakajima, T., Hashimoto, N., 1999; *Erythroxylaceae* Methylecogonone reductase - Jirschitzka, J., *et al*. 2012), and phenylpropanoid biosynthesis (Weng, J., *et al*. 2008; Weng, J., *et al*. 2010). For example; while the tropane alkaloid biochemical pathways in *Solanaceae*, and in *Erythroxylaceae* are not fully elucidated, the tropinone reductase gene families involved in both plant families have been identified and their corresponding enzymes have been biochemically characterized. The result of these independent investigations was the determination that distinct reductase families are responsible for reducing tropanone to

pseudotropine, or methylecogonone to methylecogonine, and the reverse reactions in the

presence of appropriate cofactors, in *Solanaceae* and *Erythroxylaceae* respectively

(Jirschitzka, J., *et al*. 2012).

The Tree of Life (ToL) holds many new, completely uncharacterized gene

families involved in specialized metabolite biosynthesis (Liscombe, D., *et al.* 2010;

Jirschitzka, J., *et al*. 2012; Geu-Flores, F*., et al*. 2012). As intelligent, mindful

researchers we must engineer programs that account for our incomplete knowledge if we

hope to maximize the probability of project success. This thesis documents the

development of six deep, 454 Roche pyrosequencing transcriptomes enriched in

monoterpene indole alkaloid and iriodoid pathways, and their assemblies in to useful

databases for mining genes involved in interesting biochemical reactions.

The evolution of MIA biosynthesis may represent a good model for studying how

pathway diversification has led to the appearance of several thousand MIAs within

members of the Apocynaceae, Loganiaceae and Rubiaceae plant families. The

aspidosperma, corynantheine and iboga structural classes of MIAs are derived from the

central biochemical intermediate strictosidine (Facchini, P., De Luca, V., 2008). The

biochemical steps which yield strictosidine involve the decarboxylation of tryptophan by

tryptophan decarboxylase (De Luca, V., *et al.* 1989) whose product is then condensed

with a glucoiridoid molecule, secologanin, by strictosidine synthase (McKnight, T., *et al.*

1990) to yield strictosidine. The chemical activation of strictosidine, triggered by

strictosidine β-glucosidase (Geerlings, A., *et al.* 2000), makes it possible to produce over

130 different MIAs in plant species like *Catharanthus roseus* that has been used as a

model system for studying this pathway. The reactive aglycone then undergoes a series

of uncharacterized biochemical ring rearrangement leading to the formation of the aspidosperma (tabersonine, vindoline), corynantheine (ajmalicine) and iboga (catharanthine) backbones (Facchini, P., De Luca, V., 2008).

One of the best characterized components of MIA biosynthesis involves the six step conversion of tabersonine to vindoline that is under tight development-, environment- and cell-specific control in Catharanthus leaves (Dethier, M., De Luca, V., 1993; St. Pierre, B., *et al.* 1998; Oudin, A., *et al.* 2007; Levac, D., *et al.* 2008, De Luca, V., *et al.* 1988) The third to last step in vindoline biosynthesis involves the indole *N*-methylation of 16-methoxy-2,3-dihydrotabersonine (De Luca, V., *et al.* 1988), localized to internal membranes of chloroplast thylakoids (Dethier, M., De Luca, V., 1993). The membrane association of this enzyme has made it virtually impossible to purify this protein to homogeneity to produce peptide amino acid sequences that could then be used for molecular cloning of this gene. Recently, a forward genetic approach identified the dihydrotabersonine *N*-methyltransferase (DhtNMT) as a close homologue to biochemically characterized γ-tocopherol *C*-methyltransferases (Liscombe, D., *et al.* 2010). The recombinant enzyme, isolated from *E. coli*, showed remarkable specificity for the 2,3-dihydrotabersonine artificial substrate which is also accepted by the native *C. roseus* enzyme (Dethier, M., De Luca, V., 1993).

This thesis documents the exploitation of our transcriptome databases to identify candidate genes of an emergent class of indole *N*-methyltransferases, the γ-tocopherol-like *N*-methyltransferases (TLMT) (Liscombe, D., *et al.* 2010) and how we exploited taxonomical clustering of this emergent gene family to identify the alkaloid substrates for their functional characterizations.

Modification of antineoplastic small molecules through indole *N*-methylation has been shown to improve their cytotoxicity (Kumar, D., et al. 2009) and specificity (Kumar, D., et al. 2011). Therefore, combinatorial biosynthetic efforts that synthesize naturally occurring, commercially valuable small molecule derived through the addition of such indole *N*-methylations may access more targeted ant-cancer drugs with improved therapeutic indexes. This thesis describes how large scale transcriptome databases derived from several MIA producing plant species have been used to select candidate TLMTs for functional characterization. This study also describes novel strategies for identifying MIAs that may serve as appropriate substrates for candidate TLMTS that led to the biochemical and molecular characterization a) of picrinine-*N*-methyltransferases from *Rauvolfia serpentina* and *Vinca minor*; b) of perivine-*N*-methyltransferases from *Catharanthus roseus*; and c) ajmaline *N*-methyltransferase from *Rauvolfia serpentina*.

2.9 BIBLIOGRAPHY

Ambrose, S., (2001) Paleolithic Technology and Human Evolution, *Science*, **219**, 1748-1753

Armstrong, G., Rohrbaugh, L., Rice, E., Wender, S., (1970) The effect of nitrogen deficiency on the concentration of caffeoylquinic acids and scopolin in tobacco, *Phytochemistry*, **9**, 945-948

Bacerra, J., Noge, K., Venable, D., (2009) Macroevolutionary chemical escalation in an ancient plant-herbivore arms race, *Proceedings of the National Academy of Science*, **106**, 18062-18066

Baldwin, I., Halitschke, R., Kessler, A., Sachittko, U., (2001) Merging molecular and ecological approaches in plant-insect interaction, *Current opinion in plant biology*, **4**, 351-358

Becerra, J., Venable, D., (1990) Rapid-terpene-bath and quirt-gun defense in Bursera schlechtendalii and the counterploy of chrysomelid beetles, *Biotropica*, **22**, 320-323

Becerra, J., Venable, D., Evans, P., Bowers, W., (2001) Interactions between chemical and mechanical defenses in the plant genus Bursera and their implications for herbivores, *American Zoologist*, **41**, 865-876

Behie, S., Zolisko, P., Bidochka, M., (2012) Endophytic Insect-Parasitic Fungi Translocate Nitrogen Directly from Insects to Plants, *Science*, **336**, 1576-1577

Billings, W., Mooney, H., (1968) The ecology of arctic and alpine plants, *Biology Reviews*, **43**, 481-529

Bovy, A., Hughes, S., Muir, S., Van Tunen, A., Verhoeyen, M., De Vos, C., (2003) Methods for modulating flavonoid content in tomato by transformation with chalcone isomerase DNA, Lipton, Division of Conopco Inc., US6608246 B1

Bramble, D., Lieberman, D., (2004) Endurance running and the evolution of *Homo*, *Nature*, **432**, 345-352

Brant, J., Chapin, F., Klein, D., (1983) Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory, *Oikos*, **40**, 357-368

Burns, R., Hardy, R., (1975) Nitrogen fixation in bacteria and higher plants, E.I. Du Pont de Nemours & Co., Wilmington, DE

Butelli, E., Titta, L., Giorgio, M., Mock, H., Matros, A., Peterek, S., Schijlen, E., Hall, R., Bovy, A., Luo, J., Martin, C., (2008) Enrichment of tomato fruit with health-promoting anthocyanis by expression of select transcription factors, *Nature Biotechnology*, **26**, 1301-1308.

Chapin, F. (1989) The cost of tundra plant structures: evaluation of concepts and currencies, *American Naturalist*, **133**, 363-392.

Chapple, C., (1998) Molecular-genetic analysis of plant cytochrome P450-dependant monoxygenases, *Annual review of plant biology*, 49, 311-343

Chen, C., Baucher, M., Christensen, J., Boerjan, W., (2001) Biotechnology in trees: towards improved paper and pulping by lignin engineering, *Euphytica*, **118**, 185-195

Coley, P., Bryant, J., Chapin, F., (1985) Resource availability and plant antiherbivore defense, *Science*, **230**, 895-899

Croteau, R., Karp, F., Wagschal, K., Satterwite, D., Hyatt, D., Skotland, C., (1991) Biochemical Characterization of Spearmint Mutant That Resembles Peppermint in Monoterpene Content, *Plant Physiology*, **96**, 744-752

Croteau, R., Kutchan, T., Lewis, N., (2001) Natural Products (Secondary Metabolites), Biochemistry & Molecular Biology of Plants, B. Buchanan, W. Gruissem, R. Jones, Eds., American Society of Plant Physiologists

Da Silva, N., Srikrishnan, S., (2012) Introduction and expression of genes for metabolic engineering applications in Saccharomyces cerevisiae, *FEMS Yeast Research*, **12**, 197-214

Dawkins, R., Krebs, J., (1979) Arms Races between and within Species, *The Royal Society of London. Series B, Biological Sciences*, **205**, 489-511

Dayday, H., (1954a) Gene frequencies in wild populations of Trifolium repens L., Distribution by altitude, *Heredity*, **8**, 377-384

Dayday, H., (1954b) Gene frequencies in wild populations of Trifolium repens L., Distribution by latitude, *Heredity*, **8**, 61-78

Dayday, H., (1965) Gene frequency in wild population of *Trifolium repens* L., *Heredity*, **20**, 355-365

Del Moral, R., (1972) On the variability of chlorogenic acid concentration, *Oecologia*, **9,** 289-300

Desgagné-Penix, I. and Facchini, P. J. (2011) Benzylisoquinoline Alkaloid Biosynthesis, in Plant Metabolism and Biotechnology (eds H. Ashihara, A. Crozier and A. Komamine), John Wiley & Sons, Ltd, Chichester, UK

Despres, L., David, J., Gallet, C., (2007) The evolutionary ecology of insect resistance to plant chemicals, *Trends in Ecology and Evolution*, **22**, 298-307

Dial, K., (2003) Wing-Assisted Incline Running and the Evolution of Flight, *Science*, **229**, 402-404

Dixon, R. A., Achnine, L., Kota, P., Liu, C.-J., Reddy, M. S. S. and Wang, L., (2002), The phenylpropanoid pathway and plant defence—a genomics perspective. *Molecular Plant Pathology*, **3**, 371–390

Edward, E., Ryan, C., (1990) Interplant communication: Airborne methyl jasmonate induces synthesis of proteinase inhibitors in plant leaves, *Proceedings of the National Academy of Science*, **87**, 7713-7716

Facchini, P., (2001), ALKALOID BIOSYNTHESIS IN PLANTS: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications, *Annual Reviews in Plant Physiology and Plant Molecular Biology*, **52**, 26-66

Facchini, P., Bohlmann, J., Covello, P., De Luca, V., Mahadevan, R., Page, J., Ro, D., Sensen, C., Storms, R., Martin, V., (2012) Synthetic biosystems for the production of high-value plant metabolites, *Trends in Biotechnology*, **30**, 127-131

Facchini, P., Deluca, V., (2008) Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants, *The Plant Journal*, **54**, 763-784

Fahey, J., Zalcmann, A., Talalay, P., (2001) The chemical diversity and distribution of glucosinolates and isothiocyanates among plants, *Phytochemistry*, **56**, 5-51

Farmer, E., Ryan, C., (1991) Interplant communication: Airborne methyljasmonate induces synthesis of proteinase inhibitors in plant leaves, *Proceedings of the National Academy of Science*, **87**, 7713-7716

Ferrer, J., Jez, J., Bowman, M., Dixon, R., Noel, J., (1999) Structure of chalcone synthase and the molecular basis for plant polyketide biosynthesis, *Nature Structural Biology*, **6**, 775-784

Forterre, Y., Skothelm, J., Dumals, J., Mahadevan, L., (2005) How the venus fly trap snaps, *Nature*, **433**, 421-425

Fraenkel, G., (1969) Evaluation of our thoughts of secondary plant substances, *Entomologia Experimentalias et Applicata*, **12**, 473-486

Freeman, G., Mossadeghi, N., (1971) Water regime as a factor in determining flavor strength in vegetables, *Biochemical journal*, **124**, 61-62

Gershenzon, J., (1984), Changes in the levels of plant secondary metabolites under water and nutrient stress, Phytochemical adaptations to stress, Plenum Press, New York

Geu-Flores, F., Sherden, N., Courdavault, V., Burlat, V., Glenn, W. S., Wu, C., Nims, E., Cui, Y. and O'Connor, S. E. (2012) An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis, *Nature,.* **492**, 138-142.

Giusto, B., Bessiere, J., Gueroult, M., Lim, L., Marshall, D., Hossaert-McKey, M., Gaume, L., (2010) Flower-scent mimicry masks a deadly trap in the carnivorous plant *Napenthese rafflesiana*, *Journal of Ecology*, **98**, 845-856

Giusto, B., Grosbois, V., Fargeas, E., Marshall, D., Gaume, L., (2008) Contribution of pitcher fragrance and fluid viscosity to high prey diversity in *Napenthes* carnivorous plant from Borneo*, Journal of Biosciences*, 33:1, 121-136

Glenn, W., Runguphan, W., O'Connor, S., (2012) recent progress in the metabolic engineering of alkaloids in plant systems, *Current Opinion in Biotechnology*, In Press

Gonzalez, F., Nebert, D., (1990) Evolution of the p450 gene superfamily:: animal-plant 'warfare,' molecular drive and human genetic differences in drug oxidation, *Trends in Genetics*, **6**, 182-186

Grime, J., (1979) Plant strategies and vegetation processes, John Wiley & Sons, New York

Hall, A., Blum, U., Fites, R., (1982) Stress modification of allelopathy of *Helianthus annus* L., debris on seed germination, *American Journal of Botany*, **69**, 776-783

Harborne, J., (1988) Cyanogenic glycosides, trefoils and snails, Introduction to ecological biochemistry, Academic Press 3ed, London

Herms, D., Mattson, W., (1992) The dilemma of plants – to grow or defend, *The Quarterly Review of Biology*, **67**, 283-335

 Hertox, M., Fekens, E., Hollman, P, Katman, M., Kromhout, D., (1993) Dietary antioxidant flavonols and risk of coronary heart disease risk – the Zutphen elderly study, *Lancet*, **342**, 1007-1011

Hierro, J., Callaway, R., (2003) Allelopathy and exotic plant invasion, *Plant and Soil,* **256**, 29-39

Hofman, P., Menary, R., (1979) Variations in morphine, codeine and thebaine in the capsules of Papaver somniferum L. during maturation, *Australian Journal of Agricultural Research*, **31**, 313-326

Humphreys, M., Chapple, C., (2002) Rewriting the lignin roadmap, *Current opinion in plant biolog*y**, 5**, 224-229

Jennewein, S., Rithner, C., Williams, R., Croteau, R., (2003) Taxoid metabolism: Taxoid 14β-hydroxylase is a cytochrome P450-dependent monooxygenase, *Archives of Biochemistry and Biophysics*, **413**, 262-270

Jirschitzka, J., Schmidt, G., Reichelt, M., Schneider, B., Gershenzon, J., D'Auria, J., (2012) Plant tropane alkaloid biosynthesis evolved independently in Solanaceae and Erythroxylaceae, *Proceedings of the National Academy of Science*, **109**, 10304-10309

Jones, C., Firn, R., (1991) On the evolution of plant secondary chemical diversity, *Phil. Trans. R. Soc. Lond*. B, **333**, 273-280

Jones, C., Firn, R., (2003) Natural products –a simple model to explain chemical diversity, Natural Product Reports, **20**, 382-391

Jones, D., (1972) Cyanogenic glycosides and their function, in Phytochemical Ecology, Academic Press inc, London

Jones, M., Bickar, D., Wilson, M., Brunori, M., Colosimo, A., Sarti, P., (1984) A re-examination of the reaction of cyanide with cytochrome c oxidase, *Biochemistry Journal*, **220**, 57-66

Kappelman, J., (1996) The evolution of body mass and relative brain size in fossil hominids, *Journal of Human Evolution*, **30**, 243-276

Karban, R., Baldwin, I., Baxter, K., Laue, G., Felton, G., (2000), Communication between Plants: Induced Resistance in Wild Tobacco Plants following Clipping of Neighboring Sagebrush, *Oecologia,* **125**, 66-71

Kessler, D., Baldwin, I., (2006) Making sense of nectar scents: the effects of nectar secondary metabolites on floral visitors of *Nicotiana attenuata*, *The Plant Journal*, **49**, 840-854

Knobloch, K., Berlin, J., (1980) Influence of medium composition on secondary compounds in cell suspension cultures of *Catharanthus roseus* (L.) G. Don., *Zeitschrift fuer Naturforschung, Section c Biosciences*, **35**, 551-556

Knobloch, K., Berlin, J., (1981) Phosphate mediated regulation fo cinnamoyl putrescine biosynthesis in cell suspension cultures of Nicotiana tabacum, *Planta medica*, **42** 167-172

Kumar, D., Kumar, N., Chang, K., Gupta, R., Shah, K., (2011) Synthesis and in-vitro anticancer activity of 3,5-bis(indolyl)-1,2,4-thiadiazoles, *Bioorganic and Medicinal Chemistry Letters*, **21**, 5897-5900

Kumar, D., Sandaree, S., Johnson, E., Shah, K., (2009) An efficient synthesis and biological study of novel indolyl-1,3,4-oxadiazoles as potent anticancer agents, *Bioorganic and Medicinal Chemistry Letters*, 19, 4492-4494

Lerdau, M. (1992) Future discounts and resource allocation in plants, *Functional Ecology,* **6**, 371-375.

Lerdau, M., (1993) Formal Equivalence Among Resource Allocation Models: What is the Appropriate Currency? *Functional Ecology*, **7,** 507-508

Lewinsohn, E., Gijzen, M., (2008) Phytochemical diversity: The sounds of silent metabolism, *Plant Science*, **176,** 161-169

Liscombe,D.K., Usera,A.R. and O'Connor,S.E., (2010), Homolog of tocopherol C methyltransferases catalyzes N methylation in anticancer alkaloid biosynthesis, *Proceedings of the National Academy of Science*, **107** , 18793-18798

Loomis, W., (1932) Growth-differentiation balance vs. Carbohydrate-nitrogen ration, *Proceedings of American Horticultural Science*, **29**, 240-245

Matsuda, J., Okabe, S., Hashimoto, T., Yamada, Y., (1991) Molecular Cloning of Hyoscyamine Beta-Hydroxylase, a 2-Oxoglutarate-dependent Dioxygenase, from cultured Roots of *Hyoscyamus niger*, *Journal of Biological Chemistry*, **266**, 9460-9464

Mattson, W., (1980) Herbivory in relation to plant nitrogen content, Annual *Review in Ecological Systems*, **11**, 119-161

McCaskill, D.J. and Croteau, R. (1997) Prospects for bioengineering of isoprenoid biosynthesis. In Advances in Biochemical Engineering/Biotechnology (Berger, R.G., de Bont, J.A.M., Cheetham, P.S.J. and Croteau, R., eds). Berlin: Springer-Verlag, pp. 107–146.

Me**rbach, M., Merbach, D., Maschwitz, U., Booth, W., Fiala, B., Zizka, G.,** (2002) Mass march of termites into the deadly trap, *Nature*, **415**, 36-37

Monteith, J., Elston, J., (1983) Performance and productivity of foliage in the field. The Growth and Functioning of Leaves, eds. J. Dale & F. Milthorpe, Cambridge University Press, Cambridge

Muir, S., Collins, G., Robinson, S., Hughes, S., Bovy, A., Ric De Vos, C., van Tunen, A., Verhoeyen, M., (2001) Overexpression of petunia chalcone isomerise in tomato results in fruit containing increased levels of flavonols, *Nature* Biotechnology, **19**, 470-474

Nakajima, K., Hashimoto, T., (1999) Two tropinone reductases, that catalyze opposite stereospecific reductions in tropane alkaloids biosynthesis, are localized in plant root and different cell-specific patterns, *Plant Cell Physiology*, **40**, 1099-1107

Nelson, C., (1953) Hydrocyanic acid content of certain sorghums under irrigation as affected by nitrogen fertilizer and moisture stress, *Agronomy Journal*, **45**, 615-617

Nowacki, E., Jurzysta, M., Forski, P., Nowacka, D., Waller, G., (1976) Effect of nitrogen nutrition on alkaloid metabolism in plants, *Biochemie und Physiologie der Pflanzen*, **169**, 231-240

Nozzolillo, C., (1978) The effects of mineral nutrient deficiencies on anthocyanin pigmentation in vegetative tissues, Phytochemical Bulletin, 11

Obeso, J., (2002) The cost of reproduction in plants, *New Phytologist*, **125**, 325-3483 p. 499-518.

Pare, P., Tumlinson, J., (1999) Plant Volatiles as a Defense against Insect Herbivores, *Plant Physiology*, **121**, 325-331

Perrin, N. (1992) Optimal resource allocation and the marginal value of organs. *American Naturalist*, **139**, 1344-1369.

Perrin, N. (1993) On future discounts and economic analogy in life-history studies. *Functional Ecology* **7**, 506-507

Pichersky, E., Gang, D., (2000) Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective, *Trends in Plant Science*, **5**, 439-445

Pichersky, E., Sharkey, T., Gershenzon, J., (2006) Plant volatiles: a lack of function or a lack of knowledge? *Trends in Plant Science*, **11**

Razdan, M., Mattoo, A.,(2008) in Genetic improvements in Solenaceous crops, Volume 2: Tomato, Science Publishers, British isles

Robinson, T., (1974) Metabolism and Function of Alkaloids in Plants, *Science*, **184**, 430-435

Saslis-Lagoudakis, C., Savolainen, V., Williamson, E., Forest, F., Wagstaff, S., Baral, S., Watson, M., Pendry, C., Hawkins, J., (2012) Phylogenies reveal predictive power of traditional medicine in biprospecting, *Proceedings of the National Academy of Science*, **109**, 15835-15840

Schaefer, H., Schaefer, V., Levey, D., (2004) How plant-animal interaction signal new insights in communication, *Trends in Ecology and Evolution*, **19**, 577-584

Schwab, W., (2003) Meabolome diversity: too few genes, too many metabolites? *Phytochemistry*, **62**, 837-849

Suffness, M., Douros, J., (1982) Current status of the NCI plant and animal product program, *Journal of Natural Products*, **45**, 1-14

Taylor, R., Pearcy, R., (1976) Seasonal patterns in the CO2 exchange characteristics of understory plants from decisuous forest, *Canadian Journal of Botany*, **54**, 1094-1103

Tingey, W., Singh, S., (1980) Environmental factors influencing the magnitude and expression of resistance, Breeding plants resistant to insects, John Wiley & Sons, New York

Valkenburgh, B., (1987) Canine tooth strength in large carnivores, *Journal of Zoology*, **212**, 379-397

Verporte, R., (1998) Exploration of Natures Chemodiversity: The role of secondary metabolites in lead drug discovery, *Drug Discovery Today*, **3**, 232-238

Vetter, J., (2000) Plant cyanogenic glucosides, *Toxicon*, **38**, 11-36

Weissman, K., Leadlay, P., (2005) Combinatorial biosynthesis of reduced polyketides, *Nature Reviews Microbiology*, **3**, 925-936

Weng, J., Mo, H., Chapple, C., (2012) Assembly of an Evolutionarily new Pathway for Alpha-Pyroes Biosynthesis in Arabidopsis, *Science*, **337**, 960-964

Weng, J., Stout, J., Chapple, C., (2008) Independent origin of syringyl lingnin in vascular plants, *Proceedings of the National Academy of Science*, **105**, 7887-7892

Wink, M., (1988) Plant Breeding: The importance of plant secondary metabolites for protection against pathogens and herbivores, *Theoretical applied genetics*, **75**, 225-233

Wink, M., (1993) The plant vacuole; a multifunctional compartment, *Journal of Experimental Botany*, **44**, 231-246

Wink, M., (2003) Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective, *Phytochemistry*, **64**, 3-19

Yi, D., Ellis, H., Lee, E., Jenkins, N., Copeland, N., Court, D., (2000) An efficient recombination system for chromosome engineering in Escherichia coli, *Proceedings of the National Academy of Science*, **97**, 5978-5983

Zubieta, C., He, X., Dixon, R., Noel, J., (2001) Structure of two natural product methyltransferases reveal the basis of substrate specificity in plant O-methyltransferases, *Nature Structural Biology*, **8**, 271-279

CHAPTER 3 – EPIDERMOME ENRICHED BIOSYNTHESIS OF MONOTERPENOID INDOLE ALKALOIDS AND SECRETION TO THE PLANT SURFACE MAY BE COMMON IN THE APOCYNACEAE FAMILY OF PLANTS

AUTHORS: Sayaka Masada-Atsumi, Dylan Levac, Elizabeth Edmunds, Kyung-Hee Kim, Vincenzo De Luca

3.1 – INTRODUCTION

The monoterpenoid indole alkaloids (MIAs) make up one of the most structurally and pharmacologically diverse classes of plant secondary metabolites, including the most prominent anti-cancer drugs vinblastine from *Catharanthus roseus* and camptothecin from *Camptotheca acuminata*, the anti-hypertensive agent, reserpine and the anti-arrhythmia agent, ajmaline from *Rauwolfia serpentina*, and the anti-malarial drug quinine from *Cinchona officinalis* and related species. In spite of their pharmacological and therapeutic value, plants remain the only commercial source for these low yield compounds since their chemical complexity makes total synthesis uneconomical. Therefore, with over 2000 known structures, their pharmacological and economical importance has motivated the characterization of the MIA pathways involved, together with investigations of the specialized cells and organs participating in their biosynthesis (reviewed by Facchini, P., and De Luca, V., 2008).

MIAs originate from the assembly of the shikimate and non-mevalonate pathways that supply the indole precursor tryptamine and monoterpene-secoiridoid precursor secologanin, respectively (Figure 3-1). Tryptophan decarboxylase (TDC; EC 4.1.1.28)

converts tryptophan to tryptamine, while secologanin is derived from the plastidic 2-*C*-methyl-D-erythritol 4-phosphate (MEP) pathway via multiple steps that have partially been characterized at the molecular level. Strictosidine synthase (STR; EC 4.3.3.2) catalyzes the condensation of tryptamine with secologanin to form strictosidine, the common precursor for thousands of Corynanthe, Iboga and Aspidosperma MIAs. The metabolites typically found in *Catharanthus roseus*, *Vinca minor*, *Amsonia hubrichtii* and *Tabernamontana elegans* are representative examples of MIAs derived from strictosidine (Figure 3-1). The highly regulated expression of MIA biosynthesis in *Catharanthus roseus* by development-, environment-, organ- and cell-specific controls have been well documented (De Luca, V., 2011; Facchini P., and De Luca, V., 2008, Guirimand, G., *et al*., 2010; Guirimand, G., *et al*., 2011). All known genes involved in the MEP pathway as well as geraniol 10-hydroxylase (G10H; CYP76B6; EC 1.14.14.1) (Burlat, V., *et al*., 2004; Courdavault *et al*., 2005; Mahroug, S., *et al*., 2007; Oudin, A., *et al*., 2007) were shown to be preferentially expressed in internal phloem-associated parenchyma (IPAP) cells in *C. roseus* leaves, while secologanin synthase (SLS; CYP71A1; EC 1.3.3.9) and loganic acid methyltransferase (LAMT; EC 2.1.1.50) encoding the two terminal steps in secologanin biosynthesis, TDC, STR and strictosidine β-glucosidase (SGD; EC 3.2.1.105) were expressed exclusively in leaf epidermal cells (St-Pierre, B., *et al*., 1999; Irmler, S., *et al*., 2000; Murata, J., and De Luca, V., 2005, Murata, J., *et al.*, 2008). The leaf epidermis also preferentially expresses tabersonine 16-hydroxylase and 16-hydroxytabersonine 16-*O*-methyltransferase (Murata, J., and De Luca, V., 2005) while the terminal steps in vindoline biosynthesis are expressed in leaf mesophyll idioblasts and laticifers (St-Pierre, B., *et al*., 1999). This highly specialized nature of *C. roseus* leaf

epidermis in young leaves led to the term 'epidermome' (Murata, J., *et al.,* 2008), coined

for the apparent organization of this cell type for simultaneous biosynthesis of several

different small molecules, many of which are secreted to the leaf surface.

Most recently, the need for complex development-, environment-, organ-, and

cell-specific regulation of MIA biosynthesis was partially explained by the discovery that

catharanthine and vindoline accumulated in different locations in Catharanthus leaves

(Roepke, J., *et al*., 2010). Although the entire production of catharanthine and vindoline

occurs in young developing leaves, catharanthine accumulated in leaf wax exudates of

leaves, while vindoline is found within leaf cells of 4 separate species of Catharanthus (*C.*

roseus, *C. longifolius*, *C. ovalis* and *C. trichophyllus*). The spatial separation of these two

MIAs provided a biological explanation for the low level production of dimeric

anticancer drugs found in the plant that result in their high cost of commercial

production. The ability of catharanthine to inhibit the growth of fungal zoospores at

physiological concentrations found on the surface of *Catharanthus* leaves, as well as its

insect toxicity, provide an additional biological role for its secretion. This spatial

separation also raised the possibility that dimer formation might be triggered by

herbivory that would mix catharanthine and vindoline to form toxic vinblastine or related

dimers within the intestinal tract of the herbivore. The involvement of the nucleus in the

subcellular organization of strictosidine biosynthesis in Catharanthus was also proposed

to provide a mechanism for the release of reactive strictosidine breakdown products that

might be involved in defense against insect herbivores (Guirimand, G., *et al*., 2010;

Guirimand, G., *et al*., 2011). Other studies in *Camptotheca acuminata* used indirect

florescence detection to suggest that camptothecin accumulated in glandular trichomes on

the abaxial side of the leaf mid-rib and in epidermal cells surrounding them, as well as in spongy parenchyma idioblasts adjacent to the abaxial epidermis (Valletta, A., *et al.*, 2010).

The discovery of catharanthine on the leaf surfaces of four separate *Catharanthus* species suggests that many more plant species secrete alkaloids for defensive reasons, as well as for other functions that remain to be discovered in Nature. MIAs are found mainly in plant families of the Apocynaceae, Loganiaceae and Rubiaceae. Botanically, these three are closely related families in which the Apocynaceae and Rubiaceae have emerged from the Loganiaceae. Since a large variety of MIAs have been extensively characterized in the Apocynaceae plant family, three species from widely different geographic regions (*Vinca minor* from Eurasia; *Tabernaemontana elegans* from Africa; *Amsonia hubrichitii* from North America), were selected to test the hypothesis that they might secrete MIAs to their respective leaf surfaces. The present study demonstrates that all 3 species commonly secret various MIAs to leaf surface, and that the leaf epidermis or closely related cells are responsible for MIA biosynthesis.

3.2 – MATERIALS AND METHODS

3.2.1 – PLANT MATERIAL AND ALKALOID EXTRACTION

Catharanthus roseus, *Vinca minor*, and *Tabernaemontana elegans* plants were grown in a greenhouse under a 16/8 h day photoperiod at 30 °C. Fresh leaves and other tissues of *Amsonia hubrichtii*, *A. tabernaemontana*, *A. jonesii*, *A. ciliata* and *A. orientalis* were harvested in the Niagara Parks Botanical Garden or from the Montreal Botanical

Gardens. The surface extracts were obtained by dipping each plant tissue in 3-5 mL

chloroform at room temperature for 30 min to 1 hr. The surface stripped materials were

then air dried and dipped in 3-5 mL methanol at room temperature for 1 hr to collect

alkaloids accumulated within the internal cells. These chloroform and methanol extracts

were evaporated by vacuum centrifugation using a SPD SpeedVac (Thermo Savant). The

dry samples were resuspended in 500 mL (chloroform extracts) or 1-3 mL (methanol

extracts) of methanol and filtered through 0.22 mm PALL filter (VWR) before UPLC-

MS analysis.

3.2.2 – ALKALOID ANALYSIS

Thin-layer chromatography (TLC) was performed on Polygram Sil G/UV$_{254}$

(Macherey-Nagel and Co.) developed with ethyl acetate/methanol, 90:10 v/v. Silica gel

plates were visualized under UV light to detect MIAs or the plates were treated with CAS

(1 % ceric ammonium sulfate in 85% phosphoric acid) spray reagent to develop typical

MIA colors that could be visualized. An ACQUITY UPLC system (Waters) equipped

with a Single Quadruple Detector (SQD) mass detector and a photo diode array (PDA)

was used for alkaloid analysis. The analytes were separated using an ACQUITY UPLC

BEH C$_{18}$ column (1.0 × 50 mm i.d., 1.7mm, Waters). Ultra Performance Liquid

Chromatography (UPLC) Mass Spectrometry (MS) analysis was carried out with the

same condition as described previously (Roepke, J., *et al*., 2010). Chromatographic peaks

were identified by the diode array profiles and mass of each compound.

3.2.3 – ISOLATION AND IDENTIFICATION OF VINCADIFFORMINE FROM
AMSONIA HUBRICHTII

A chloroform extract was obtained from 768 g of young *Amsonia hubrichtii* leaves and stems by immersion in chloroform (4 L) over a 3 hr period at room temperature with periodic shaking. After the solvent was evaporated *in vacuo*, the residue was suspended in 200 mL of water:methanol (80:20), acidified to pH 2 with 10% H_2SO_4 and subsequently washed several times with ethyl acetate. The resultant aqueous phase was then basified to pH 12 with 10 N NaOH and subsequently washed several times with ethyl acetate. The ethyl acetate phase was separated from the aqueous phase and evaporated *in vacuo,* yielding 1.6 g of total alkaloids. 800 mg of the residue was subjected to silica gel column chromatography with isocratic elution of chloroform:methanol (9:1) and the first yellow colored fractions were collected. 12 mL of the colored fractions was evaporated using a SPD SpeedVac (Thermo Savant), yielding 180 mg of hydrophobic alkaloids that contained approximately 60% vincadifformine. The residue was subjected to preparative TLC (TLC Silica gel 60 F254, EMD) with hexane:ethyl acetate (8:2) as the eluent. The bands which developed at Rf 0.49 were scraped and eluted with methanol. The removal of the methanol produced approximately 10 mg of a white crystal powder. The nuclear magnetic resonance (NMR) spectrum of this powder was identical to that of a reference sample of vincadifformine (Kalaus *et al.* 1993). ^{1}H NMR spectra were measured on a Bruker Avance AV 600 Digital NMR spectrometer (Bruker) with a 14.1 Tesla Ultrashield Plus magnet. ^{1}H NMR spectra (in $CDCl_3$, 600 MHz) δ (ppm): 0.58 (3H, m; C20-CH_2CH_3), 3.2 (3H, s, $COOCH_3$), 6.8-7.3 (4H, m; aromatic H), 8.9 (1H, br s; indole NH).

3.2.4 – CRUDE PROTEIN PREPARATION AND ENZYME ASSAYS

Leaf epidermis-enriched and whole leaf extracts protein extracts were prepared as described previously (Murata, J., and De Luca, V., 2005; Levac, D., *et al.*, 2008) with minor modifications for each plant organ. The protein concentration was determined using a protein assay kit (Bio-Rad). TDC assays were performed in 200 µL reaction mixture containing buffer (100 mM Tris-HCl, pH 8.0, 14 mM 2-mercaptoethanol), crude protein extract, 1mM pyridoxal phosphate and 6.25 mM (0.06 Ci) [3-^{14}C] L-tryptophan (Moravek). After incubation at 30 °C for 1 hr, the reaction was terminated by adding 30 µL of 10 N NaOH. The products were extracted with 500 µL of ethyl acetate, then 50 µL of ethyl acetate extracts were mixed into 5 mL of SCINTISAFE cocktail (Fischer Scientific) and radioactivity was detected using a Beckman LS 6500 scintillation counter.

3.2.5 – CONSTRUCTION AND SEQUENCE ANALYSIS OF cDNA

V. minor and *T. elegans* EST libraries were established as described previously (Facchini *et. al.* 2012). Total RNAs were extracted from young *V. minor* leaves and roots and *T. elegans* leaves using Trizol (Invitrogen) according to the manufacturer's instruction. Double strand cDNA libraries were prepared with the optimized protocols and sequenced on a Roche 454 GS-FLX Titanium platform (Roche) and an Illumina Genome Analyzer IIX platform (Illumina) at McGill University and Genome Quebec Innovation Centre. After assembling and annotating using BLAST vs nr, RefSeq (plant), NCBI CDD, as well as HMMSearch against Interpro suite of protein motif databases and sequences were archived in the MAGPIE software package (http://magpie.ucalgary.ca/magpie/) at University of Calgary. Reverse transcription was

performed using AMV Reverse Transcriptase (Promega) according to the manufacturer's instruction and obtained cDNAs for following experiments.

3.2.6 – REAL-TIME SEMI-QUANTITATIVE PCR

The *TDC* and *STR* mRNA transcripts were quantified by real-time quantitative PCR using iTaq SYBR Green Supermix with ROX (Bio-Rad) according to the manufacturer's instruction. The real-time quantitative PCR was performed in triplicate with a 25 µL reaction mixture containing 150 nM of each primer, 12.5 µL iTaq SYBR Green Supermix with ROX and 1 µL (80 ng) of cDNA under the following conditions: 95 °C for 3 min, then 40 cycles of 95 °C for 10 sec, 55 °C for 15 sec and 72 °C for 30 sec. Primer concentrations, annealing temperatures, and cycle numbers were optimized for each primer pair as follows: *VmTDC*-RT-F 5′- , *VmTDC*-RT-R 5′- , *VmSTR*-RT-F 5′- , *VmSTR*-RT-R 5′- , *VmACT*-RT-F 5′- , *VmACT*-RT-R 5′- , *TeTDC*-RT-F 5′- GATTGTGATGGATTGGTTTGCTC-3′, *TeTDC*-RT-R 5′- TTGACTCGCTGGTGGTGTTC-3′, *TeSTR*-RT-F 5'- CCCTCCACCAAAGAAACAACA-3′, *TeSTR*-RT-R 5′- CAAGAACTCGGCCACAAGAAC-3′, *TeUBQ*-RT-F 5′- TCTTACTGGCAAGACCATCACA-3′, *TeUBQ*-RT-R 5′- TGCCAGCGAAAATCAACC-3′. The expression levels of each target gene were analyzed with the Bio-Rad CFX Manager Software (Bio-Rad), and normalized to the housekeeping gene using the $2^{-\Delta\Delta Ct}$ method (Livak and Schmittgen, 2001).

3.2.7 – MOLECULAR CLONING AND FUNCTIONAL CHARACTERIZATION

Full length ORFs for *TDC* and *STR* were amplified using gene specific primers as follows: *VmTDC*-ORF-F 5′- *VmTDC*-ORF-R 5′- , *VmSTR*-ORF-F 5′- , *VmSTR*-ORF-R 5′- , *TeTDC*-ORF-F 5′-GAAAATGGGCAGCATTGATT-3′ , *TeTDC*-ORF-R 5′-GCTCACTTCAGGCTTCCTTG-3′ , *TeSTR*-ORF-F 5′-CAGTCTGAACATGGCAAACTC-3′ , *TeSTR*-ORF-R 5′-CAATATGTCCGTCCATGACTCT-3'. PCR products were cloned into the pGEM-T easy vector (Promega) and were sequenced to verify their identities. Inserts were then mobilized to the pET 30b vector (Novagen) by digestion with appropriate restriction enzymes and transformed into *E. coli* (DE3) pLysS cells. Transformed cells were cultivated at 37 °C until an OD_{600} of 0.6 and incubated at 25 °C overnight. After harvesting recombinant cells and obtaining the soluble proteins, enzyme assays for TDC and STR were performed as described previously (Dethier, M., De Luca, V., 1993). The identity of strictosidine produced in the STR assay was confirmed by converting it to strictosamide (strictosidine lactam) with 10 N NaOH.

UPLC-MS analysis was carried out with the same equipment and column for alkaloid analysis. The solvent systems for enzyme assay containing A (methanol: acetonitrile: 5 mM ammonium aetate (6:14:80)) and B [methanol: acetonitrile: 5 mM ammonium aetate (25:65:10)] formed with the following linear gradient at 0.3 mL/min in between the time points: 0.2 min 1% B, 1.0 min 12% B, 2.5 min 35% B, 2.8 min 50% B, 3.2 min 35% B, 3.8 min 30% B, 4.1 min 12% B, 5.0 min 1% B. The mass spectrometer was operated with an Electrospray Ionization (ESI) ion source of positive ionization

mode. A capillary voltage of 3.0 kV, cone voltage of 30 V, cone gas flow of 2 L/hr, desolvation gas flow of 650 L/hr, desolvation temperature of 350°C, and a source temperature of 150°C were applied. Conversion of tryptophan to tryptamine was verified by the diode array profile, mass (m/z 161) and retention time (0.96 min) compared to authentic standard. Production of strictosamide from tryptamine and secologanin was verified by the diode array profile, mass (m/z 499) and monitoring the tryptamine reduction.

3.2.8 – TISSUE FIXATION AND EMBEDDING

The first pair of leaves from *V. minor* and *T. elegans* were fixed in FAA (50% ethanol, 5% acetic acid, and 5% formaldehyde), dehydrated through an ethanol and tert-butanol series and then embedded in Paraplast Xtra (Fisher Scientific) as described previously (St-Pierre, B., *et al.*, 1999; Murata, J., and De Luca, V., 2005). The embedded samples were sectioned into 10 µm thick slices using a rotary microtome (Reichert Jung) and sections were carefully spread onto slides previously treated with 2% (v/v) 3-aminopropyltriethoxysilane (AES; Sigma) in acetone, incubated for 24 hr at 40°C and stored at 4°C until use. Serial sections were deparaffinized by two incubations of 15 min each in xylene before rehydration in an ethanol gradient series up to diethylpyrocarbonate-treated water.

3.2.9 – IN SITU HYBRIDIZATION

The *in situ* RNA hybridization was performed basically as described previously (St-Pierre, B., *et al.*, 1999) with some modifications. Full-length of *VmTDC*, *TeTDC*, *VmSTR* and *TeSTR* in the pGEM-T easy vector (Promega) were used for the synthesis of

sense and antisense digoxigenin-labeled RNA probes using DIG RNA Labeling

Kit(SP6/T7) (Roche Canada) according to the manufacturer's instructions. The RNA

probes were submitted to partial alkaline hydrolysis for 20 min at 60 °C. After pre-

hybridization, hybridization of the digoxigenin-labeled RNA probes and washing, the

slides were stained with alkaline phosphatase-conjugated anti-digoxigenin antibodies

(Roche Canada). For color development, the conjugates were visualized by incubation for

3 hr (*V. minor*) and 5 hr (*T. elegans*) in 5-bromo-4-chloro-3'-indolylphosphate (BCIP)

and nitro blue tetrazolium chloride (NBT).

3.3 - RESULTS

3.3.1 – FIVE AMSONIA SPECIES ACCUMULATE THEIR ALKALOIDS EXCLUSIVELY ON THEIR LEAF SURFACE

The genus *Amsonia* (bluestars) contains 22 species, with the majority found in a

wide range of habitats throughout North America, except for *A. orientalis* (native to

Mediterranean) and *A. elliptica* (native to Eastern Asia). Like other *Apocynaceae* plants,

several MIAs have been isolated from *A. elliptica* and *A. tabernaemontana* in the past

several decades (Aimi, N., *et al.*, 1978). Leaves from *Amsonia orientalis*, *A. ciliata*, *A.

tabernaemontana*, *A. hubrichtii, A. illustris* and *A. jonesii* were harvested and dipped in

chloroform to obtain leaf surface waxes and other chemicals. When the chloroform

soluble fractions were separated by TLC and visualized with ceric ammonium sulphate

spray reagent (Figure 3-S1), *Amsonia ciliata*, *A. tabernaemontana*, *A. hubrichtii, A.

illustris* and *A. jonesii* showed identical simple patterns of two MIAs that were tentatively

identified as strictamine (blue spot, Rf 0.04) and vincadifformine (blue spot, Rf 0.65) by

UPLC MS. In contrast three spots (Rf 0.04, 0.11 and 0.20) were detected in *A. orientalis*

extracts and the major compound was tentatively identified as isoeburnamine (orange

spot, Rf 0.20) (Liu, H., *et al.,* 1991). These preliminary results clearly showed that all

Amsonia species irrespective of their geographic origin appear to secrete their MIAs onto

the leaf surface. More detailed studies were preformed with *Amsonia hubrichtii* by

collecting leaves (from 4 separate stages of growth: the youngest leaves at stage 1 to

oldest leaves at stage 4), stems and flowers that were extracted with chloroform to

harvest leaf surface MIAs. Chloroform stripped plant parts were then extracted in

methanol to harvest MIAs remaining in the leaves as described in experimental

procedures. UPLC-MS analysis identified vincadifformine (**1**), based on its mass (339

m/z) and its absorption spectrum (Figure 3-2A) as the major MIA occurring on the

surfaces of flowers, stage 1 to 4 leaves and stems, while the methanol extracts contained

virtually no MIAs (Figure 3-2B). Vincadifformine was purified and identified by [1]H-

NMR analysis as described in experimental procedures. These results suggest that the

biosynthesis of MIAs in Amsonia species from different geographic origins takes place in

or near the epidermis of different above ground plant organs and that secretion of MIAs

is the default pathway that leads to the exclusive accumulation of vincadifformine on the

surface of plant organs (Figure 3-2) together with waxes and other surface molecules

such as triterpenes (Roepke, J., *et al.*, 2010).

3.3.2 – *VINCA MINOR* ACCUMULATES PLUMERAN – AND EBURNAN – TYPE MIAs ON THE LEAF AND STEM SURFACES

Vinca minor L. (lesser periwinkle) is a popular fast-growing ground cover that is distributed throughout middle and southern Europe to the Caucasus. This herb has been known to humans for thousands of years for its properties in the treatment of circulatory disorders and to stimulate brain metabolism. More recently it has been studied extensively for its potential use to treat high blood pressure. *Vinca minor* is closely related to *Catharanthus roseus* and phytochemical studies have identified more than 50 MIAs, including vincamine and vincadifformine (Farananika, P., *et al.*, 2011) in this species. Metabolites extracted by chloroform dipping of leaf pairs 1 to 3 (Figure 3-3) and stems showed almost identical MIA metabolite profiles that included vincamine (**2**), minovincinine (**3**), lochnericine (**4**), vincadifformine (**1**), ervinceine (**1'**), ervinidinine (**5**) and vincaminoreine (**6**) as predicted by their UV absorbance and UPLC MS profiles (Figure 3-3A) (Malikov, V., and Yunusov, S., 1977). The majority (89 to over 99 %) of the plumeran- and eburnan-type MIAs of *V. minor* appeared to be secreted to the leaf surface (Figure 3-3B) and only low levels of a few MIAs were detected in the methanol extracts of chloroform stripped leaves.

3.3.3 - MIA BIOSYNTHETIC ENZYMES ARE PREFERENTIALLY EXPRESSED IN YOUNG DEVELOPING LEAVES IN V. MINOR

Since MIA biosynthesis is highly regulated during plant growth and development in *Catharanthus roseus* (Murata, J., and De Luca, V., 2005; Levac, D., *et al.*, 2008; Murata, J., *et al.*, 2008), this was investigated in *V. minor* by measuring TDC

enzyme activity together with *TDC* and *STR* gene expression profiles in leaves of different ages. Analyses of TDC enzyme activity showed that while the youngest leaves (Figure 3-3C, Leaf Pair 1) were most active, leaf pair 2 was only 22 % as active and leaf pair 3 displayed no TDC activity. A *V. minor* EST library prepared from mRNA extracted from leaf pair 1 was submitted to large scale 454 sequencing and the complete sequences of *VmTDC* and *VmSTR* were obtained by homology searches using *Catharanthus* genes as queries (Figure 3-S2 and 3-S3). Real time RT-PCR analysis showed that *VmTDC* and *VmSTR* gene expression was strictly limited to leaf pair 1 when compared to the results obtained in leaf pairs 2 and 3 (Figure 3-3D). Analysis of roots suggested that while *VmSTR* expression in roots (Figure 3-3D, RT) was similar to the levels found in leaf pair 1, expression of *VmTDC* was insignificant (Figure 3-3C, RT) and this profile was consistent with the lack of MIAs occurring in the roots. These results clearly suggest that as in *C. roseus*, MIA biosynthesis takes place in early stages of leaf development in *V. minor* and that the majority of MIAs being produced are secreted to the leaf surface as leaf expansion and maturation is completed.

3.3.4 – *TABERNAEMONTANA ELEGANS* ACCUMULATES THE MAJORITY OF MIAs ON THE LEAF SURFACE

Tabernaemontana elegans (milkwood) is a small tree that grows up to 15 m tall, originating in tropical east Africa, South Africa and Swaziland that has been used in traditional medicine for treating heart disease, cancer, pulmonary disease and tuberculosis (Neuwinger, H., 2000). Phytochemical analyses of this traditional medicinal plant have characterized at least 66 MIAs (Pratchayasakul,W., *et al*., 2008) that include a number of monomers and several dimers (Figure 3-1). In contrast to *V. minor* (Figure 3-3C) and *C.*

roseus that produce rather small leaves, *T. elegans* leaves are much larger (Figure 3-4Aa, inset). Leaf pairs 1, 2 and 3 harvested from 4 month-old *T. elegans* plants were dipped in chloroform and stripped leaves were extracted in methanol to harvest MIAs. The surface of leaf pair 1 contained vobasine (**8**), tabernaemontanine (**9**) and akuammiline-*N*-(4)-oxide (**10**) (Figure 3-4Aa, B), while leaf pairs 2 (Figure 3-4Ab) and 3 (Figure 3-4Ac) showed more complex MIA profiles containing apparicine (**7**) and deoxyvobtusine (**11**). While extracts from chloroform stripped leaves contained most of MIAs **7**, **8**, **9**, **10** and **11**, the dimeric MIAs conophylline (**12**) and conophyllidine (**14**) were found inside these leaves (Figure 3-4Ad) (Majumder, P., et al., 1974; van Beek, T., *et al*., 1984; Kam, K., *et al*., 1993). Ibogan-type alkaloids, such as tabernaemontanine (**9**), dregamine and pericyclivine, were also identified from the root surface extracts (Majumder, P., *et al*., 1974). Interestingly, more than 90% of the vobasine (**8**), tabernaemontanine (**9**), akuammiline-*N*-(4)-oxide (**10**) and deoxyvobtusine (**11**) were detected from chloroform extracts, while only 25% of the conophylline (**12**) and 1.5% of conophyllidine (**14**) were extracted from chloroform (Figure 4B). Two other unknown compounds present in chloroform (**13**) and methanol (**15**) extracts of chloroform stripped leaves remain to be tentatively identified. These results indicate that formation of some dimers in *T. elegans* may restrict secretion of the dimer to the leaf surface as already described for *C. roseus* (Roepke, J., *et al*. 2010). The expanded MIA profile of older leaves suggest that MIA biosynthesis in *T. elegans* may take place over a broader range of leaf developmental stages than those of *V. minor* (Figure 3-3C) or *C. roseus* (De Luca, V., 2011; Facchini, P., and De Luca, V., 2008).

3.3.5 - MIA BIOSYNTHESIS IS REGULATED DIFFERENTIALLY IN

***TABERNAEMONTANA ELEGANS* COMPARED TO *VINCA MINOR* AND**

CATHARANTHUS ROSEUS

Analyses of TDC enzyme activity showed that older leaves (leaf pairs 2 and 3) were quite active (Figure 3-4C) compared leaf pair 1. Unlike *V. minor* where there was a >75% loss and a complete disappearance of TDC activity in leaf pairs 2 and 3, the *Te*TDC activities gradually decreased from 21.4 to 8.9 to 3.1 pmol/mg in leaf pairs 1, 2 and 3, respectively (Figure 3-4C). In addition, *TeTDC* expression was quite stable (Figure 3-4D *TDC*) while *TeSTR* expression gradually decreased with leaf age (Figure 3-4D *STR*) in *T. elegans*. This is of particular relevance since growth, development and expansion of *T. elegans* produces leaves that reach up to 20 cm in length (Figure 4A, inset) compared to those of *V. minor* (2 cm) (Figure 3-3C, inset) or *C. roseus* (4 cm). The results suggest that MIA biosynthesis in *T. elegans* may take place over a broader range of leaf developmental stages than in *V. minor* or in *C. roseus*. A *T. elegans* EST library prepared from mRNA extracted from leaf pair 1 was submitted to large scale 454 sequencing and the complete sequences of *TeTDC* and *TeSTR* were obtained by homology searches using *Catharanthus* genes as queries (Figure 3-S2 and 3-S3). Real time PCR analysis showed that *TeTDC* and *TeSTR* gene expression (Figure 3-4D) was well correlated with the expanded TDC enzyme activity (Figure 3-4C) and MIA profiles (Figure 3-4Aa-d, B) that were observed with leaf age and confirm that MIA biosynthesis is regulated differently in *T. elegans* compared to *V. minor* or *C. roseus*.

In order to establish the enzymatic functions of *TeTDC*, *VmTDC*, *TeSTR* and *VmSTR*, their corresponding ORFs of were cloned and expressed in *E. coli*. The

sequences of cloned *TeTDC*, *VmTDC*, *TeSTR* and *VmSTR* were identical to those derived

from 454 sequencing. Recombinant TeTDC (Figure 3-5Ab) and VmTDC (Figure 3-5Ac)

enzymes converted tryptophan (Rt, 0.64 min) into tryptamine (Rt, 0.96 min) compared to

bacterial extracts expressing the empty vector (Figure 3-5Ad). Recombinant TeSTR

(Figure 5Ba) and VmSTR (Figure 5Bb) enzymes converted secologanin (Rt, 1.46 min)

and tryptamine (Rt, 0.92 min) into strictosamide (Rt, 2.33 min) compared to bacterial

extracts expressing the empty vector (Figure 3-5Bc).

3.3.6 – BIOCHEMICAL AND IN SITU HYBRIDIZATION LOCALIZED TDC AND STR GENE EXPRESSION CLOSE TO THE UPPER AND LOWER LEAF EPIDERMIS IN *T. ELEGANS* AND *V. MINOR*.

The large scale secretion of MIAs to the surfaces of *Amsonia* species, *T. elegans*

and *V. minor* strongly suggest that the MIA pathways in these species are expressed in

epidermal or associated cells, as shown previously in *C. roseus* where *in situ*

hybridization studies showed that *CrTDC* and *CrSTR* genes are preferentially expressed

in epidermal cells of young growing leaves, stems and flower buds (St-Pierre, B., *et al.*,

1999). Whole and epidermis enriched 1st leaf pairs of *V. minor* were extracted and

assayed for TDC activity (Figure 6A). The specific activity of TDC was ca 5-fold greater

within leaf epidermis-enriched extracts compared with whole leaves and these results

were corroborated when extracts were submitted to SDS-PAGE, electrophoretic transfer

to Polyvinylidene Difluoride (PVDF) membranes and immunoblot with TDC antibodies

(Figure 6A, inset).

Very young leaves (leaf pair 1) of *T. elegans* and *V. minor* were prepared for *in situ* RNA hybridization studies to localize transcripts of *VmTDC* and *VmSTR* genes. *In situ* RNA hybridization of *V. minor* longitudinal sections using *VmTDC* antisense probes suggest that palisade mesophyll cells as well as those from the upper and lower leaf epidermis preferentially express TDC (Figure 3-6B), since no hybridization signal was observed with *VmTDC* sense probes (Figure 3-6C). The *VmSTR* antisense probes hybridized with varying intensity to several leaf cell types including the palisade and spongy mesophyll as well as the upper and lower leaf epidermis (Figure 3-S4A), while sense probes did not (Figure 3-S4B). *In situ* RNA hybridization of longitudinal sections using *TeTDC* antisense probes (Figure 3-6D) localized *TDC* transcripts to the upper and lower leaf epidermis when compared to sections treated with *TeTDC* sense probes (Figure 3-6E) that produce no signals. The *VmSTR* antisense probes (Figure 3-S4C) hybridize with palisade mesophyll cells together with those of upper and lower leaf epidermis, while *TeSTR* sense probe did not (Figure 3-S4D). Together (Figures 6A-E & Figure 3-S4) these results suggest that the upper and lower epidermis of *V. minor* and *T. elegans* could be important sites of MIA biosynthesis for alkaloids that subsequently accumulate on the plant surface, but the results do not eliminate other cell types in the leaves that might be involved.

3.4 – DISCUSSION

The Apocynaceae family is one of the largest angiosperm families, being composed of over 5000 species found mainly in tropical and subtropical areas around the

world. Many species from this family grow as small or large forest trees or as climbing

vines attached to trees, shrubs or other physical structures. This adaptability combined

with their capacity to biosynthesize many different biologically active secondary

metabolites has allowed members of this family to occupy a range of habitats that has

contributed to their biodiversity and ecological success. The Apocynaceae also differ

from other families in their order (Gentianales) since they accumulate latex in specialized

cells known as laticifers, where MIAs, cardiac glycosides and various other secondary

metabolites are assumed to accumulate. The preferential expression of MIA biosynthetic

enzymes encoded by *LAMT*, *SLS*, *TDC*, *STR*, *SGD*, *T16H* and *16OMT* in *C. roseus* leaf

epidermis (De Luca, V., 2011; Facchini, P., and De Luca, V., 2008) and the exclusive

accumulation of catharanthine as an exudate on the surface of the plant (Roepke, J., et al.,

2010), has raised the possibility that specialized epidermal biosynthesis and secretion of

MIAs to the surface may be quite common among the MIA-producing members of the

Apocynaceae. Cellular specialization for MIA biosynthesis coupled with surface

secretion has also been suggested for camptothecin biosynthesis in glandular trichomes,

the surrounding epidermal cells and adjacent spongy parenchyma idioblasts (Valletta, A.,

et al., 2010). The present study has expanded this investigation to *Vinca minor* from

Eurasia, *Tabernaemontana elegans* from Africa and several Amsonia species mostly

from North America to suggest that the biosynthesis of MIAs in epidermal cells and their

secretion to plant surfaces may have evolved as a common feature in MIA-producing

members of the Apocynaceae.

Initial qualitative studies using simple TLC combined with CAS spray reagent

(Figure S1) demonstrated that the leaf surfaces of *Amsonia orientalis*, *A. ciliata*, *A.*

tabernaemontana, *A. hubrichtii, A. illustris* and *A. jonesii* accumulate the majority of the MIAs that they produce, whereas the leaves that contain laticifers exude large amounts of latex upon wounding which is virtually devoid of MIAs. The results obtained with these five species suggest that the 22 known Amsonia species probably secrete their MIAs to the plant surface. This is corroborated by the detailed and more quantitative studies that were preformed with different stages of growth and development harvested from *Amsonia hubrichtii* that its major MIA, vincadifformine, is virtually all secreted to the plant surface (Figure 3-1A, B). These studies clearly suggest that the plant epidermis of Amsonia species may be specialized for MIA biosynthesis and secretion.

Remarkably, the plant surface of *V. minor* accumulated virtually all of the MIAs produced with the profiles from leaf pairs 2 and 3 being virtually identical to the profiles shown in Figure 3A , B for leaf pair 1. Enzymatic analyses of TDC together with the expression profiles of *VmTDC* and *VmSTR* genes suggested that young leaves are most active in MIA biosynthesis (leaf pair 1, Figure 3-3C and D) and that leaf epidermal cells (Figure 3-6 and Figure 3-S4) may be primary, but not exclusive sites for expression of these pathways.

T. elegans secretes MIAs of increasing complexity on the leaf surface when comparing the profiles from leaf pairs 1 to 3 (Figure 3-4Aa,b,c,d; B). In contrast to the results obtained with *V. minor,* TeTDC enzyme activity, as well as *TeTDC* and *TeSTR* gene expression decreased gradually with leaf age (Figure 3-4C, D). *In situ* hybridization studies also suggested that leaf epidermal cells (Figure 3-6B and C) are preferred sites of MIA gene expression. These observations validate the use of simple chloroform dipping to identify MIAs that accumulate on the surface of leaves/stems and that the exclusive

accumulation catharanthine in the leaf wax exudates in four separate *Catharanthus* species (*C. roseus, C. longifolius, C. ovalis*, and *C. trichophyllus*) (Roepke, J., *et al.*, 2010) is not an isolated case.

T. elegans also accumulates the dimeric MIAs including conophylline (**12**) and conophyllidine (**14**) within the leaves (Figure 3-4Ad), while deoxyvobtusine (**11**) was only found on the leaf surface (Figure 3-4Ac). The biosynthesis of conophylline (**12**) would require the formation of 16-hydroxytabersonine and taberhanine precursors, while conophyllidine (**14**) would be derived from taberhanine and 16-hydroxylochnericine. The biosynthesis of each precursor involves various oxidations and in the case of taberhanine two separate *O*-methylations. The mechanisms of dimer formation, as well as those that allow some dimers to accumulate in the leaf or on the surface in *T. elegans,* remain to be determined. The directional transport mechanisms of MIAs are obscure, but 16-hydroxylation might be a key step for the transportation of plumeran-type alkaloids from epidermis to mesophyll cells. Together, these results provide strong evidence that members of MIA producing Apocynaceae have evolved specialized MIA biosynthesis within the leaf epidermis coupled to secretory mechanisms that allow preferential surface accumulation of MIAs.

3.5 – ACKNOWLEDGEMENTS

3.6 – ACCESSION NUMBERS

The nucleotide sequences in this paper can be found in the GenBank database under accession numbers JN644945 (*Vinca minor* TDC), JN644946 (*Tabernaemontana elegans* TDC), JN644947 (*Vinca mino*r STR), and JN644948 (*Tabernaemontana elegans* STR).
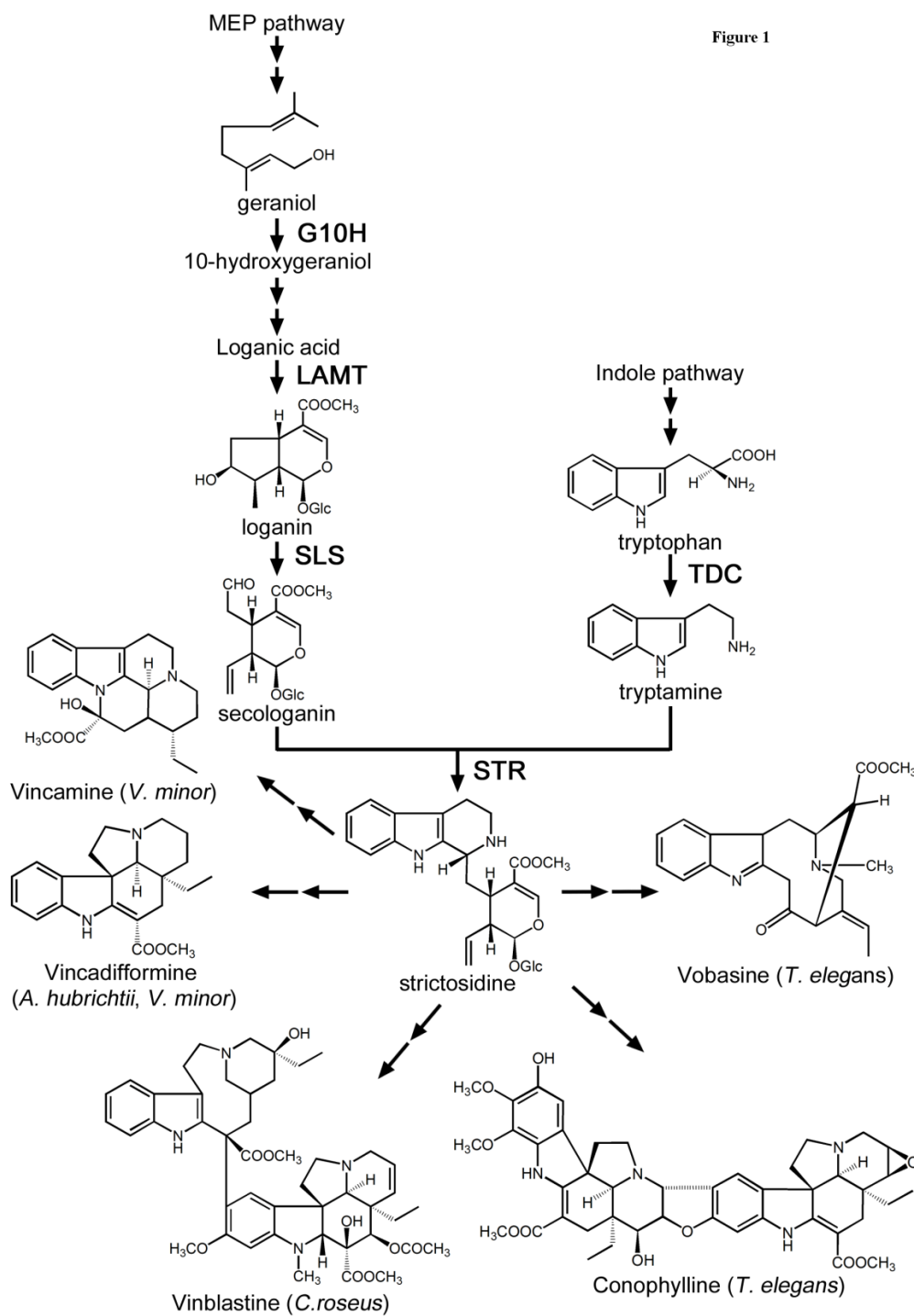
Figure 3-1: MIAs, from strictosidine, in Apocynaceae. The biosynthetic of typical *V. minor*, *A. hubrichtii*, *C. roseus* and *T. elegans* MIAs from the central strictosidine pathway. The post-strictosidine steps for each pathway have not been characterized.
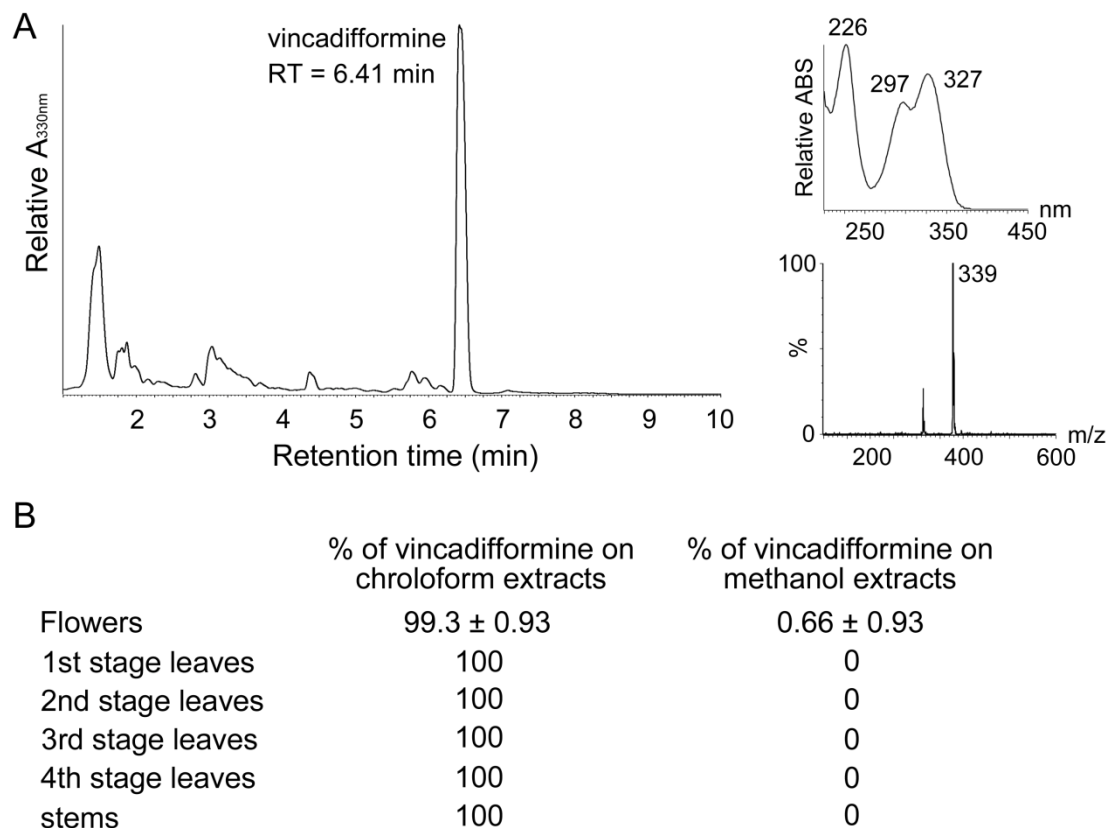
Figure 3-2: *Amsonia hubrichtii* secretes vincadifformine into leaf surface.

(A) UPLC-MS chromatogram of the chloroform extracts from 2nd stage leaves. The major peak was tentatively identified by their absorption and mass spectra as vincadifformine (RT = 6.41 min, m/z = 339) and by ^{1}H-NMR (Kalaus *et al.* 1993).

(B) Distribution of vincadifformine content in different *A. hubrichtii* organs. Chromatographic peak areas of vincadifformine on the surface of flowers, leaves of different ages and stems were measured by UPLC-MS. The experiment for each organ was performed in quintuplicate.

A

B

	% of MIA on CHCl3 extracts	% of MIA on MeOH extracts
Vincamine (**2**)	98.94	1.06
Minovincinine (**3**)	99.43	0.57
Lochnericine (**4**)	96.42	3.58
Vincadifformine (**1**) } Ervinceine (**1'**)	89.24	10.76
Ervinidinine (**5**)	98.16	1.84
Vincaminoreine (**6**)	99.27	0.73

3   R1 = R2 = H; R3 = OH
4   R1 = R2 = R3 = H; C6,7 -O-
1   R1 = R2 = R3 = H
1'   R1 = OCH3; R2 = R3 = H
5   R1 = R3 = H; R2 = O

C

D

tdc

str
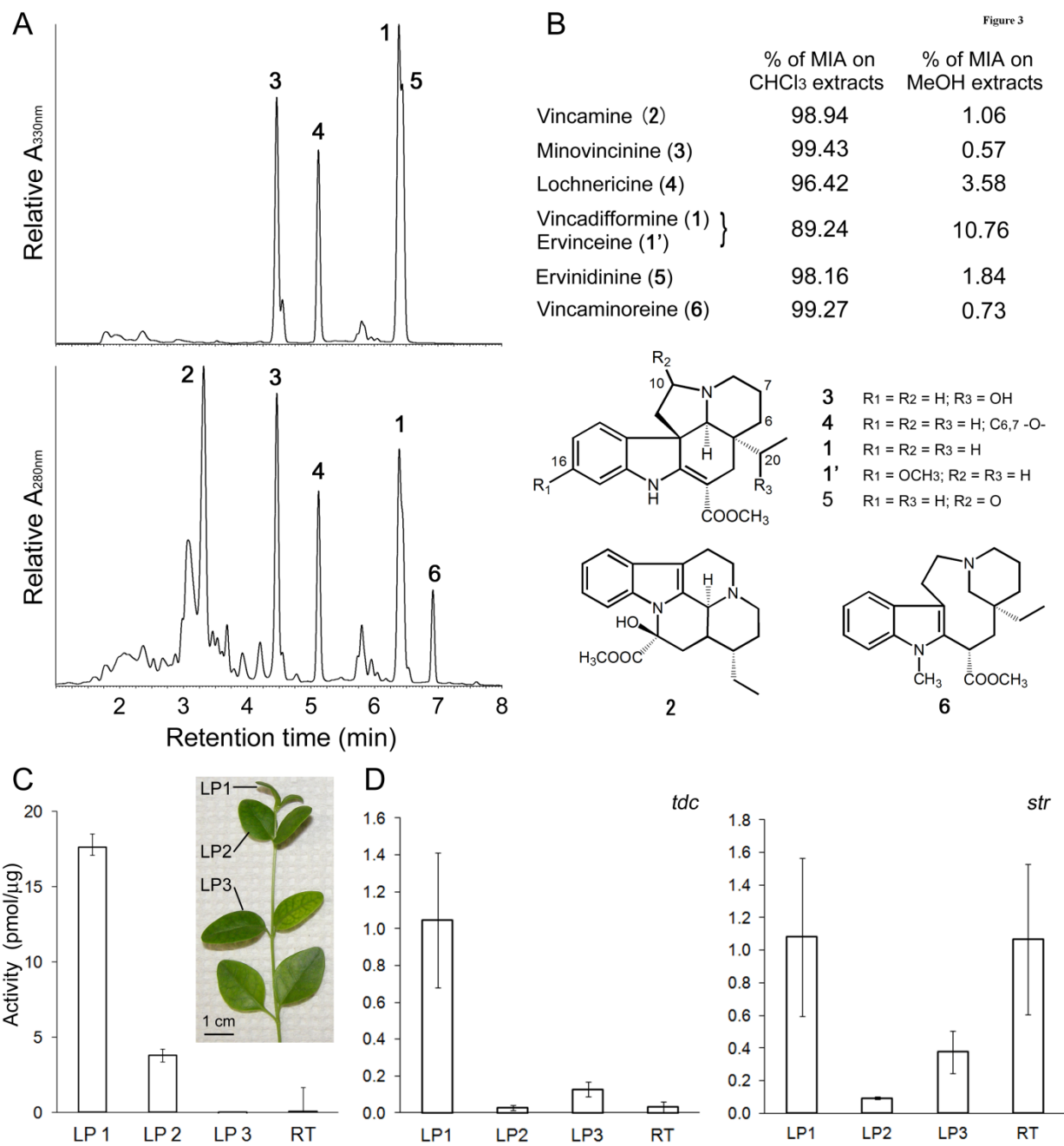
Figure 3-3: *Vinca minor* secretes most MIAs into leaf surface.

(**A**) UPLC-MS chromatogram of chloroform extracts from 1st pair of leaves at 330 nm (upper panel) and 280 nm (lower panel). The major peaks were tentatively identified by their absorption and mass spectra as vincamine (**2**, RT = 3.32 min, m/z = 355), minovincinine (**3**, RT = 4.47 min, m/z = 355), lochnericine (**4**, RT = 5.12, m/z = 353), vincadifformine (**1**, RT = 6.38 min, m/z = 339), ervinidinine (**5**, RT = 6.43, m/z = 353) and vincaminoreine (**6**, RT = 6.91, m/z = 355). (**B**) The % of each MIA found on the surface in *Vinca minor* 1st leaf pairs. This was estimated by measuring UPLC chromatographic peak areas of each MIA on the leaf surface compared to those of the methanol extract of whole leaves after stripping their surfaces with chloroform. (**C**) Distribution of TDC activity in *Vinca minor* in leaves of different ages (LP1 to LP3) and roots (RT). Assays containing crude extracts were incubated for 1 hr in the presence of [3-^{14}C] L-tryptophan and processed as described in materials and methods. Each point represents the mean of three biological replicate assays ± SD. (D) Real-time PCR quantification of *TDC* and *STR* in leaf pairs 1, 2 and 3 and in roots of *Vinca minor*. β-actin was used as a reference for the quantitative RT-PCR. Each data point represents the mean of triplicate measurements ± SD.

A CHCl₃ extracts — MeOH extracts — Figure 4

(a) LP1

(b) LP2

(c) LP3

(d) LP2

Relative A₃₁₀ₙₘ

Retention time (min)

2 cm — LP1 LP2 LP3

B

	% of metabolite in CHCl₃ extracts
Apparicine (**7**)	77.29 ± 20.28
Vobasine (**8**)	99.73 ± 0.47
Tabernaemontanine (**9**)	93.09 ± 9.20
Akuammiline N(4)-oxide (**10**)	97.16 ± 1.23
Deoxyvobtusine (**11**)	100
Conophylline (**12**)	24.78 ± 14.69
unknown (**13**)	30.39 ± 9.48
Conophyllidine (**14**)	1.46 ± 1.83
unknown (**15**)	4.41 ± 3.68

C

Activity (pmol/µg)

LP1 LP2 LP3

D

tdc

LP 1 LP2 LP3

str

LP 1 LP2 LP3

Figure 3-4: *Tabernaemontana elegans* secretes MIAs into leaf surface except for some dimers.

(A) UPLC-MS chromatogram of the extracts from leaves of different ages. Chloroform extracts of leaf pairs 1 (a), 2 (b), 3(c) and methanol extract of leaf pair 2 (d). The major peaks were tentatively identified by their known absorption and mass spectra as apparicine (**7**, RT = 3.06 min, m/z = 265), vobasine (**8**, RT = 3.71 min, m/z = 353), tabernaemontanine (**9**, RT = 4.14, m/z = 355), akuammiline N (4)-oxide (**10**, RT = 4.57 min, m/z = 411), anhydrovobtusine (**11**, RT = 5.31, m/z = 702), conophylline (**12**, RT = 5.54, m/z = 796), unknown (**13**, RT = 6.10, m/z = 837), conophyllidine (**14**, RT = 6.42, m/z = 780) and unknown (**15**, RT = 8.61 min, m/z = 793). **(B)** The % of each MIA found on the surface of 2nd *T. elegans* leaf pairs. This was estimated by measuring UPLC chromatographic peak areas of each MIA on the leaf surface compared to those of the methanol extract of whole leaves after stripping their surfaces with chloroform. The experiment represents data obtained from 3 biological replicates. **(C)** Distribution of TDC activity in *T. elegans* leaves of different ages (LP1 to LP3). Assays containing crude extracts were incubated for 10 minute in the presence of [3-^{14}C] L-tryptophan and processed as described in materials and methods. Each data point represents the mean of three separate assays ± SD. **(D)** Real-time PCR quantification of *TDC* and *STR* in leaf pairs 1, 2 and 3 of *T. elegans*. Ubiquitin was used as a reference for the quantitative RT-PCR. Each data point represents the mean of triplicate measurements ± SD.

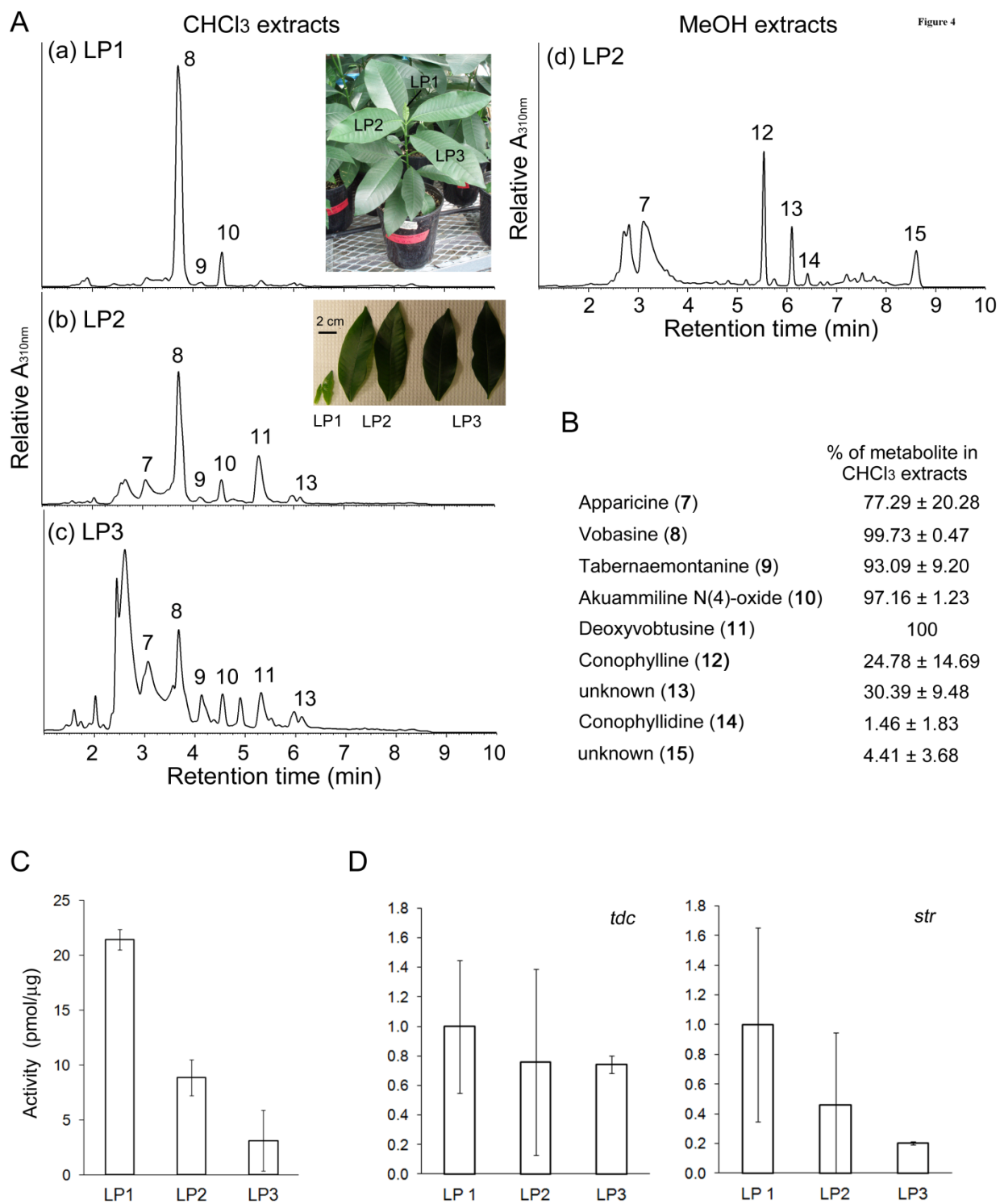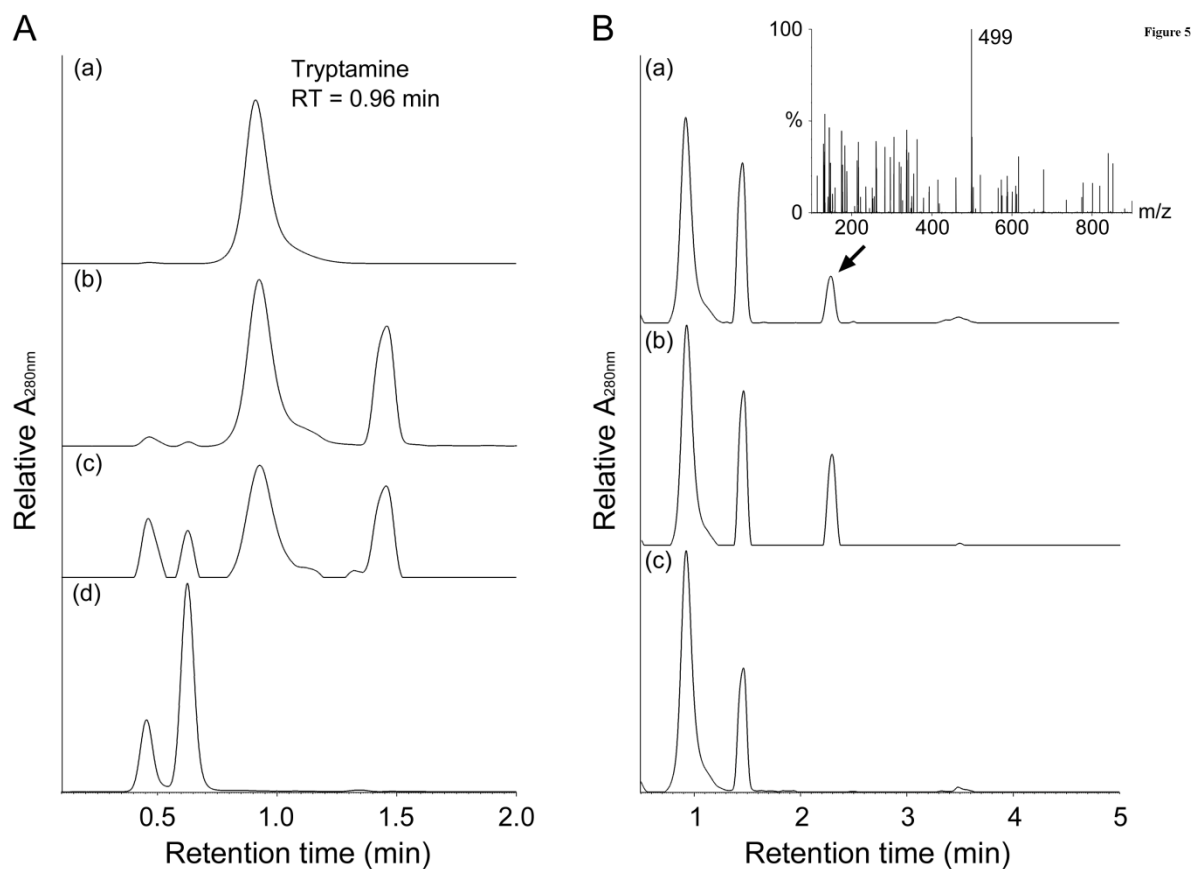**Figure 3-5:** Functional verification of *Vinca minor* and *Tabernaemontana elegans* recombinant TDCs and STRs expressed in *E. coli*.

(A) *E. coli* cell-free extracts expressing recombinant *Te*TDC (b) and *Vm*TDC (c) converted tryptophan to tryptamine (RT; 0.96 min) compared to those expressing empty vector (d). Chromatography of tryptamine standard is in panel (a)

(B) *E. coli* cell-free extracts expressing *Te*STR (a) and *Vm*STR (b) convert tryptamine and secologanin to the strictosidine by-product strictosamide (RT = 2.33 min, m/z = 499) compared with those expressing empty vector (c).

Vinca minor

A

TDC Specific activity pmol/μg protein/10 min

Whole leaf

Leaf epidermis

→ 55 Kda

30

20

10

0

Whole leaf   **Leaf epidermis**

B   **C**

←UE→
←PM→

←LE→

50 μm

Tabernamontana elegans

D   **E**

←UE→

←PM→

←LE→
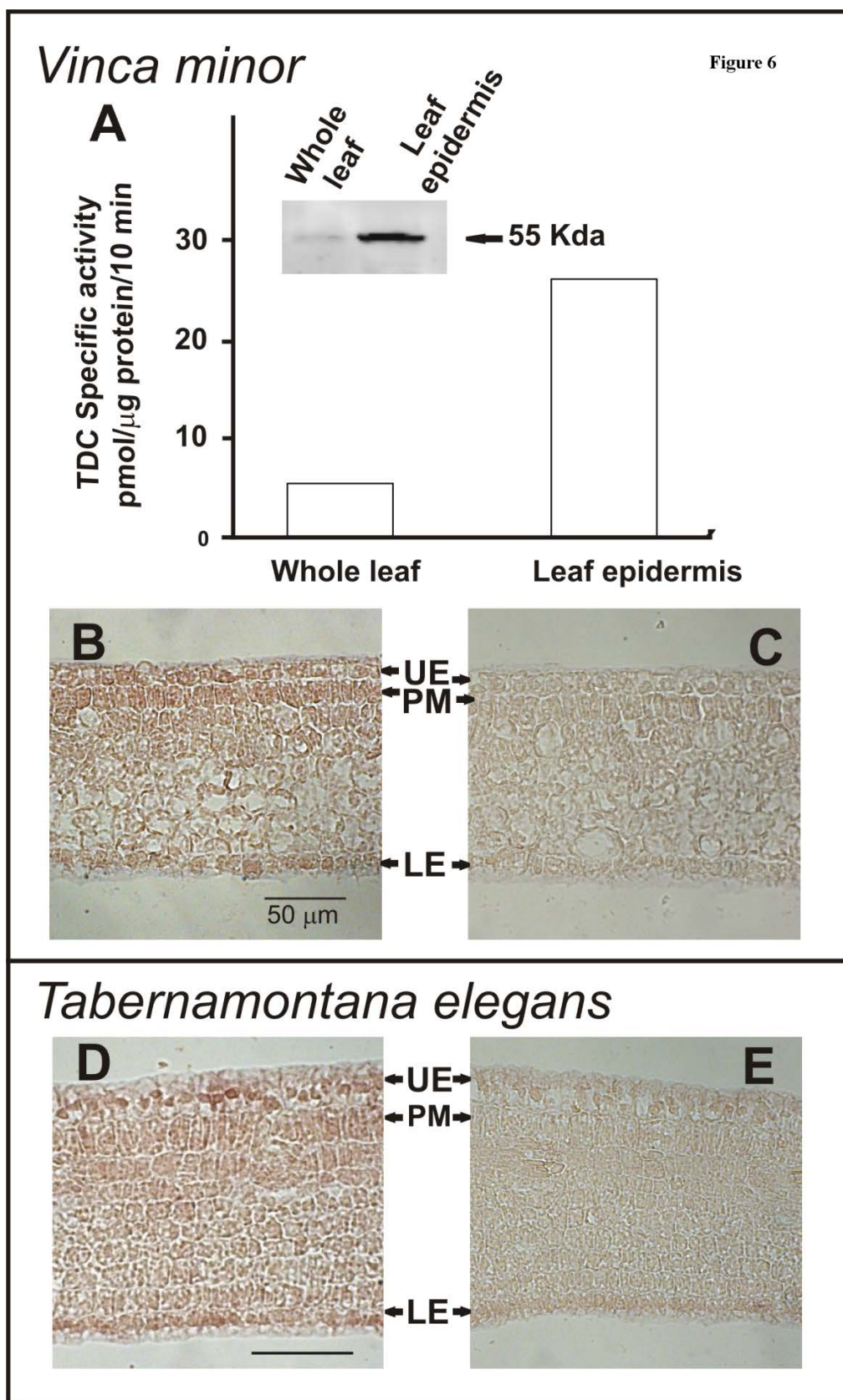
Figure 3- 6: Biochemical and in situ localization of *tdc* and *str* mRNA in young

developing leaves of *Tabernanamonta elegans* and *Vinca minor.*

A) Leaf epidermis-enriched and whole leaf extracts of *Vinca minor* were assayed for

TDC enzyme activity and for TDC antigen abundance (inset) within each preparation.

Paraffin-embedded serial longitudinal 10 µm sections were made from 2-3 cm long

leaves of *T. elegans* and 6-10 mm long leaves of *V. minor.* The slides containing *V. minor*

(B, C) and *T. elegans* (D, E) sections were hybridized with antisense (B, D) and control

sense (C, E) digoxigenin-labeled TDC transcripts.
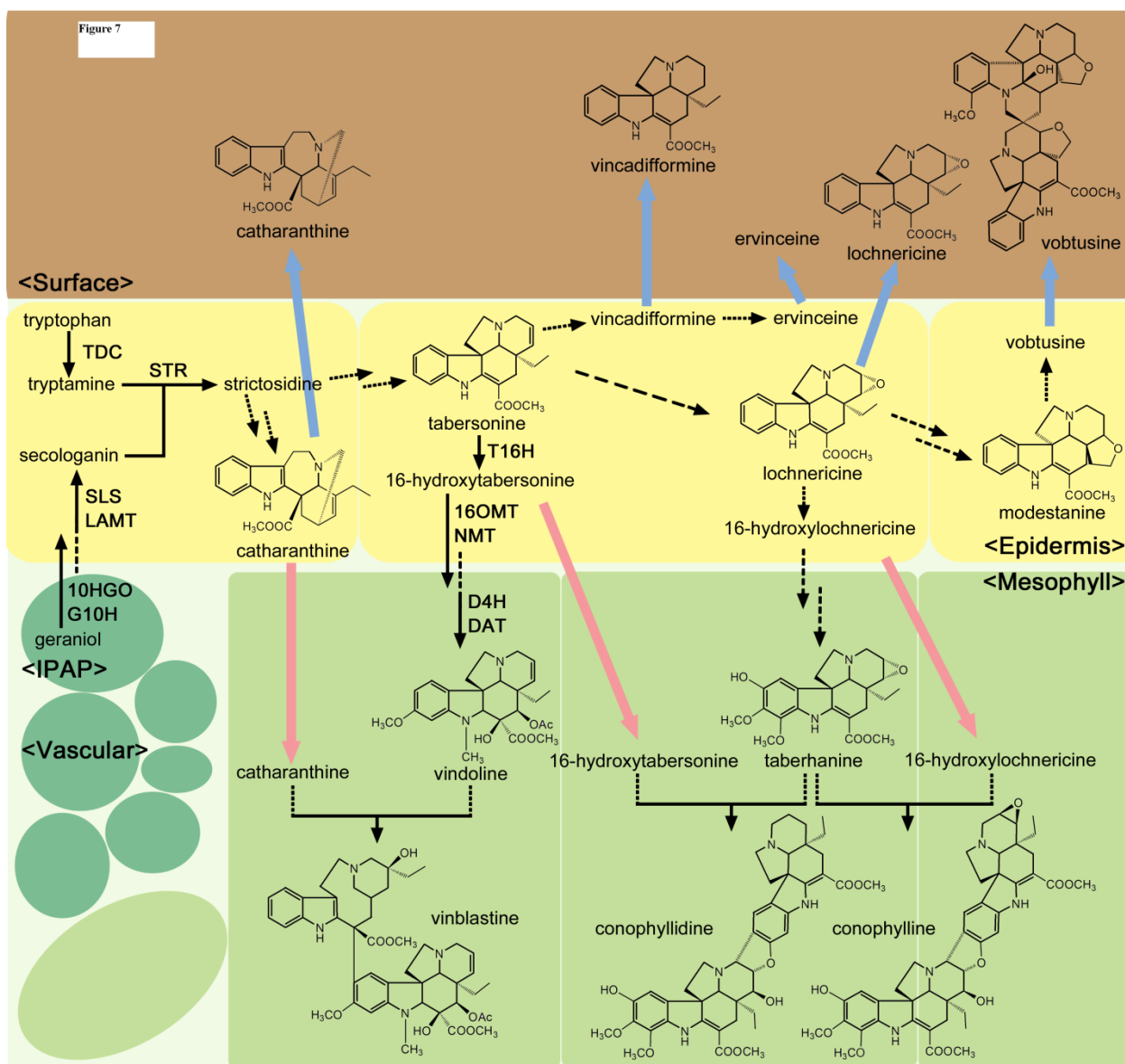
Figure 3-7: Model for the MIA biosynthesis in Apocynaceae leaves

The leaf epidermis may to be the site of MIA assembly in the Apocynaceae since several

Catharanthus and *Amsonia* species, as well as those of *V. minor*, and *T. elegan,s* secrete

and accumulate their MIAs (catharanthine, vincadifformine, ervinceine, lochnericine and

vobtusine) on the plant surface. However, some MIAs (vindoline, vinblastine,

conophyllidine and conophylline) occur mostly within the leaves and are assembled by

processes that remain to be described. In *C. roseus*, IPAP cells whose are responsible for the formation of 10-hydroxygeraniol as suggested by the preferential expression of G10H (geraniol-10-hydroxylase) and 10HGO (10-hydroxy- geraniol-oxidoreductase) in this cell type. An uncharacterized intermediate is transported from the IPAP cells to the leaf epidermis the last 2 steps (LAMT; loganic acid *O*-methyltransferase and SLS; secologanin synthase) in secologanin biosynthesis are preferentially expressed, as are TDC (tryptophan decarboxylase), STR (strictosidine synthase), T16H (tabersonine-16-hydroxylase), 16OMT (16-hydroxytabersonine- 16-*O*-methyltransferase) and possibly NMT (16-methoxy-2,3-dihydrotabersonine- N-methyl-transferase). Since the last 2 steps in vindoline biosynthesis (D4H desacetylvindoline-4-hydroxylase and DAT deacetylvindoline- 4-*O*-acetyltransferase) have been localized to specialized mesophyl idioblasts and laticifers, transport of a late vindoline pathway intermediate from the leaf epidermis must take place. The dotted arrows indicate uncharacterized reactions in different plant species. The wide arrows indicate the putative transportation of MIAs from the epidermis to the leaf surface or to leaf mesophyll.

3.7 – BIBLIOGRAPHY

Aimi N, Asada Y, Sakai S, Haginiwa J. 1978. Studies on Plants containing Indole Alkaloids. VII. Isolation of several Aspidosperma- and Vincamine-type alkaloids from the seeds of Amsonia elliptica ROEM. et Schult. *Chemical Pharmaceutical Bulletin* **26,** 1182-1187.

Aerts RJ, Gisi D, De Carolis E, De Luca V, Baumann TW. 1994. Methyl jasmonate vapor increases the developmentally controlled synthesis of alkaloid in *Catharanthus* and *Cinchona* seedlings. *Plant J*ournal **5**, 635–643.

Burlat V, Oudin A, Courtois M, Rideau M, St-Pierre B. 2004. Co-expression of three MEP pathway genes and geraniol 10-hydroxylase in internal phloem parenchyma of *Catharanthus roseus* implicates multicellular translocation of intermediates during the biosynthesis of monoterpene indole alkaloids and isoprenoid-derived primary metabolites. *Plant J*ournal **38**, 131–141.

Collu G, Garcia AA, van der Heijden R, Verpoorte R. 2002. Activity of the cytochrome P450 enzyme geraniol 10-hydroxylase and alkaloid production in plant cell cultures. *Plant Sci*ence **162,** 165–172.

Courdavault V, Burlat V, St-Pierre B, Giglioli-Guivarc'h N. 2005. Characterisation of CaaX-prenyltransferases in *Catharanthus roseus*: relationships with the expression of genes involved in the early stages of monoterpenoid biosynthetic pathway. *Plant Science* **168,** 1097–1107.

Courdavault V. 2011. Spatial organization of the vindoline biosynthetic pathway in *Catharanthus roseus*. *Journal of Plant Physiology* **168,** 549-557.

De Luca V. 2011. Monoterpenoid indole alkaloid biosynthesis. *Plant Metabolism and Biotechnology* 263-292. (eds. Ashihara, Crozier, Komamine).

De Luca V. 1993. Indole alkaloid biosynthesis. *Methods in Plant Biochemistry*. *Enzymes of Secondary Metabolism* **9,** 345-368 (Lea, P., ed. London: Academic Press).

De Luca V, Balsevich J, Tyler RT, Eilert U, Panchuk BD, Kurz WGW. 1986. Biosynthesis of indole alkaloids: developmental regulation of the biosynthetic pathway from tabersonine to vindoline in *Catharanthus roseus*. *Journal of Plant Physiol*ogy **125,** 417-156.

Endress ME, van der Ham RWJM, Nilsson S, Civeyrel L, Chase MW, Sennblad B, Potgieter K, Joseph J, Powell M, Lorence D, Zimmerman YM, Albert VA. 2007. A Phylogenetic Analysis of Alyxieae (Apocynaceae) Based on rbcL, matK, trnL Intron, trnL-F Spacer Sequences, and Morphological Characters. *Annals of the Missouri Botanical Garden*. **94,** 1-34.

Facchini PJ, De Luca V. 2008. Opium poppy and Madagascar periwinkle: Model non-model systems to investigate alkaloid biosynthesis in plants. *Plant J*ournal **54,** 763–784.

Facchini PJ, Bohlmann J., Covello PS, De Luca V, Mahadevan R, Page JE, Ro DK, Sensen CW, Storms R, Martin VJ (2012) Synthetic biosystems for the production of high-value plant metabolites. *Trends in Biotechnoogy* **30,** 127-131.

Farananika B, Akbarzadeh T, Jahangirzadeh A, Yassa N, Ardekani MRS, Mirnezami A, Khanavi M. 2011. Phytochemical investigation of *Vinca minor* cultivated in Iran. *Iranian Journal of Pharmaceutical Research* **10,** 777-785.

Guirimand G, Courdavault V, Lanoue A, Mahroug S, Guihur A, Blanc N, Giglioli-Guivarc'h N, St-Pierre B, Burlat V. 2010. Strictosidine activation in Apocynaceae: towards a 'nuclear time bomb'? *BMC Plant Biol*ogy **10,** 182.

Guirimand G, Guihur A, Ginis O, Poutrain P, Héricourt F, Oudin A, Lanoue A, St-Pierre B, Burlat V, Courdavault V. 2011. The subcellular organization of strictosidine biosynthesis in *Catharanthus roseus* epidermis highlights several trans-tonoplast translocations of intermediate metabolites. *FEBS Journal,* **278,** 749-763.

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

Kalaus G, Greiner I, Kajatar-Peredy M, Brlik J, Szabo L, Szantay C. 1993. Synthesis of Vinca Alkaloids and Related Compounds. A new synthetic pathway for preparing alkaloids and related compounds with Aspidosperma skeleton. Total syntheses of (+)-vincadifformine, (+)-tabersonine, and (+)-3-oxotabersonine, *Journal of Organic Chemistry* **58,** 1434-1442.

Kam TS, Loh KY, Wei C. 1993. Conophylline and conophyllidine: new dimeric alkaloids from *Tabernaemontana divaricata. Journal of Natural Products* **56,** 1865-1871.

Kutney JP, Chan KK, Failli A, Fromson JM, Gletsos C, Nelson VR. 1968. Total synthesis of some monomeric vinca alkaloids: dl-vincadine, dlvincaminoreine, dl-vincaminorine, dl-vincadifformine, dl-minovine, and dl-vincaminoridine. *Journal of the American Chemical Society* **90,** 3891-3893.

Levac D, Murata J, Kim WS, De Luca V. 2008. Application of carborundum abrasion for investigating the leaf epidermis: Molecular cloning of *Catharanthus roseus* 16-hydroxytabersonine-16-*O*-methyltransferase. *Plant J*ournal **53,** 225–236.

Liscombe DK, Usera AR, O'Connor SE. 2010. Homolog of tocopherol *C* methyltransferases catalyzes *N* methylation in anticancer alkaloid biosynthesis. *Proceedings of the National Academy of Sciences USA* **107,** 18793-18798.

Liu HM, Wu B, Zheng QT, Feng XZ. (1991) New Indole Alkaloids from *Amsonia sinensis. Planta Med*ica **57,** 566-568.

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

Mahroug S, Burlat V, St-Pierre B. 2007. Cellular and sub-cellular organisation of the monoterpenoid indole alkaloid pathway in *Catharanthus roseus*. *Phytochemical Review* **6,** 363–381.

Majumder PL, Chanda TK, Dinda BN. 1974. Alkaloids of the leaves of Voacanga grandifolia. *Phytochemistry* **13,** 1261-1264.

Malkov VM, Yunusov SY. 1977. Vinca alkaloids. *Chemistry of Natural Compounds* **13,** 497-512.

Mansoor TA, Ramalho RM, Mulhovo S, Rodrigues CMP, Ferreira MJU 2009. Induction of apoptosis in HuH-7 cancer cells by monoterpene and ☐-carboline indole alkaloids isolated from the leaves of *Tabernaemontana elegans*. *Bioorganic and Medicinal Chemistry Lett*ers, **19,** 4255–4258.

Murata J, De Luca V. 2005. Localization of tabersonine 16-hydroxylase and 16-OH tabersonine-16-*O*-methyltransferase to leaf epidermal cells defines them as a major site of precursor biosynthesis in the vindoline pathway in *Catharanthus roseus*. *Plant J*ournal **44,** 581–594.

Murata J, Roepke J, Gordon H, De Luca V. 2008. The leaf epidermome of *Catharanthus roseus* reveals its biochemical specialization. *Plant Cell* **20,** 524–542.

Neuwinger HD. 2000. African traditional medicine: a dictionary of plant use and applications. Stuttgart: Medpharm Scientific, pp.58

Oudin A, Mahroug S, Courdavault V, Hervouet N, Zelwer C, Rodríguez-Concepción M, St-Pierre B, Burlat V. 2007. Spatial distribution and hormonal regulation of gene products from methyl erythritol phosphate and monoterpene-secoiridoid pathways in *Catharanthus roseus*. *Plant Molecular Biology* **65,** 13–30.

Pratchayasakul W, Pongchaidecha A, Chattipakorn N, Chattipakorn S. 2008. Ethnobotany and ethnopharmacology of *Tabernaemontana divaricata*. *Indian Journal of Medicinal Research* **127,** 317-335.

Rakhimov DA, Malikov VM, Yagudaev M.R, Yunusov SY 1970. Structure of ervinceine, ervamicine, and ervincinine. *Khimiya Prirodnykh Soedinenii* **6,** 226-231.

Roepke J, Salim V, Wu M, Thamm AMK, Murata J, Ploss K, Boland W, De Luca V. 2010. Vinca drug components accumulate exclusively in leaf exudates of Madagascar periwinkle. *Proceedings of the National Academy of Sciences, USA* **107,** 15287-15292.

St-Pierre B, Vazquez-Flota FA, De Luca V. 1999. Multicellular compartmentation of *Catharanthus roseus* alkaloid biosynthesis predicts intercellular translocation of a pathway intermediate. *Plant Cell* **11,** 887-900.

Valletta A, Trainotti L, Santamaria AR, Pasqua G. 2010. Cell-specific expression of tryptophan decarboxylase and 10-hydroxygeraniol oxidoreductase, key genes involved in camptothecin biosynthesis in *Camptotheca acuminata* Decne (Nyssaceae). *BMC Plant Biology* **10,** 69-76.

van Beek TA, van Gessel MAJT. 1988. Alkaloids of Tabernaemontana species. In *Alkaloids: Chemical and biological perspectives* **6,** 75–226 (Pelletier, S.W. ed. New York: Wiley).

van Beek TA, Verpoorte R, Baerheim Svendsen A, Leeuwenberg AJM, Bisset NG. 1984. Tabernaemontana L. (Apocynaceae): a review of its taxonomy, phytochemistry, ethnobotany and pharmacology. *Journal of Ethnopharmacology* **10,** 1–156.

van der Heijden R, **Louwe CL**, **Verhey ER**, **Harkes PA, Verpoorte R**. 1989. Characterization of a Suspension Culture of *Tabernaemontana elegans* on Growth, Nutrient Uptake, and Accumulation of Indole Alkaloids. *Plant Medica* **55,** 158-162.

van der Heijden R, **Brouwer RL**, **Verpoorte R, Wijnsma R, van Beek TA, Haekes PAA, Baerheim Svendsen A.** 1986. Indole alkaloids from a callus culture of *Tabernaemontana elegans. Phytochemistry* **25,** 843-846.

van der Heijden R, **Brouwer RL**, **Verpoorte R, van Beek TA, Haekes PAA, and Baerheim Svendsen A.** 1986. Indole alkaloids from *Tabernaemontana elegans. Planta Medica* **52,** 144-147.

CHAPTER 4 – TAXONOMICAL CLUSTERING OF *N*-METHYLATED MONOTERPENOID INDOLE ALKALOIDS IN APOCYNACEAE

AUTHORS: Dylan Levac, Vincenzo De Luca

4.1 – INTRODUCTION

In the last century, most drugs approved by the U.S. Food and Drug Administration (FDA) were from natural sources or were derived from compounds first isolated from Nature. However, in recent decades the proportion of approved drugs discovered from Nature has plummeted to just below 50 percent (Li, J., Vederas, J., 2009). During this period the pharmaceutical industry shifted research efforts from screening for drug activities from natural sources to the screening of synthetic libraries against defined biological targets (Paul, S., *et al*. 2010). With the development of high throughput screening (HTS) methodologies, it was thought that natural extracts were incompatible with HTS, as they are highly complex chemical mixtures composed of many structurally related molecules, at times present in only trace amounts, and therefore more HTS compatible samples should be pursued for hit identification (Broach, J., Thorner, J., 1996). Chemists, on the other hand, could produce massive libraries of defined chemical composition and quantity, whose components could be systematically decorated, and were perfectly suited for HTS. While combinatorial libraries were less structurally diverse, and complex natural extracts (Feher, M., Schmidt, M., 2003), HTS would guarantee and enhance the rates of drug discovery (Paul, S., *et al*. 2010).

Unfortunately, the speed of HTS has not been able to complement the quality of the combinatorial libraries with only 0.4 chiral centers per molecule that were being screened. In comparison, FDA approved drugs have, on average, 2.3 chiral centers per molecule and naturally derived molecules have, on average, 6.2 chiral centers per molecule (Feher, M., Schmidt, M., 2003). Since the shift away from screening naturally derived drugs, the approval rate of new molecular entities (NMEs) has dropped dramatically (1996 to 2011), bottoming out at only 17 NMEs in 2002 (FDA, 2013). The failure of combinatorial chemistry is perfectly illustrated by the fact that only one NME has ever been brought to market through this type of screening when the protein kinase inhibitor, Nexavar was approved for the treatment of kidney and liver cancer (Mewmann, D., Cragg, G., 2007).

In an attempt to recapture the chemical complexity of natural products the pharmaceutical industry has been exploring "diversity oriented syntheses" (Schreiber, S., 2000; Tan, D., 2005). While an improvement, the chemical complexity of these newer synthetic libraries is still nowhere near that of Nature. As a result, there is renewed interest in bioprospecting and the development of natural products as pharmaceuticals (Li, J., Vederas, J., 2009; De Luca, V., *et al*. 2012; Zhu, F., *et al.* 2011; Saslis-Lagoudakis, C., *et al.* 2012; Malik, N., 2008). Shareholders, who have grown accustomed to enormous growth rates and profits during the 'golden age' of pharma, would reactive negatively to a perceived shift back into the pharmaceutical 'stone age' (Malik, M., 2008); however, modern bioprospecting is anything but primitive. It is no longer necessary for ethnobotanists to drudge through tropical rain forests to identify plants containing interesting, biologically active compounds. Saslis-Lagoudakis *et al.*

have shown that traditionally used medicinal flora from South Africa, New Zealand, and Nepal assemble into phylogenetic clusters described as 'hot nodes' (Saslis-Langoudakis, C., *et al*. 2012). Systematic comparative analyses using organisms from which ca ~1300 FDA approved and clinical trial drugs are derived, clustered in only 144 organism families; a hit rate of approximately 9 drugs per family, where a random distribution of drug accumulation through the tree-of-life would predict only one drug per five families (Zhu, F., *et al.* 2011). Clearly, privileged structures accumulate in evolutionarily closely related species, and therefore combined large scale sequencing with phylogenetic analysis of taxa may be useful to identify pharmaceutically useful but yet unexplored organisms. While much more systematic then past ethnobotanical efforts, this chemical and phylogenetic approach still succumbs to the reality that many, potentially valuable natural products accumulate to levels that preclude their testing against biological targets.

Combinatorial biosynthesis, first employed to engineer polyketide synthases to make over 200 new polyketides (Weissman, K., Leadlay,P., 2005), are now being used to engineer benzylisoquinoline, monoterpene indole alkaloid (MIA) and terpene biosynthesis in modified yeast strains (Facchini, P., *et al*. 2012; Xiao, M., *et al*. 2013). This platform shows remarkable potential to generate high levels of naturally occurring compounds, with appropriate stereocenters, that are otherwise inaccessible to drug development pipelines due to their low natural abundance (Glenn, W., *et al*.2012). For example; conolidine, a MIA from *Tabernaemontana diverticata*, accumulates to only 0.00014% yield in stem bark and was identified as a potent non-opioid analgesic only after an efficient *de novo* chemical synthesis was developed (Tarselli, *et al*. 2011). As the polymerase chain reaction revolutionized the amplification of low abundance nucleic

acids from natural sources, combinatorial biosynthesis may provide the solution for producing sufficient amounts of low abundance natural products suitable for drug testing. What combinatorial biosynthetic initiatives need are properly annotated, repositories of functionally characterized enzymes and their corresponding genes.

Since the development of next-generation DNA sequencing technologies the cost of sequencing genomes and transcriptomes has been dramatically reduced (Shoner, A., *et al.* 2011). In order to realize the potential of combinatorial biosynthesis, deep transcriptome sequencing of phylogenetically distant species should be performed in order to maximize the number of genes that will capture the biodiversity of selected classes of molecules. This approach has yielded large publicly available and searchable transcriptome databases for many different medicinal plant species in Canada (PhytoMetaSyn, http://www.phytometasyn.ca/; Facchini, P., *et al*. 2012; Xiao, M., *et al*. 2013), and in the USA (Medicinal Plant Genomics Resource, http://medicinalplantgenomics.msu.edu/; MedPlants, http://medplants.ncgr.org/). Using these resources to perform phylogenetic clustering (Zhu, F., *et al*. 2011; Saslis-Lagoudakis, C., *et al.* 2012, 10) makes it possible to target and identify 'hot nodes' of genes and gene products responsible for unique biochemical reactions and pathways. This will speed up the cloning and functional characterization of genes required for producing a combinatorial biosynthesis pipeline of testable drug candidates.

This study uses RNA-seq to generate deep transcriptome assemblies from plant tissues of five members of the Apocynaceae and one member of the Caprifoliaceae families, specialized for the biosynthesis of medicinally important MIAs and iridoids. The more than 40 publically available transcriptomes generated by the PhytoMetaSyn,

Medicinal Plant Genomics, and MedPlants consortia have been probed for molecular and biochemical events that control the taxonomical clustering of pharmaceutically important indole *N*-methylated alkaloids. We are interested in this specific modification because *N*-methylation of indole nitrogens has been shown to dramatically improve the cytotoxicity (Kumar, D., *et al*. 2009), and specificity (Kumar, D., *et al*. 2011) of antineoplastic drugs.

The conversion of tabersonine to vindoline in *Catharanthus roseus* involves a key *N*-methylation (Dethier, M., De Luca V., 1993) catalysed by an indole *N*-methyltransferase (NMT) belonging to a newly discovered class of γ-tocopherol-like-methyltransferase (TLMT) (Liscombe, D., *et al*. 2010) that may be taxonomically limited to the Vinceae tribe of Apocynaceae. This limited, phylogenetic distribution of TLMTs provides a compelling biochemical argument for why the biosynthesis of indole *N*-methylated medicinal products like vinblastine, an anti-cancer drug isolated from *C.roseus* leaves, and ajmaline, an antiarrhythmic drug isolated from *R. serpentina* roots, appear to be clustered in the Apocynaceae. This study also describes the molecular identification, functional expression and developmental characterization of two new TLMTs; *CrPeNMT* from *C. roseus* that *N*-methylates perivine, and RsAjNMT, from *R. serpentina* that *N*-methylates ajmaline.

4.2 – MATERIALS AND METHODS

4.2.1 – PLANT MATERIAL

Catharanthus roseus (cv Little Delicata), *Vinca minor*, *Rauvolfia serpentina*, *Tabernaemontana elegans*, *Amsonia hubrichtii* (*Amsonia tabernaemontana*) and

Lonicera japonica were grown in the greenhouse under a long-day (16/8 h) photoperiod at $30^{o}C$. Young 1st leaf pairs (< 1.5 cm in length), subsequent leaf pairs, flowers and whole roots, respectively were harvested for total RNA isolation intended for next generation 454 Roche pyrosequencing.   Other tissues were also used for other experiments.

4.2.2 – ALKALOID ISOLATION FROM LEAF EXUDATES

Freshly harvested *Catharanthus roseus* leaves were placed in Erlenmeyer flasks with sufficient chloroform to completely submerge them in solvent to havest surface extracts over 2 hours at room temperature with shaking (100 RPM). Chloroform was then decanted into a round bottom flask and extracts were taken to dryness by flash evaporation (Buchi Rotavapor R-205, Heating Bath B-490, Brinkmann WLK 230 LAUDA, Thermosavant GP 110 GelPump). The residue was dissolved in methanol and water was added (85:15 v/v water:methanol). This aqueous extract was acidified to pH 3 with 10% H_2SO_4 and ethylacetate was added to extract contaminating small molecules to organic. The aqueous phase was kept, basified to pH 12 using 10N NaOH, and ethylacetate was added to extract alkaloids to organic. Ethylacetate was transferred to a round bottom flask and taken to dryness by flash evaporation. Alkaloid surface extracts were dissolved in methanol and stored at $-20^{o}C$ until use.

4.2.3 – TOTAL RNA ISOLATION

Freshly harvested plant tissues were homogenized in a mortar and pestle with liquid nitrogen to produce a fine powder and 200 mg were extracted in 1.5 mL microcentrifuge tubes containing 0.225 mL extraction buffer (0.1 M NaCl, 0.01 M Tris-

Cl pH 7.5, 1 mM EDTA, 1% SDS) and 0.15 mL phenol:chloroform:isoamyl alcohol

(25:24:1) pH 6.6. Samples were then shaken vigorously for 15 minutes at room

temperature, and centrifuged (9500 g, 10 minutes). The aqueous phase was transferred to

fresh 1.5 mL tubes and mixed with 0.1 volume 3 mM NaOAc pH 5.1 after which 2

volumes cold 95% ethanol was added. After incubation for 20 min at -80 ^{o}C and

precipitation of nucleic acids by centrifugation (10000 xg, 15 minutes, 4 ^{o}C), they were

washed with 70% ethanol to remove salts and dried in a SAVANT DNA120 centrifuge

(Thermo ElectroCompany). Pellets dissolved in 0.5 mL sterile DEPC treated water were

mixed by gentle shaking with 1 volume 4 M cold LiCl, incubated on ice for 3 hours and

nucleic acids were pelleted by centrifugation (10000 xg, 10 minutes, 4 ^{o}C). Pellets were

then dissolved in sterile water, then immediately treated with DNase for 30 minutes at 37

^{o}C to remove genomic DNA and purified total RNA was mixed with 0.1 volume 3 mM

NaOAc pH 5.1, and 4 volumes 100% ethanol, incubated for 20 min at -80 C and

precipitated by centrifugation (10000 xg, 15 minutes, 4 ^{o}C). After washing the pellet 3

times with 75% ethanol to remove salts it was stored in 100% ethanol at -80 ^{o}C until

needed.

4.2.4 – NEXT-GENERATION 454 PYROSEQUENCING

Databases enriched in sequences encoding genes involved in MIA biosynthesis

were generated by extracting RNA from highly active tissues (*C.roseus* young leaves,

R.serpentina whole root, and *V.minor* 1st leaf pair) and by subjecting them to

pyrosequencing on a Genome Sequencer FLX (454 Life Sciences Corp., Bradford, CT) at

McGill University Genome Quebec Innovation Centre (Montreal, QC, Canada). Sequence fidelity was verified by traditional Sanger sequencing at the Genome Quebec Innovation Center for all PCR products used to establish gene expression profiles or for cloning open reading frames (ORFs) that could be used for biochemical characterization of recombinant enzymes.

4.2.5 - DATABASE ASSEMBLY AND TRANSCRIPT ANNOTATION

The data obtained from pyrosequencing of *Catharanthus*, *Rauvolfia* and *Vinca* materials was assembled using our bioinformatics pipeline as described previously (www.phytometasyn.ca, Facchini, P., *et al*. 2012; Xiao, M., *et al*. 2013). Sequence fidelity was verified by traditional Sanger sequencing at the Genome Quebec Innovation Center for all PCR products used to establish gene expression profiles or cloning open reading frames (ORFs) for biochemical characterization of recombinant enzymes,.

4.2.6 – DATABASE MINING FOR GENES INVOLVED IN IRIDOID AND MIA BIOCHEMICAL PATHWAYS

BLASTn was performed on our *C. roseus*, *V. minor*, *R. serpentina*, *T. elegans*, *A. hubrichtii*, and *L. japonica* transcriptome databases. Known, and putative candidate genes involved in early iridoid and MIA biochemical pathways were identified by searching for homologs of geraniol-10-hydroxylase (CrG10H, AJ251269), ferrihemoprotein-reductase (CrCPR, X69791), loganic acid methyltransferase (CrLAMT, EU057974), secologanin synthase (CrSLS, L10081), tryptophan decarboxylase (CrTDC, M25151), strictosidine synthase (RsSTR, Y00756), and strictosidine beta-glucosidase

(RsSG, (AJ302044). Taxonomically specific MIA pathway steps were also used in this search, including *C. roseus* tabersonine-16-hydroxylase (*CrT16H*, FJ647194), 16-hydroxytabersonine-16-O-methyltransferase (*Cr16OMT*, EF444544), 2,3-dihydrotabersonine-N-methyltransferase (*CrDhtNMT*, HM584929), desacetoxyvindoline-4-hydroxylase (*CrD4H*, U71604), deacetylvindoline-acetyltransferase (*CrDAT*, AF053307), and *R. sepentina* polyneuridine aldehyde esterase (*RsPNAE*, AF178576), vinorine synthase (*RsVS*, AJ556780), acetylajmaline esterase (*RsAAE*, AY762990), raucaffricine beta-glucosidase (*RsRG*, AF149311), perakine reductase (*RsPR*, AY766462). To establish if a homologue was likely to be involved in iridoid or alkaloid biosynthesis arbitrary threshold cutoffs were set high (Score (bits) = 500, E-value = 10e-50, %ID = 85%, Coverage = 100%).

4.2.7 – PHYLOGENETIC ANALYSIS OF APOCYNACEAE AND RELATED PLANT SPECIES

The *matK* chloroplast gene has been used extensively for plant phylogenetic analyses. The *matK* genes from members of the Apocynaceae (aspidospermeae, alstonieae, vinceae, tabernaemontaneae, hunterieae, melodoneae), Apocynoideae, Periplocoideae, Secamonoideae, Asclepiadoideae with known MIA chemistries as well as one out group member of the Loganaceae that only makes Iridoids (AB636281, DQ660520, DQ660517, DQ660502, Z70189, DQ660552, DQ660553, DQ660527, DQ660528, DQ660507, DQ660538, Z70178, DQ66542, DQ660547, AM295067, AM295072, DQ660513, DQ660524, DQ660498, DQ660511, DQ660521, Z70179, DQ660535, DQ660536, DQ660555, DQ660516, DQ6605500, DQ522659, DQ026716) were copied in to the Molecular evolutionary Genetics Analysis version 5 application

(MEGA 5) integrated tool for phylogenetic analysis (Felsenstein, J., 1985). The sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MuSCLE) (Nei, M., Kumar, S., 2000) with the following constraints; Gap Penalties: Open = -400, Extend = 0. Memory/Iterations: Max Memory in Mb = 3354, Max iterations = 8. Clustering Method UPGMB.

The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa being analyzed (Tamura, K., *et al*. 2011). Branches corresponding to partitions reproduced in less than 40% bootstrap replicates are collapsed. The percentage of replicate trees, with values greater than 40%, in which the associated taxa clustered together are shown below the branches (Tamura, K., *et al.* 2011). The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Buckingham, J., *et al.* 2010) with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis involved all 29 *mat*K nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 1248 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Nei, M., Kumar, S., 2000).

4.2.8 – PHYTOCHEMICAL COMPOSITION OF APOCYNACEAE AND RELATED PLANT SPECIES

The Dictionary of Alkaloids (Buckingham, T., *et al*. 2010) was surveyed for the MIAs that include indole *N*-methylated MIA that accumulate in the selected genera that make up the phylogeny. The total number of alkaloids and indole *N*-methylated alkaloids

for each genus was noted for statistical analysis. The relative phylogenetic clustering of indole N-methylated alkaloid accumulation was derived from the indole *N*-methylated molecule index (NMI) based on the following formula; NMI =(number of N-methylated molecules within clade/total N-methylated molecules)/(number of molecules within clade/total molecules). The phylogenetic NMI was analyzed for statistical significance, and because our phytochemical sample size was greater than 30, we calculated the Z-statistic for each genus.

4.2.9 – IDENTIFICATION OF γ-TOCOPHEROL AND TOCOPHEROL-LIKE METHYLTRANSFERASES

The nucleotide sequence of the CrDhtNMT (HM584929) and AtγTMT (AF104220) ORFs was used to independently query all PhytoMetaSyn databases (www.phytometasyn.ca) using the BLASTn Portal and an E-value threshold cut off of 10. Nucleotide sequences were converted to FASTA format, and full ORFs were identified using ALL-IN-ONE-SEQ-ANALYZER version 1.36 ORF prediction utility. ORFs were translated to their corresponding peptide sequences using ALL-IN-ONE-SEQ-ANALYZER, and these peptide sequences were then aligned against each other.

4.2.10 – AMINO-ACID ALIGNMENT OF γ-TOCOPHEROL AND TOCOPHEROL-LIKE METHYLTRANSFERASES

Tocopherol and tocopherol-like amino acid sequences (Table 4-S1) were copied into the MEGA 5 integrated tool for phylogenetic analysis (Felstenstein, J., 1985), and protein sequences were aligned using MuSCLE (Mei, J., Kumar, S., 2000) with the following constraints; Gap Penalties: Open = -2.9, Extend = 0, Hydrophobicity Multiplier

= 1.2. Memory/Iterations: Max Memory in Mb = 4095, Max iterations = 50. Clustering Method UPGMB.

The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa analyzed (Tamura, K., *et al*. 2011). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) is shown next to the branches (Tamura, K., *et al*. 2011). The MP tree was obtained using the Close-Neighbor-Interchange algorithm (Buckingham, J., *et al*. 2010) with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis involved 34 tocopherol *C*-methyltransferase, and tocopherol-like *N*-methyltransferase amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 205 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Felsenstein, J., 1985).

4.2.11 –PREDICTION OF PEPTIDE LOCALIZATION

Tocopherol and tocopherol-like peptide sequences were converted to FASTA format and input into WoLF PSORT online subcellular localization prediction tool for plants (wolfpsort.org, Horton, P., et al. 2007).

4.2.12 – INDEPENDENT VALIDATION OF ASSEMBLED TLMT

TRANSCRIPTS AND RECOMBINANT PROTEIN EXPRESSION

The full-length *CrPeNMT* (Accession number KC708453) and *RsAjNMT*

(Accession number KC708445) sequences were identified from the respective *C.roseus*

and *R. serpentina* 454 MIRA databases by BLASTp search using the published amino

acid sequence of CrDhtNMT (ADP00410) as query.   PCR primers were designed to

amplify the *CrPeNMT* open reading frame (CrPeNMT_FW 5'

TTCATATGGGAGAGAAGGAGGCAGTGG 3', CrPeNMT_Rv 5'

TTGCGGCCGCATTTAGTTTTGCGAAATGTAACTGC 3') and the *RsAjNMT* open

reading frame (RsAjNMT_FW 5' TTCCATGGCAGAGAACCAGGAGGC 3',

RsAjNMT_Rv 5' TTGCGGCCGCAATTAGATTTCCGGCAAGTCAGT 3'). The

presence of 5' NdeI and 3' NotI restriction sites in the *CrPeNMT* primer set, and the 5'

NcoI and 3' NotI in the *RsAjNMT* primer set facilitated directional cloning of each gene

into the pET30b (Invitrogen) *E.coli* expression vector

The desired ORFs were amplified by PCR with pFusion high fidelity DNA

polymerase (New England Biolabs) using the following conditions; 2 minute initial

denaturation 95°C, 35 cycles of 20 second 95°C, 20 seconds 57°C, 60 seconds 72°C,

followed by a 5 minute final extension at 72°C. PCR products were separated by 1.2 %

agarose gel with ethidium bromide. PCR products were then excised from the gel and

purified using a liquid nitrogen gel purification technique. Briefly; gel slices were

quickly frozen in liquid nitrogen, and then placed in a 1.5 mL tube which had a hole and

glass wool in the bottom (reservoir tube). The reservoir tubes were then placed on top of

a new 1.5 mL tube (collection tube) and gel slices were then thawed at 37 ^{o}C. Once gel

slices had thawed the tubes were centrifuged at max speed for 30 seconds to collect PCR

products in the collection tubes. PCR products were then precipitated and washed by

ethanol, dissolved in MilliQ water, and used directly for A-tailing reactions. A-tailing

reactions were performed using ExTaq DNA polymerase (TaKaRa), using all standard

conditions except dNTPs were substituted for dATP, and the reaction was a single step 1

hour at 72°C. A-tailed PCR products were used directly for ligation into pGEM-T Easy

T-A cloning vector (Promega). The ligation reaction was then transformed into XL-10

Gold ultracompotent *E.coli* cells, and plated on LB ampicillin (100 μg / mL), agar plates

with XGAL and IPTG to facilitate blue-white colony selection after being grown

overnight at 37°C. White colonies from each plate were screened by colony PCR to

verify presence of inserts, and colonies verified to contain inserts were sequenced by

GenomeQuebec DNA sequencing services (McGill University, Montreal, Quebec) to

verify the cloned DNA sequence. One colony was then selected for large scale plasmid

purification, NdeI-NotI double digestion, and directional cloning into pET30b *E.coli*

expression vector. Presence of desired ORF in pET30b, and proper direction was

verified by PCR using a combination of gene specific and vector specific PCR primers.

pET30b vectors harboring desired methyltransferase ORFs were then

transformed into Bl21 DE3 Codon+ *E.coli* and plated on LB kanamycin (50 μg / mL)

agar plates and grown overnight at 37°C. The following day single colonies were

selected to produce a 5 mL saturated liquid culture, 500 μl of which was used to inoculate

50 mL liquid LB kanamycin cultures. 50 mL liquid cultures were then grown to OD =

0.6 at 37°C with shaking. At this point recombinant protein expression was induced

using 2 mM final concentration IPTG. Proteins were expressed overnight at room

temperature (~22°C) with shaking. The following day liquid cultures were transferred to 50 mL conical tubes and centrifuged at 3000 xg to pellet *E.coli* cells and LB media was removed. Cell pellets were then re-suspended in 3 mL 100 mL Tris-Cl pH 7.7, with 14 mM mercaptoethanol. Resuspended cells were lysed by sonication. Lysates were again centrifuged to pellet cell debris, and the supernatant was desalted by PD-10 columns (GE Healthcare) to remove small molecules using. The desalted protein extracts were then used directly for enzyme assays and enzyme kinetics.

4.2.13 – TLMT EXTRACTION FROM LEAVES

5 grams of first leaf pair *C.roseus* was harvested and homogenized in a cooled mortar and pestle with 15 mL ice cold buffer (100 mM Tris-Cl pH 7.7, 14 mM mercaptoethanol). Leaf homogenates were then filtered into a 15 mL conical tube with cheese cloth to remove large debris. Filtrates were then centrifuged at 1000 xg to pellet debris. The resulting supernatant was desalted to remove small molecules, and used directly for enzyme assays.

4.2.14 – *N*-METHYLTRANSFERASE ASSAYS

The standard radioactive enzyme assay (150 μL) contained crude desalted protein (recombinant or native), 2.5 nCi S-Adenosyl-L-[14C]-methionine (specific activity 58 mCi/mmol; GE Healthcare Canada), and 5 μg substrate (perivine for PiNMTs)*,* and NMT enzyme assay buffer (100 mM Tris-HCl, pH 7.7, 14 mM mercaptoethanol). Enzyme assays were incubated at 37°C for 1 hour. Assays were stopped by basification (10% v/v 10 N NaOH), and reaction products were extracted to 500 µl ethyl acetate. 50 µL of the

reaction products was used for quantification by scintillation (Beckman LS 6500 Scintillation Counter).

The standard nonradioactive enzyme assay (200 μL) contained crude desalted recombinant NMT protein, 2 mM AdoMet, 5 µg substrate, and enzyme assay buffer (100 mM Tris-HCl, pH 7.7, 14 mM mercaptoethanol). Enzyme assays were incubated at 37$^\text{o}$C for 3 hours. Assays were stopped by basification (10% v/v 10N NaOH), and reaction products were extracted to 500 µl ethylacetate. Organic phases of replicate assays were combined into a single 2 mL tube, and taken to dryness using a SAVANT Speedvac (Thermo Scientific). Dried alkaloid products were re-suspended in 200 μL methanol, and filtered through a 0.22 μm PALL filter (VWR Canada).

4.2.15 – SUBSTRATE SPECIFICITY ASSAYS

Desalted crude *E. coli* protein extract containing recombinant protein (rPeNMT or rAjNMT) was used to assay pure alkaloid as substrates. Initial screening used the standard radioactive enzyme assay with an incubation time of 3 hours. Any substrates that yielded greater than 3x background radioactive counts were repeated using nonradioactive assay conditions in batches of 3 technical replicates. Replicates of nonradioactive enzyme assay products were extracted to organic, pooled, dried and prepared for UPLC-MS analysis as described previously.

Reaction products obtained from rPeNMT, or empty vector, methyltransferase assays with perivine as substrate were analyzed with UPLC (Waters) according to published methods (Roepke, J., *et al*. 1010). Briefly; The analytes were separated using an Aquity UPLC BEH C18 column with a particle size of 1.7 µm and column dimensions

of 1.0×50 mm. Samples were maintained at 4 °C and 5-µL injections were made into the column. The analytes were detected by photodiode array and MS. The solvent systems for alkaloid analysis were as follows: solvent A, methanol: acetonitrile:5-mM ammonium acetate and 6:14:80; solvent B, methanol: acetonitrile:5-mM ammonium acetate at 25:65:10. The following linear elution gradient was used: 0–0.5 min 99% A, 1% B at 0.3 mL/min; 0.5–0.6 min 99% A, 1% B at 0.4 mL/min; 0.6–7.0 min 1% A, 99% B at 0.4 mL/min; 7.0–8.0 min 1% A, 99% B at 0.4 mL/min; 8.0–8.3 min 99% A, 1% B at 0.4 mL/min; 8.3–8.5 min 99% A, 1% B at 0.3 mL/min; and 8.5–10.0 min 99% A, 1% B at 0.3 mL/min. The mass spectrometer was operated with a capillary voltage of 2.5 kV, cone voltage of 34 V, cone gas flow of 2 L/h, desolvation gas flow of 460 L/h, desolvation temperature of 400 °C, and a source temperature of 150 °C.

Reaction products obtained from rAjNMT, or empty vector, methyltransferase assays with ajmaline as substrate were analyzed with UPLC (Waters). Briefly; The analytes were separated using an Aquity UPLC BEH C18 column with a particle size of 1.7 µm and column dimensions of 1.0×50 mm. Samples were maintained at 4 °C and 5-µL injections were made into the column. The analytes were detected by photodiode array and MS. The solvent systems for alkaloid analysis were as follows: solvent A, methanol: acetonitrile: 5-mM ammonium acetate and 6:14:80; solvent B, methanol: acetonitrile:5-mM ammonium acetate at 25:65:10. The following linear elution gradient was used: 0–0.5 min 99% A, 1% B at 0.3 mL/min; 0.5–0.6 min 99% A, 1% B at 0.4 mL/min; 0.6–7.0 min 1% A, 99% B at 0.4 mL/min; 7.0–8.0 min 1% A, 99% B at 0.4 mL/min; 8.0–8.3 min 99% A, 1% B at 0.4 mL/min; 8.3–8.5 min 99% A, 1% B at 0.3 mL/min; and 8.5–10.0 min 99% A, 1% B at 0.3 mL/min. The mass spectrometer was

operated with a capillary voltage of 2.5 kV, cone voltage of 34 V, cone gas flow of 2 L/h, desolvation gas flow of 460 L/h, desolvation temperature of 400 °C, and a source temperature of 150 °C.

4.2.16 – SATURATION KINETICS

For *Km* determinations, crude *E. coli* protein extracts containing recombinant protein (rPeNMT or rAjNMT) were desalted on PD-10 columns and used directly for kinetic analysis. Enzyme assays were performed with 105, 63, 21, 6, 4, and 2 μM perivine or ajmaline at constant concentration of 42 μM (0.075 μCi) (methyl-14C)S-adenosyl-L-methionine, for CrPeNMT, or 125 μM (0.075 μCi) (methyl-14C)S-adenosyl-L-methionine, for RsAjNMT, and with 1, 2, 5, 10 and 20 μM (methyl-14C)S-adenosyl-L-methionine at a constant concentration of 0.2 mM perivine or ajmaline. All assays were performed in triplicate at room temperature for 30 min at pH 7.7 (100 mM Tris–Cl, 14 mM mercaptoethanol). Kinetic constants were calculated using GraphPad Prism 5.

4.2.17 – REAL-TIME QUANTITATIVE PCR

Primer design was performed using with the software PRIMER 3 (Whitehead Institute, MIT, Cambridge, MA, USA) specifying that PCR primers must flank the stop codon such that the 3' UTR of the transcript assists in increasing the specificity of the PCR reaction. The primer pairs Actin-F (5'-GGAGCTGAGAGATTCCGTTG-3') and Actin-R (5'-GAATTCCTGCAGCTTCCATC-3'), CrPeNMT-F (5'-GGAAGCTGCAAGATGTTTAG-3'), and CrPeNMT-R (5'-TTGAGGAACTCTTTGGAGAA -3'), RsAjNMT-F (5'-TGATCCCTCTTCTCGTCTAA-3') and RsAjNMT-R (5'-

CCTCGTCTTACAATGAAACC-3') were used to generate 73 bp, 274 bp and 284 bp

length PCR products respectively. The real-time quantitative PCR reaction was carried

out in a final 25 μl containing 200 nM of each primer, 12 μl iQ^{TM} SYBR Green PCR

Master Mix (Bio-Rad), and 1 μl of cDNA (corresponding to approximately 286 ng

cDNA). Real-time PCR conditions were as follow: 95°C for 15 min, then 40 cycles of

95°C for 10 sec, 55°C for 15 sec and 72°C for 30 sec. After all 40 PCR cycles were

completed a melting curve was performed to establish if the amplified PCR product was

homogeneous. All real-time PCR experiments were run in triplicate for each biological

replicate of cDNA produced from total RNA isolated from fresh *C. roseus*, or *R.

serpentina*, tissues. PCR was controlled using a C1000 thermocycler (Bio-Rad) and

luminescence was detected at the end of each PCR cycle by a CF96 Real-Time System

(Bio-Rad). The average threshold cycle (Ct) and relative quantities were calculated using

CFX manager version 2.1 (Bio-Rad). β-Actin was used as an internal standard to

calculate the relative fold difference based on the comparative Ct method. To determine

relative fold differences for each sample in each experiment, the Ct value for the TLMT

amplicons was normalized to the Ct value for β-Actin, and was calculated relative to a

first leaf pair cDNA, using the formula 2^DDCt (Livak, K., Schmittgen, T., 2001).

4.3 – RESULTS

4.3.1 - PYROSEQUENCING AND TRANSCRIPTOME ASSEMBLY OF cDNA LIBRARIES FROM SIX MEDICINAL PLANTS

Total RNA from young leaves of *C. roseus*, *V. minor*, *T. elegans*, *A. hubrichtii*, and *L. japonica* and from roots of *R. serpentina* were provided to McGill University Genome Quebec Innovation Centre where quality control, cDNA library preparation, and 454 pyrosequencing (GS-FLX Titanium 454 pyrosequencing). Sequencing of the *C. roseus*, *V. minor*, *R. serpentina*, *T. elegans*, *A. hubrichtii* and *L. japonica* cDNA libraries yielded on average 522,561 sequence reads per library, and each library had mean read lengths of 365 nt, 675 nt, 438 nt, 355 nt, 450 nt, 470 nt respectively. The combined total sequence output was 149,308,344 nt, with each library having on average 24,884,724 nt of sequence data (Table 4-1).

4.3.2 – VALIDATING THE TRANSCRIPTOME ASSEMBLIES; RICH RESOURCES TO MINE GENES INVOLVED IN IRIDOID AND MIA BIOSYNTHESIS

To demonstrate that the transcriptome assemblies from each species were rich resources for mining genes involved in iridoid and MIA biochemical pathways, BLASTn searches for functionally characterized iridoid pathway [*CrG10H* (Collu, G., *et al*. 2001), *CrCPR* (Meijer, A., *et al*. 1993), *CrLAMT* (Murata, J., *et al*. 2008), *CrSLS* (Vetter, H., *et al.* 1992)], early MIA pathway [*CrTDC* (De Luca, V., *et al*. 1989), *RsSTR* (Kutchan, T., *et al*. 1988), *RsRG* (Warzecha, H., *et al*. 2000)] and late MIA pathway gene sequences [*CrT16H* (Schroeder, G., *et al*. 1999), *Cr16OMT* (Levac, D., *et al*. 2008), *CrDhtNMT* (Liscombe, D., *et al*. 2010), *CrD4H* (Vazquez-Flota, F., *et al*. 1997), *CrDAT* (St-Pierre, G., *et al*. 1998), *RsPNAE* (Dogru, E., *et al*. 2000), *RsVS* (Bayer, A., *et al*. 2004), *RsAAE* (Ruppert, M., *et al.* 2005), *RsPR* (Sun, L., *et al*. 2008)] were performed. Every member of the Apocynaceae contained contigs corresponding to all published genes, or their close

homologues, known to be involved in iridoid and early MIA biosynthesis (Table 4-2).

Not surprisingly only the *C. roseus* and *R. serpentina* transcriptomes contained genes

known to be involved in the genus specific biochemical pathways of vindoline and

ajmaline biosynthesis, respectively. Remarkably, *L. japonica* produces significant

amounts of secologanin and some other iridoids (Kawai, H., *et al.* 1988), did not contain

easily identifiable iridoid biosynthesis pathway transcripts. However it is possible that

candidate genes from Lonicera could be identified if less stringent sequence identity

conditions were used for the search.

4.3.3 – THE *N*-METHYLATED ALKALOID TRAIT IS TAXONOMICALLY CLUSTERED IN APOCYNACEAE

Evolutionary relationships among the Apocynaceae (Rauvolfiodeae), and closely

related plant families, have previously been reported based on chloroplast DNA

sequences and morphological characters (Simoes, A., *et al*. 2007; Simoes, A., *et al*.

2010). These phylogenetic relationships were reproduced using the chloroplast *matK*

nucleotide sequence marker in representative plant species with well documented MIA

chemistries (Figure 4-1). Inspection of the literature shows that the 29 selected genera

contain at least 2554 documented MIAs with 150 structures being indole *N*-methylated.

These phylogenetic relationships together with documented MIAs were used to calculate

the expected distribution of *N*-methylated structures for each genus, based on a normal

distribution, and contrast it with the observed number of indole *N*-methylated structures

for each genus. This contrast is represented through the NMI where values less than or

greater than 1.00 represent fewer or more indole *N*-methylated structures, respectively

per genus than expected. Indole *N*-methylated structures appear to be clustered within

Alstonieae (NMI = 2.63), Vinceae (NMI = 1.50), Tabernaemontaneae (NMI = 0.96), and Hunterieae (NMI = 2.48) taxonomical clades, and the deviation from the expected number of indole *N*-methylated structures observed in all clades, with the exception of Tabernaemontaneae, was statistically significant ($p < 0.001$) (Figure 4-1). Structures that are methylated at the N4 position of MIAs also appear to be clustered within Strychnos (NMI = 1.41), Aspidospermeae (NMI = 0.96), Alstonieae (NMI = 1.80), Tabernaemontaneae (NMI = 1.32) and Hunterieae (NMI = 2.77) taxonomical clades and the deviation from the expected number of N4 methylated structures, observed in all clades, with the exception of Tabernaemontaneae, was statistically significant ($p < 0.03$) (Figure 4-1).

4.3.4 - TLMTs ARE ONLY PRESENT IN THE TRANSCRIPTOMES OF THE VINCEAE TRIBE OF APOCYNACEAE

Of the plants we selected to sequence only *L. japonica* belongs to the plant family Caprifoliaceae (Kawai, H., *et al*. 1988), whereas *V. minor*, *C. roseus*, *R. serpentina*, *T. elegans*, and *A. hubrichtii* all belong to the plant family Apocynaceae. *V. minor*, *C. roseus* and *R. serpentina* are members of the Vinceae tribe of Apocynaceae, *T. elegans* is a single representative of the Tabernamontaneae tribe which comprises 170 different tabernamontana species (Simoes, A., *et al*. 2007; Simoes, A., *et al*. 2010), and *A. hubrichtii* is a single example of the small Amsonia tribe (Figure 4-1).

Using the published *C. roseus* 16-methoxy-2,3-dihydro-3-hydroxytabersonine *N*-methyltransferase (*CrDhtNMT*) nucleotide sequence (HM584929) as query we performed BLASTn searches of our transcriptome databases. This search revealed hits of

significance (E value <e-10) in the *C. roseus*, *V. minor* and *R. serpentina* databases, whereas the *T. elegans*, *A. hubrichtii*, and *L. japonica* databases do not contain contigs of significant homology. We then performed further BLASTn analysis using the same NMT nucleotide query to search all PhytoMetaSyn assemblies. This publically available BLAST portal contains 36 unique medicinal plant transcriptomes from the Papaveraceae, Asteraceae, Cannabaceae, Mackinlayaceae, Caryophyllaceae, Euphorbiaceae, Ranunculaceae, Guttiferae, Lamiaceae, Verbenaceae, Berberidaceae, Menispermaceae, Platanaceae, Primulaceae, Pinaceae, Celastraceae, Valerianaceae plant families. This extended BLASTn analysis revealed no transcripts of significant similarity to the published *CrDhtNMT* outside of the Vinceae tribe of Apocynaceae, whereas BLASTn searches of this portal using the *Arabidopsis thaliana* γ-tocopherol *C*-methyltransferase (γTMT) nucleotide sequence as query revealed the expected γTMT homologues in all databases. The protein sequences of 10 TLMTs, 5 TLMT-γTMT and 20 representative γTMT homologues were predicted by aligning them against each other using MuSCLE, and a phylogeny of protein relatedness was constructed using a Maximum Parsimony statistical method (Figure 4-2). TLMTs and γTMTs form independent monophyletic clades that are sister, and whose separation is supported by bootstrap analysis. While only Vinceae derived databases possess transcripts that form the TLMT monophyletic clade (Figure 4-2), genes from *T. elegans* (tribe Tabernamontanae) and *A. hubrichtii* (tribe Amsoniae) display intermediate amino acid sequence identities between TLMTs and γTMTs.

4.3.5 – MOLECULAR CLONING OF TLMTs FROM CATHARANTHUS AND RAUVOLFIA

Probing our *C. roseus* young leaf 454 pyrosequencing databases with the *CrDhtNMT* nucleotide sequence identified a unique TLMT, *Cr706*, which was made up of 32 reads, and shows 70% nucleotide sequence and 57 % amino acid sequence identity within the predicted *DhtNMT* open reading frame. The *Cr706* sequence is 1109 nucleotides in length with an ORF of 879 nucleotides. The translated peptide sequence is expected to be 32 kDa and possess an apparent isoelectric point of 6.02.   There was no apparent amino-terminal transit peptide and WolfP analysis suggests that Cr706 protein may be a cytosolic enzyme.

The full *Cr706* ORF was amplified by PCR from cDNA prepared from *C. roseus* first leaf pair total RNA. The amplicon was subcloned into the pGEM-T Easy TA cloning vector and sequenced to verify that there were no mutations incorporated during amplification. A single clone having 100% sequence fidelity to the database *Cr706* sequence ORF was selected, restricted out of the cloning vector using *NdeI* and *NotI* endonucleases and subcloned directionally into the pET30b *E. coli* expression vector.

Probing our *R. serpentina* root pyrosequencing database with the *CrDhtNMT* nucleotide sequence identified a unique TLMT, *Rs820*, which was made up of 346 reads, and shows 72% nucleotide identity within the predicted open reading frame when aligned to the *DhtNMT* sequence. The *Rs820* sequence is 1131 nucleotides in length with an ORF of 876 nucleotides. The translated protein sequence of Rs820 encodes a putative 32 kDa

protein with a pI of 7.1 and it is 54 % identical at the amino acid level to CrDhtNMT. WolfP analysis predicted a cytosolic subcellular location of this protein (score 6.0).

The full *Rs820* ORF was amplified by PCR from cDNA prepared from total RNA isolated from *R. serpentina* roots. The amplicon was subcloned into the pGEM-T Easy TA cloning vector and sequenced to verify that there were no mutations incorporated during amplification. A single clone encoding a protein with 100% DNA sequence fidelity to the database was selected, restricted out of the cloning vector using *NcoI* and *NotI* endonucleases and directionally subcloned into the pET30b *E. coli* expression vector.

4.3.6 – FUNCTIONAL CHARACTERIZATION OF TLMTs

In efforts to identify the biochemical role of *Cr706* potential MIA substrates were extracted from the leaf surfaces of various plant members of the Apocynaceae. Our standard radioactive enzyme assay consisted of recombinant Cr706 protein produced in bacteria, MIA substrates from different plants and ($^{14}CH_3$)-AdoMet. Remarkably, only MIAs from *C. roseus* became labelled, incorporating 28% ($^{14}CH_3$)-AdoMet into the alkaloid fraction. Non radioactive methyltransferase reactions, repeated to produce sufficient product for analysis by UPLC-MS, showed that one *Catharanthus* MIA (Rt, 1.9 min; m/z+ 339) found in enzyme assays using bacteria expressing empty pET30b vector (Figure 4-3) was converted into a methylated product (Rt, 3.4 min; m/z+ 353) in enzyme assays using extracts of bacteria expressing Cr706 (Figure 4-S1). The molecular mass and ultraviolet spectra of the MIA substrate suggested that it could be perivine that is commercially available. Enzyme assays with standard perivine confirmed the *N-*

methylation of perivine (Figure 4-3) and confirmed that *Catharanthus* extracts contain small amounts of perivine. More detailed substrate specificity analysis of recombinant perivine *N*-methyltransferase (CrPeNMT) with 29 MIA substrates (Table 4-S2) revealed its specificity for Perivine since none were accepted for *N*-methylation. Substrate saturation kinetics for the CrPeNMT produced Kms of 35 ± 8.9 μM and 3.3 ± 0.97 μM for picrinine and AdoMet, respectively (Table 4- 4 & 4-5). The pH and temperature optima for this enzyme were 7.5 and 30°C, respectively.

The biochemical role of RsAjNMT was more difficult to identify, since none of the MIA substrates from different plants could be labelled using our standard radioactive enzyme assay. However when substrates from our MIA collection were tested (Table 4-S2), only ajmaline was accepted as a substrate. Standard non-radioactive assays with ajmaline substrate revealed that ajmaline (Rt, 4.71 min; m/z+ 327) present in our empty vector enzyme assays was converted to an novel product with the same retention time (Rt, 4.71) but with a novel mass (m/z+ 341) (Figure 4-4). We have decided to name this enzyme the ajmaline *N*-methyltransferase (RsAjNMT). Substrate Saturation kinetics for RsAjNMT produced Kms of 203.5 ± 34.6 μM and 25 ± 10 μM for ajmaline and AdoMet, respectively (Table 4-5). The pH and temperature optima for this enzyme were 7.5 and 30°C respectively.

4.3.7 – BOTH TLMTs ARE DEVELOPMENTALLY REGULATED

Crude whole *C. roseus* tissue protein extracts (1st, 2nd, 3rd leaf pairs, flower, stem, and root (Figure 4-5C) were assayed for CrPeNMT activities. While young leaves (1st leaf pair) had the highest specific activities (0.15 ± 0.006 pmol/mg/hr) mature leaves (2nd

and 3rd leaf pair, 0.06 – 0.07 $\pm$ 0.01-0.005 pmol/mg/hr), flowers (0.06 $\pm$ 0.005

pmol/mg/hr) and stems (0.03 $\pm$ 0.02 pmol/mg/hr) showed reduced CrPeNMT activities

compared to roots were it was below detectable limits (Figure 5A). The levels of

CrPeNMT transcripts analysed by Real-Time PCR revealed that the stems were highest in

relative transcript abundance (5.56 $\pm$ 0.57 DDCt) followed by second and third leaf pairs

(2.33 $\pm$ 0.15 DDCt and 2.51 $\pm$ 0.68) while first leaf pair (1.46 $\pm$ 0.20 DDCt) and flowers

(1.13 $\pm$ 0.39 DDCt) showed less relative transcript abundance. roots had no detectable

messenger (Figure 4-5B). All reactions were performed in triplicate and NMT transcript

Ct values were normalized to β-Actin Ct values in their respective tissue. The resulting

value (delta Ct) was then normalized to the NMT delta Ct value from leaf pair 1 of the

appropriate plant type to yield the final DDCt for each transcript in each tissue type. This

RT-PCR data analysis was performed previously described (Livak, K., Schmittgen, T.,

2001).

Crude whole *R. serpentina* tissue protein extracts (1st, 2nd, 3rd leaf pairs, flower,

stem, root, Figure 5F) were assayed for RsAjMT activities. *R. serpentina* roots and young

leaves (1st leaf pair) had higher relative activities (0.063 $\pm$ 0.030 and 0.07 $\pm$ 0.006

pmol/mg/hr respectively) than mature mature leaves (2nd and 3rd leaf pair, 0.045 – 0.052

$\pm$ 0.011-0.003 pmol/mg/hr). (Figure 4-5D). In contrast the corresponding transcripts were

not visible in above ground tissues, while this transcript was highest in roots (Figure 4-

5E).

The levels of *RsAjNMT* transcripts were quantified by Real-Time PCR for the

same tissue types analyzed for NMT biochemical activities. Total RNA was isolated to

prepare cDNA templates from each tissue, and RT-PCR reactions were performed

according to experimental procedures. The levels of *RsAjNMT* transcripts analysed by

Real-Time PCR revealed that roots were highest in relative transcript abundance

(2799.21 $\pm$ 509.03 DDCt) followed by stems (414 $\pm$ 139.20 DDCt) while first leaf pair

(1.08 $\pm$ 0.47 DDCt), second leaf pair (7.27 $\pm$ 5.85 DDCt) and third leaf pair (121.40 $\pm$

47.18 DDCt) showed less relative transcript abundance (Figure 4-5 E). All reactions were

performed in triplicate and NMT transcript Ct values were normalized to β-Actin Ct

values in their respective tissue. The resulting value (delta Ct) was then normalized to

the NMT delta Ct value from leaf pair 1 to yield the final DDCt for each transcript in

each tissue type.

4.4 – DISCUSSION

Recent studies have identified a novel class of γ-tocopherol-like methyltransferase

responsible for indole *N*-methylation of the third to last step in vindoline biosynthesis in

C. roseus (Liscombe, D., *et al*. 2010). This novel class of methyltransferase may be key

enzymes for the assembly of at least 150 known MIAs (Table 4-6) and the evolution of

the TLMT gene family could be responsible for taxonomical clustering of indole N-

methylated MIAs within the Apocynaceae. The present study describes the use of

phylogenetic clustering combined with data from the phytochemical literature to predict

that TLMTs may be clustered within certain subfamilies of the Apocynaceae (Figure 4-1

and Figure 4-2). Statistical treatment of these data suggests that the TLMT gene family

appears to be more specific to the Vinceae, the Alstonieae, and the Hunterieae. The

transcriptome databases for plant species from the Vinceae subfamily presented here

(PhytoMetaSyn) led to the identification of 10 candidate TLMTs (Figure 4-2), supporting the results obtained for the Vinceae subfamily as a source of more indole *N*-methylated structures per genus than expected (Figure 4-1, NMI, 1.50). In order to further validate the phylogenetic clustering in Figure 4-1, transcriptomes of a number of representative members of the Alstonieae (Figure 4-1, NMI, 2.63) or Hunterieae (Figure 4-1, NIM, 2.63) should be generated to identify the presence of TLMTs in these preferred clades. Remarkably, TLMTs were not identified in *T. elegans* or *A. hubrichtii* databases as would be expected from their phylogenetic clustering (Figure 4-1, NIM, 0.94 and 0.16).

Two new TLMTs [CrPeNMT (Cr706) and RsAjNMT (Rs820)] were selected from our list of candidate genes (Figure 4-2) for cloning, functional expression and to establish their functions as MIA *N*-methyltransferases.   In functionally characterizing the CrPeNMT, we presented a unique strategy for identifying substrates of recombinant enzymes where no information is available to suggest lead compounds. This strategy would be particularly useful for non-model plants, like *R. serpentina*, where transient or stable gene knock-down methodologies have not been developed, or where these gene knock-down strategies do not yield clear target molecules through the backup of precursor due to redirection of carbon flow to other biochemical pathways. Leaf exudates of *C. roseus* were isolated by performing our standard chloroform dipping protocol (Roepke, J., *et al.* 2010) to isolate crude alkaloid mixtures from different plant species to use in PeNMT enzyme assays and that led to the identification of the perivine substrate.

We also documented the biochemical and developmental characterization of RsAjNMT and CrPeNMT in different plant organs in *Rauvolfia* and in *Catharanthus*. It

is interesting that all *Rauvolfia* and *Catharanthus* tissues appear to have some PeNMT

enzyme activity; these activities did not correspond to the level of transcripts observed in

each tissue (Figure 5).

Biochemically the two TLMTs reported in this manuscript are quite similar. In

terms of saturation kinetics both enzymes, *R. serpentina* AjNMT and *C. roseus* PeNMT,

demonstrate similar affinities, Km, for ajmaline (Km = 203.5 ± 34.6 μM) and perivine

(Km = 35 ± 8.9 μM) respectively. The published DhtNMT Km value for 2, 3-

dihydrotabersonine is 8.8 ± 1.0 μM (Liscombe, D., *et al*. 2010).

The empirically determined CrPeNMT, and RsAjNMT kinetic constants suggest

that perivine and ajmaline may very well be a natural substrate for the *Catharanthus* and

Rauwolfia clones, respectively; however, the necessary transient or stable transgenic

systems are not available for *Rauwolfia* at this time, and virus induced gene silencing of

the CrPeNMT in developing *C. roseus* seedlings did not reveal any clear differences in

the MIAs accumulated relative to control experiments. So while the *C. roseus* DhtNMT

in vivo function in vindoline biosynthesis has been clearly established (Liscombe, D.,

O'Connor, S., 2011), it has yet to be proven that the biological function of the PeNMT

and the AjNMT are to synthesize *N*-methylperivine and *N*-methylajmaline in

Catharanthus and *Rauwolfia*, respectively. *N*-methylajmaline has been documented to

accumulate at very low levels in *Rauwolfia* roots, the primary organ of transcript

expression (Itoh, A., *et al*. 2005). It is unclear why this research group assigned the N4 as

methylated, instead of the indole N1 center, and based on the documented specificity of

TLMTs for indole N1 nitrogens (Liscombe, D., *et al*. 2010) we believe it is reasonable

that our AjNMT, which we have identified and functionally characterized in this

manuscript, is also involved in N1-methylajmaline biosynthesis *in vivo*. In contrast, while perivine has been reported to accumulate in *Catharanthus* (Svoboda, G., *et al.* 1959) and its accumulation in the surface exudates of *C. roseus* leaves allowed us to identify the substrate of the recombinant CrPeNMT, there are no documented biological sources of *N*-methylperivine. This phytochemical data poses an interesting question: what natural biological function do the PeNMTs serve? *Catharanthus* clearly possess a gene encoding an enzyme capable of *N*-methylating perivine, yet the product is not observed in alkaloid extracts of the plant tissues (data not shown). Does this PeNMT represent a silent biochemical pathway branch, spatially or developmentally separate from perivine and related compounds? At this point we cannot confidently answer these questions. However, what is clear is that bioprospecting, phylogenetic and statistical analysis of taxonomically specialized phytochemical profiles can be used to successfully identify novel genes that encode enzymes capable of interesting chemistries. Irrespective of their natural, *in planta*, function, the identification, and functional characterization of these gene products can be used to further combinatiorial biochemical projects purposed to generate novel, and improved pharmaceuticals.

4.5 – ACKNOWLEDGEMENTS

4.6 - ACCESSION NUMBERS

The nucleotide sequences in this paper can be found in the GenBank database under accession numbers KC708448; KC708450; KC708446; KC708452; KC708447; KC708445; KC708453; KC708451

Figure 4-1: Maximum Parsimony analysis of Apocynaceae

The evolutionary history was inferred using the Maximum Parsimony method. The

bootstrap consensus tree inferred from 10000 replicates is taken to represent the

evolutionary history of the taxa analyzed (Felsenstein, J., 1985). Branches corresponding

to partitions reproduced in less than 40% bootstrap replicates are collapsed. The

percentage of replicate trees, with values greater than 40%, in which the associated taxa

clustered together are shown below the branches (Felsenstein, J., 1985). The MP tree was

obtained using the Close-Neighbor-Interchange algorithm (pg. 128 in ref. (Nei, M.,

Kumar, S., 2000)) with search level 1 in which the initial trees were obtained with the

random addition of sequences (10 replicates). The analysis involved 29 *mat*K nucleotide

sequences (AB636281, DQ660520, DQ660517, DQ660502, Z70189, DQ660552,

DQ660553, DQ660527, DQ660528, DQ660507, DQ660538, Z70178, DQ66542,

DQ660547, AM295067, AM295072, DQ660513, DQ660524, DQ660498, DQ660511,

DQ660521, Z70179, DQ660535, DQ660536, DQ660555, DQ660516, DQ6605500,

DQ522659, DQ026716) representing all major clades from Apocynaceae

(Aspidospermeae, Alstonieae, Vinceae, Tabernaemontaneae, Hunterieae, Melodoneae),

the out group Strychnos nux vomica, as well as the ASPA clade (Apocynoideae,

Periplocoideae, Secamonoideae, and Asclepiadoideae). All positions containing gaps and

missing data were eliminated. There were a total of 1248 positions in the final dataset.

Evolutionary analyses were conducted in MEGA5 (Tamura, K., *et al.* 2011). The indole

and terpene moiety N-methylated molecule index (NMI) value for the major clades is

shown above the branches in the format; Indole NMI, terpene moiety NMI. NMI values

were calculated by using the following formula; NMI =(# N-methylated molecules

within clade/total N-methylated molecules)/(# molecules within clade/total molecules).
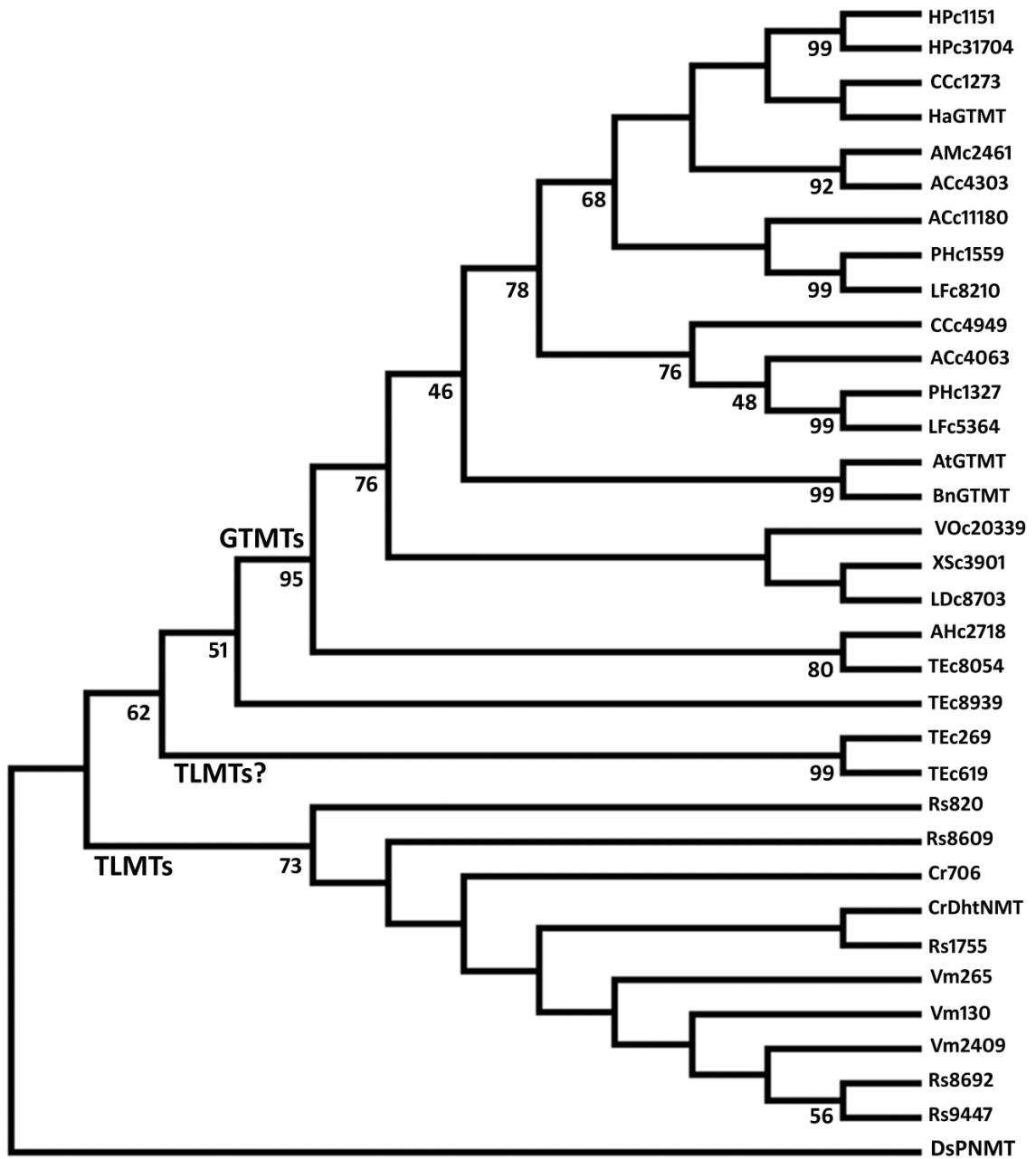
Figure 4-2. Maximum Parsimony analysis of tocopherol and tocopherol-like methyltransferases

The evolutionary history was inferred using the Maximum Parsimony method. The bootstrap consensus tree inferred from 10000 replicates is taken to represent the evolutionary history of the taxa analyzed (Felsenstein, J., 1985). Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (10000 replicates) are shown below the branches (Felsenstein, J., 1985). The MP tree was obtained using the Close-Neighbor-Interchange algorithm (pg. 128 in ref. (Nei, M., Kumar, S., 2000)) with search level 1 in which the initial trees were obtained with the random addition of sequences (10 replicates). The analysis involved 34 amino acid sequences. With the exception of biochemically characterized gamma-tocopherol C-methyltransferases from *Arabidopsis thaliana* (AtGTMT, AAD02882), *Helianthus annus* (HaGTMT, ABB52798), and *Brassica napus* (BnGTMT ACJ54674), or the biochemically characterized tocopherol-like N-methyltransferase from *Catharanthus roseus* (CrDhtNMT, *ADP00410), all amino acid sequences were derived from predicted open reading frames of isotigs identified by performing BLASTp searches of all Phytometasyn RNA-seq transcriptome databases.* All positions containing gaps and missing data were eliminated. There were a total of 205 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 (Tamura, K., *et al.* 2011). Tocopherol-like (TLMT) and tocopherol (GTMT) sister clades are noted above branches.
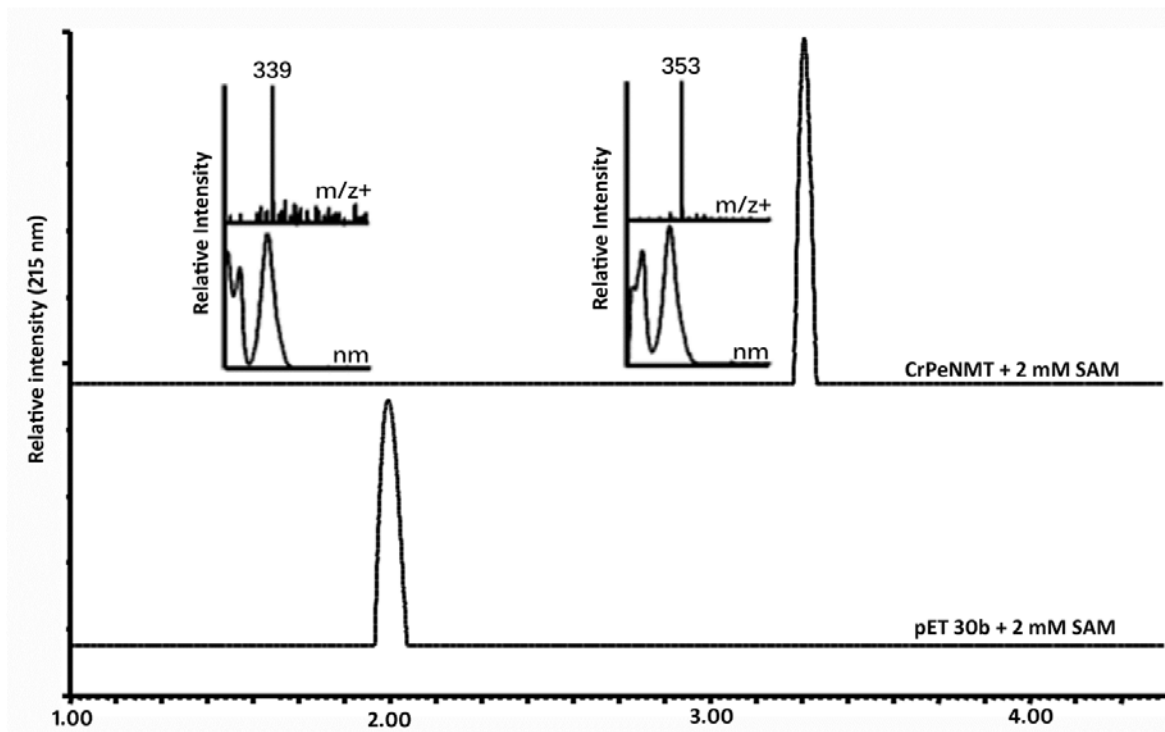
Figure 4-3. Conversion of perivine to *N*-methylperivine is catalyzed by recombinant CrPeNMT.

UPLC chromatograms showing the conversion of perivine (Rt 2.05, m/z+ 339) into N-methylperivine (Rt 3.40, m/z+ 353) by recombinant CrPeNMT in the presence of 2 mM SAM, and alkaloid substrate under standard cold methyltransferase assay conditions. UV and MS + spectra of substrate and product are inset.
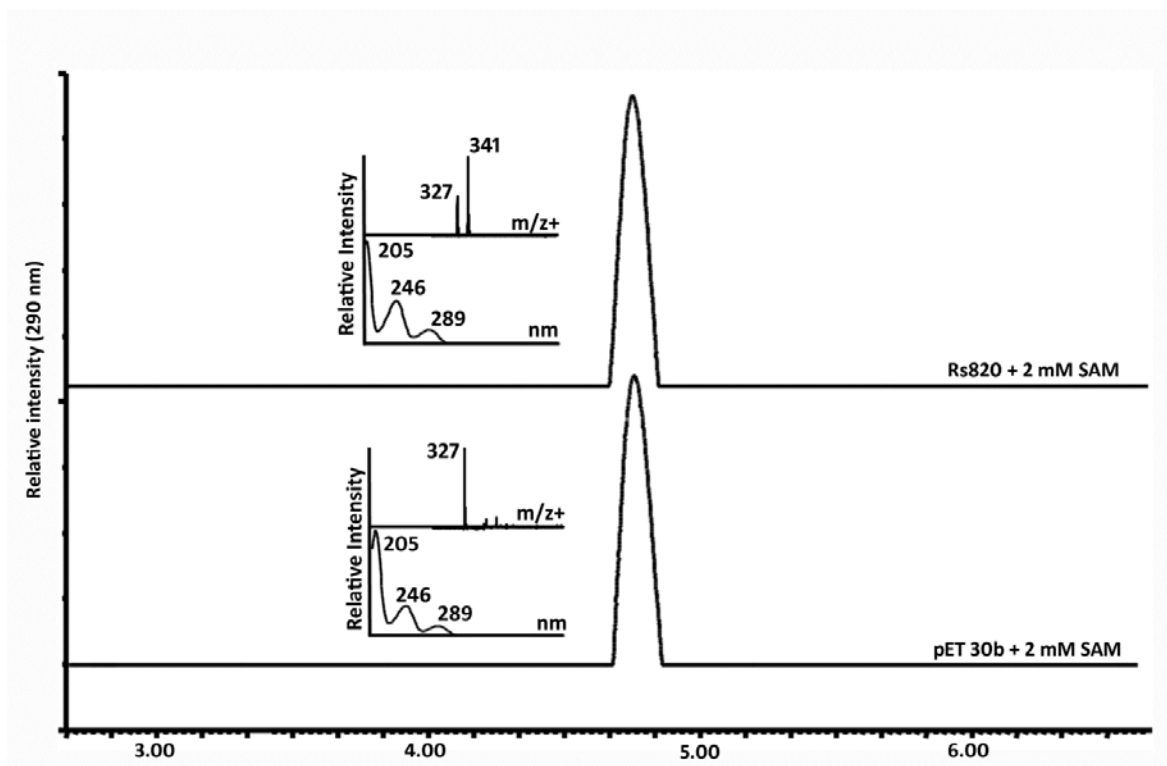
Figure 4-4. Conversion of ajmaline to *N*-methylajmaline is catalyzed by recombinant RsAjNMT.

UPLC chromatograms showing the conversion of ajmaline (Rt 4.71, m/z+ 327) into N-methylajmaline (Rt 4.71, m/z+ 341) by recombinant RsAjNMT in the presence of 2 mM SAM, and alkaloid substrate under standard cold methyltransferase assay conditions. UV and MS + spectra of substrate and product are inset.
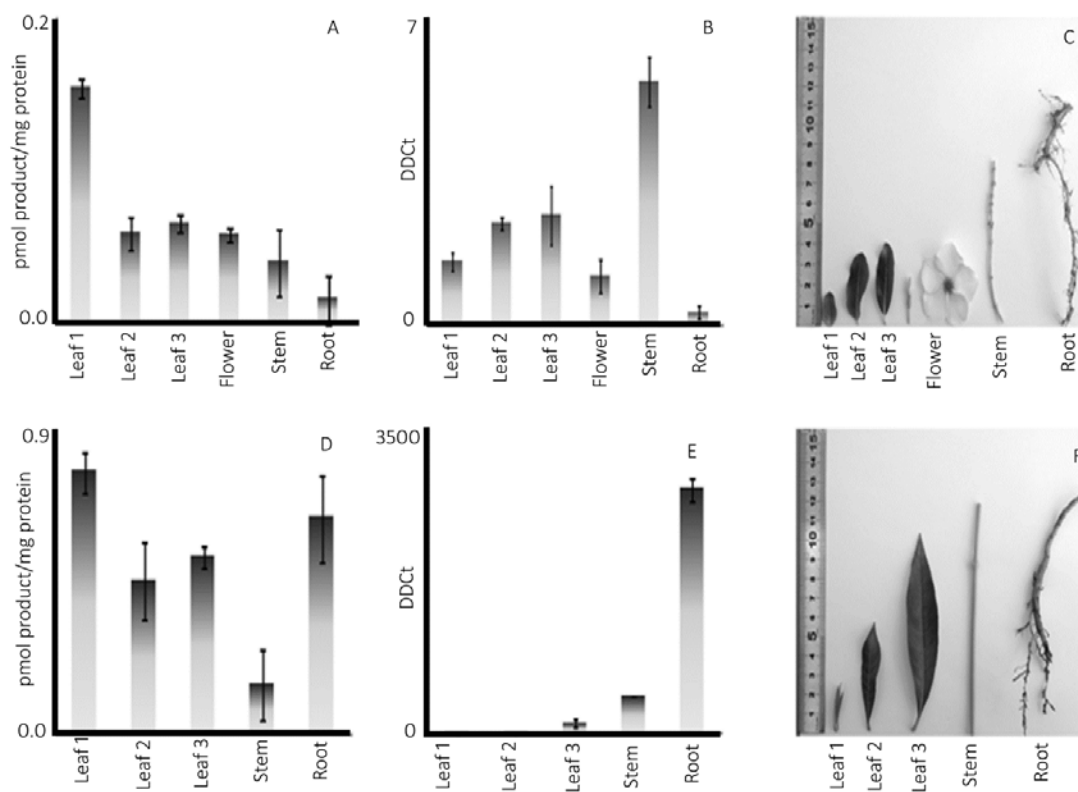
Figure 4-5. CrPeNMT and RsAjNMT enzyme activities are coordinated with gene expression in different tissues of *C. roseus* and *R. serpentina*. A) Enzyme activity profile and B) Relative abundance of *CrDhtNMT* transcripts in C) representative tissues of *C. roseus*. D) Enzyme activity profile and E) Relative abundance of *RsAjNMT* transcripts in F) Representative tissues of *R. serpentina*.
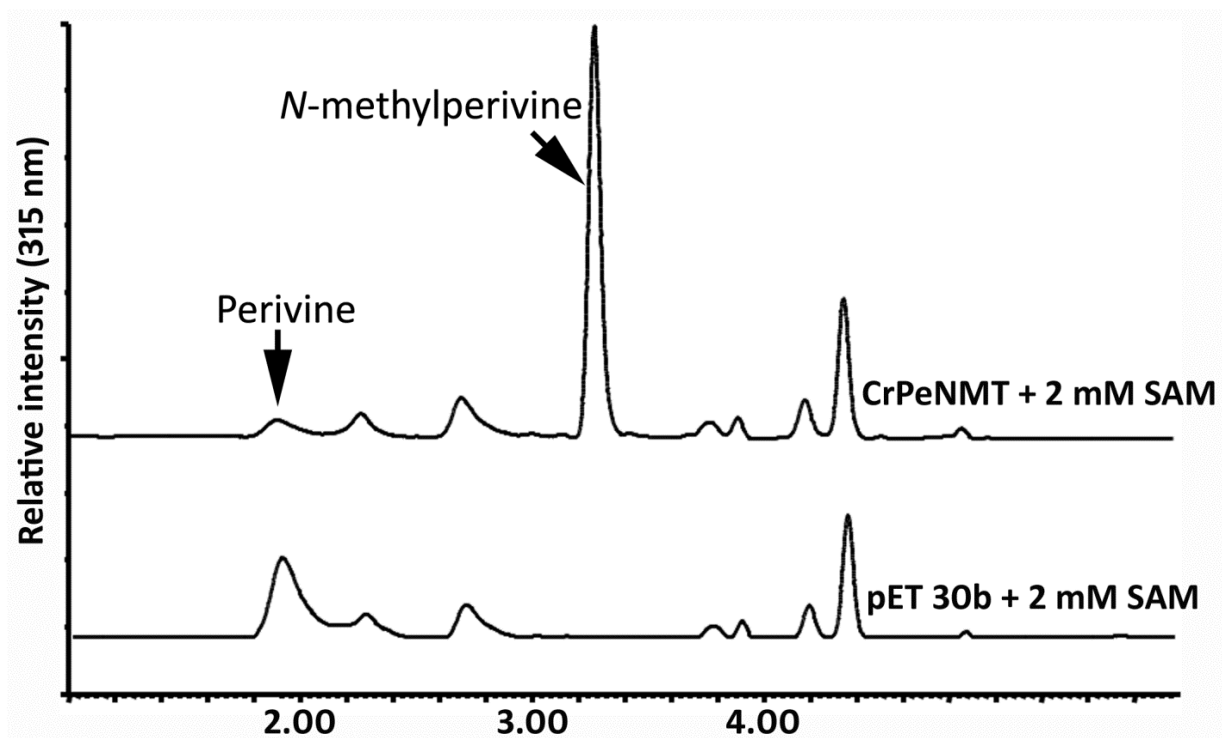
Figure 4-S1. Conversion of perivine into N-methylperivine using C. roseus surface exudates as substrate

UPLC chromatograms showing the conversion of perivine (Rt 2.05, m/z+ 339) into N-methylperivine (Rt 3.40, m/z+ 353) by recombinant CrPeNMT in the presence of 2 mM SAM, and mixed alkaloid substrate from *C. roseus* surface exudates under standard cold methyltransferase assay conditions.

Table 4-1: Metrics of MIRA transcriptome assemblies

Organism	Bases	Reads	Contigs	Mean read length (bp)	Average contig length (bp)	Mean reads per contig	Singletons
C.roseus	19993504	588162	26804	365	746	2-3	0
V.minor	36194656	576174	41053	675	882	2-3	426
R.serpentina	23451585	517054	25548	438	918	2-3	0
T. elegans	26756604	536535	28744	355	931	2-3	0
A. hubrichtii	24648381	558246	30348	450	812	2-3	0
L. japonica	18263614	359192	20828	470	877	2-3	2

Table 4-2: Transcriptome database validation. BLASTn results of our MIA and iridoid
pathway enriched transcriptome databases demonstrating presence of conserved iridoid,
and early MIA pathway between genus. BLASTn criteria: Cutoff – Score (bits) = 500, E-
value = 10e-50, %ID = 85%, Coverage = 100%, *E-value = 4e-20

Gene	*C.roseus* (cluster size)	*V.minor* (cluster size)	*R.serpentina* (cluster size)	*T. elegans* (cluster size)	*A. hubrchtii* (cluster size)	*L. japonica* (cluster size)
G10H	Cr1080 (86)	Vm373 (143)	Rs589 (102)	Te1671 (56)	Ah611 (111)	0
CPR	Cr2241 (74)	Vm1851 (59)	Rs424 (239)	Te2175 (62)	Ah201 (380)	Lj1366 (101)
LAMT	Cr240 (329)	Vm79 (301)	Rs285 (211)	Te2054 (42)	Ah66 (621)	0
SLS	Cr387 (394)	Vm1330 (116)	Rs60 (757)	Te2033 (367)	Ah136 (541)	0
STR	Cr884 (97)	Vm567 (81)	Rs454 (84)	Te3478 (23)	Ah2588 (43)	0
TDC	Cr288 (249)	Vm215 (176)	Rs348 (188)	Te3237 (31)	Ah5235 (36)	0
SGD	Cr318 (370)	Vm1191 (62)	Rs507 (133)	Te3525 (40)	Ah20 (1034)*	0
T16H	Cr1347 (62)	Vm1620 (39)	0	0	0	0
16OMT	Cr296 (243)	0	0	0	0	0
2,3-	Cr168	0	0	0	0	0

NMT	(370)					
D4H	Cr138 (623)	0	0	0	0	0
DAT	Cr484 (202)	0	0	0	0	0
PNAE	0	0	Rs580 (58)	0	0	0
VS	0	0	Rs48 (555)	0	0	0
AAE	0	0	Rs606 (122)	0	0	0
RG	0	0	Rs257 (218)	0	0	0
PR	Cr1564 (67)	Vm3205 (28)	Rs34 (675)	Te2324 (37)	Ah2768 (30)	0

Table 4-3: Number of N-methylated MIAs documented in the literature for each clade in apocynaceae.

Total number of phytochemical, and indole N-methylated structures, documented in the literature for each clade of phylogeny (Figure 1). T-statistic analysis of *N*-methylated clustering.

Clade	Total documented alkaloid structures	Number of indole *N*-methylated molecules		NMI value	Z-statistic (p value)
		Expected	Actual		
APSA	52	3	0	0.00	<0.001
Melodineae	103	6	1	0.16	<0.001
Hunterieae	137	8	20	2.48	<0.001
Tabernaemontaneae	307	18	17	0.94	0.47
Vinceae	680	39	61	1.50	<0.001
Alstonieae	291	17	45	2.63	<0.001
Aspidospermeae	232	13	5	0.37	<0.001
Loganaceae	472	28	1	0.04	<0.001

Table 4-4: Comparative kinetic parameters for CrPeNMT and RsAjNMT with CrDhNMT. Saturation kinetic parameters for alkaloid substrates for recombinant CrPeNMT and RsAjNMT enzymes were conducted at constant, saturating 42 μM and 125 μM AdoMet respectively. AdoMet saturation kinetic parameters for CrPeNMT and RsAjNMT recombinant enzymes were conducted at constant, saturating, 200 μM alkaloid substrate.

Enzyme	Substrate	Km (μM)	Vmax (fmol/s/mg)	Kcat (s^{-1})	Kcat/Km ($S^{-1} M^{-1}$)
CrPeNMT	perivine	35 $\pm$ 8.9	221.4 $\pm$ 21.8	4.4 $\pm$ 0.4	125 714
RsAjNMT	ajmaline	203.5 $\pm$ 34.6	528.9 $\pm$ 37.2	5.8 $\pm$ 0.02	28 571
CrDhNMT[1]	dihydrotabersonine	8.8 $+$ 1.0	67 000 $\pm$1800	5.4 $\pm$ 0.1	267 053
CrPeNMT	AdoMet	3.3 $\pm$ 0.97	73.8 $\pm$ 7.0	0.17 $\pm$ 0.01	56 666
RsAjNMT	AdoMet	25 $\pm$ 10	440.3 $\pm$ 115	3.6 $\pm$ 0.9	144 000
CrDhNMT	AdoMet	22.0 $\pm$ 1.8	151 600 $\pm$ 3900	5.4 $\pm$ 0.01	243053

[1]- Saturation kinetic parameters for CrDhtNMT were copied from Liscombe, D., *et al*. 2010, PNAS

Table 4-6: Indole N-methylated molecules by genera.

The Dictionary of Alkaloids (CRC Press).

Genus	Organism	Alkaloid
Loganaceae	*Strychnos usambarensis*	N4-methyl-10-hydroxyusambarine
Aspidospermeae	*Aspidosperma quebrachoblanco*	N-methylaspidospermatidine
	Aspidosperma cylindrocarpon	N-methylcylindrocarine
	Haplophyton cimicidum	Cimiciphytine
		Haplophytine
	Haplophyton crooksii	10-methoxy-N-methylpericyclivine
Alstonieae	*Alstonia agustifolia*	Alstocraline
		Alstolactone
		Angusticraline
		affinisine oxindole
	Alstonia macrophylla	Alstomacroine
		Alstomacroline
		Alstomicine
		Alstonal
		Alstonisine
		clstonoxine B

alstophyllal

alstoumerine

macralstonidine

macralstonine

macrocarpamine

macrodasine A

macrogentine

macrosalhine

macroxine

perhentinine

quaternine

10-methoxyaffinisine

strictaminolamine

affinisine

talcarpine

villastonine

Alstonia constricta	alstonidine
Alstonia lanceolifera	N-methyl-10-methoxyakuammidine
Alstonia mueleriana	alstonisidine
Alstonia scholaris	N-methylburnamine
	O-deacetyl-1,2,4,5-tetrahy dro-2-hydroxy-1-methylakuammilinium
Alstonia plumosa	desoxycabufiline
	plumocraline

		3,4-Seco-3,14-dehydrocabucraline
	Alstonia quaternata	cathafoline
		Vincamajine
	Alstonia undulata	17-hydroxydehydrovoachalotine
		Gentiacraline
		Undulatine
	Alstonia lanceolata	Lanceomigine
	Alstonia quaternata	Quaternidine
	Alstonia deplanchei	Pleiocorine
		Pleiocraline
	Alstonia legouixiae	Voachalotine
	Alstonia vitiensis	Vincorine
Vinceae	*Petchia ceylanica*	Ceylanicine
		Peceylanine
		Peceyline
		Pelankine
	Cabucala caudata	Cabufiline
	Cabucala erythrocarpa	Akuammine
		Cabuamine
		Ochropamine
		Vincorine

	caberoline
Cabucala torulosa	vincamajine
Rauwolfia serpentina	ajmalinimine
	raumacline
	rauwolfinine
	sandwicolidine
	sandwicoline
	ajmaline
	ajmalinimine
Rauwolfia reflexa	flexicorine
	raureflexine
Rauwolfia bahiensis	12-methoxyaffinisine
Rauwolfia oreogiton	quaternine
Rauwolfia canescens	raucanine
Rauwolfia sumatrana	rausutranine
	rausutrine
Rauwolfia suaveolens	suaveoline
Rauwolfia oreogiton	2,7,19,20-tetrahydro-3- hydroxy-1-methylvobasan-17-oic acid-d-lactone
Rauwolfia tetraphylla	tetraphyllicine

Rauwolfia nitida	N-methylvellosmine
Catharanthus roseus	Catharine
	Leurosinone
	Pseudovinblastinediol
	Vinamidine
	Vinblastine
	Vincathicine
	Vindolicine
	Vingramine
	Bannucine
	Vindoline
Catharanthus ovalis	Cathovaline
Ochrosia powerii	Ochropamine
Vinca herbacea	Akuammine
Vinca sardoa	N-methylaspidofractinine
Vinca minor	N-methylaspidospermidine
Vinca erecta	Ervincine
	4- methylraucubaininium(1+)
	N-methyl-2,16 -dihydroakuammicine
	Minovine
	Vincaminoreine
	Vincaminorine

		vincarubine
		vincatine
		vincorine
	Vinca libanotica	herbamine
		vincamajine
	Vinca pusilla	vindolidine
	Tonduzia longifolia	ajmaline
		18-hydroxycabucraline
		10-methoxy-11-[10-(11-methoxyvincamajinyl)]cathafoline
	Kopsia dasyrachis	pleiomutine
Tabernamontaneae	*Tabernaemontana accedens*	accedine
		16-epi-*N*-methylaffinine
		accedinisine
		16-epi-*N*-methylaffinine
	Tabernaemontana bovina	14,15-dihydroxy-*N*-methylaspidospermidine
		14,15- dihydroxy-*N*-methylaspidospermidine
		methylenebismehranine
		tabernaebovine
		tabernaemontabovine
	Tabernaemontana divaricata	conofoline

		11-methoxy-*N*-methyl dihydropericyclivine
		mehranine
	Tabernaemontana saccharatum	Pandicine
	Tabernaemontana fuchsiaefolia	Voachalotine
		Affinisine
	Stemmadenia obovata	N1-methyl-11-hydroxymacusine A
Hunterieae	*Hunteria congolana*	aberamine 1-4
	Hunteria congolana	17-hydroxy-akuammigine
		Lanciomigine
		Eripinal
	Hunteria zeylanica	Coreyzeylamine
		Deformylcoryzeylamine
		Desformocorymine
	Hunteria corymbosa	Dihydrocorymine
	Hunteria umbellata	Eripine
		isocorymine
		Umbellamine
	Picralima nitida	Akuammine
		Picratidine
	Pleoicarpa mutica	Pleiocarpinine

		N-methylsarpagine
		pleiomutine
	Pleoicarpa talbotii	talcarpine
		talpinine
Melodineae	*Melodinus balansae*	ajmaline

Table 4-S1 Abbreviations, Annotatins and Accession numbers for sequences used to

construct the MP tree in this study

Abbreviation	Annotation [organism]	Accession #
CrDhtNMT	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	HM584929
Cr7756	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	Liscombe, D., *et al.* 2010
RsPiNMT	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708448
VmPiNMT	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708450
Rs1755	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708446
Vm2409	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708452
Rs8609	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708447
Rs820	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708445
Cr706	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708453
Vm265	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708451
Cr1196	gamma-tocopherol methyltransferase-related protein [Catharanthus roseus]	HM584930
HaGTMT	Gamma-tocopherol methyltransferase [*Helianthus annuus*]	ABB52798
AtGTMT	Gamma-tocopherol methyltransferase [*Arabidopsis thaliana*]	AAD02882
BnGTMT	Gamma-tocopherol methyltransferase [*Brassica napus*]	ACJ54674
Cr2551	Putative methyltransferase [*Catharanthus roseus*]	HM584931

Table 4-S2: List of alkaloid substrates used in substrate specificity studies.
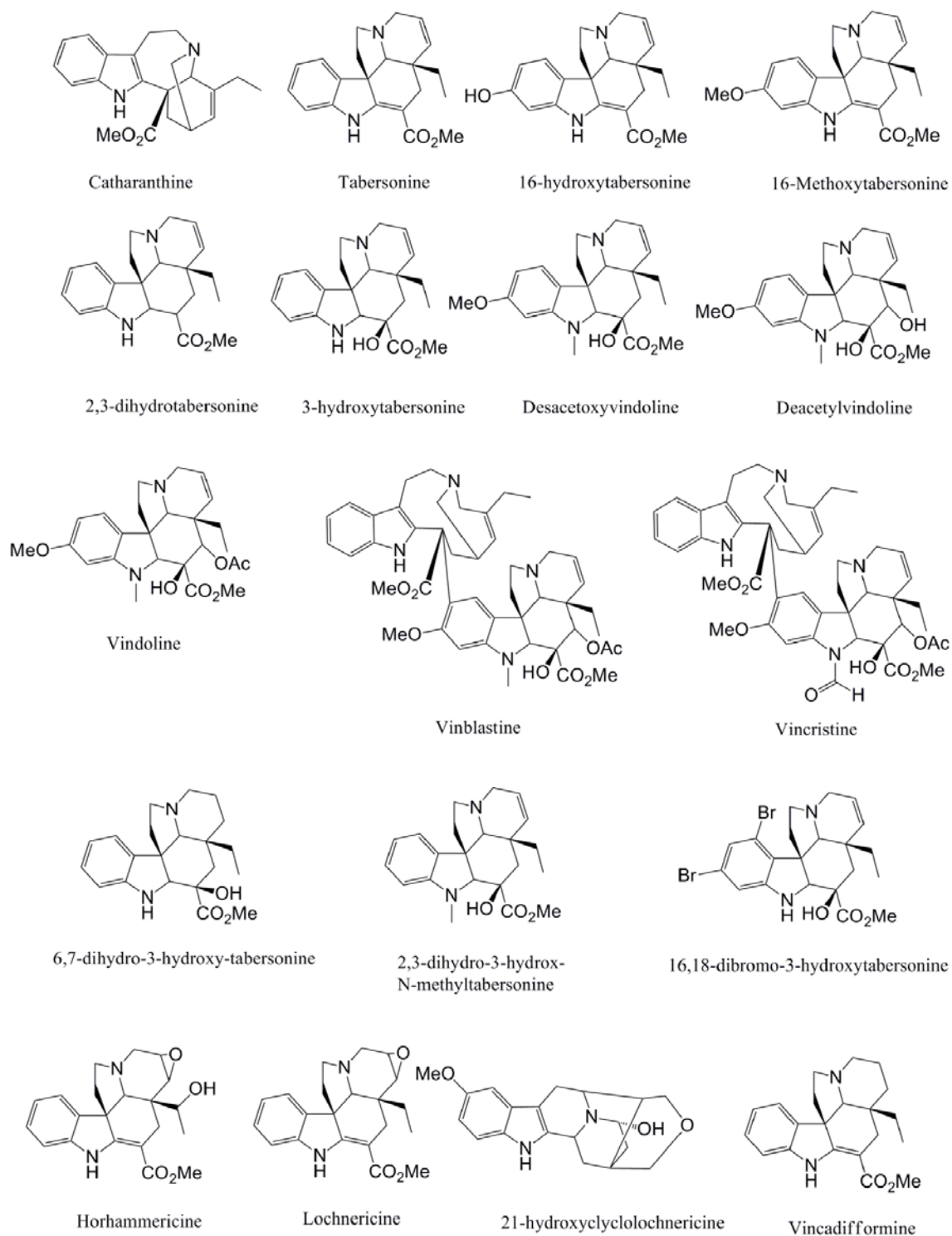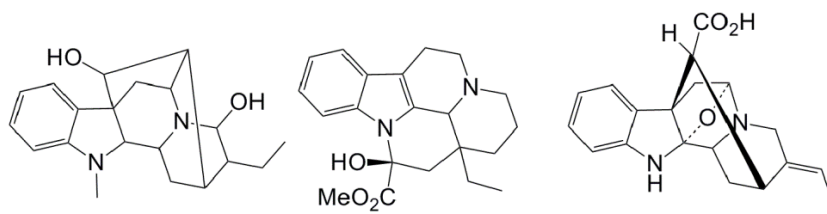
Catharanthine

Tabersonine

16-hydroxytabersonine

16-Methoxytabersonine

2,3-dihydrotabersonine

3-hydroxytabersonine

Desacetoxyvindoline

Deacetylvindoline

Vindoline

Vinblastine

Vincristine

6,7-dihydro-3-hydroxy-tabersonine

2,3-dihydro-3-hydrox-N-methyltabersonine

16,18-dibromo-3-hydroxytabersonine

Horhammericine

Lochnericine

21-hydroxyclyclolochnericine

Vincadifformine

Ajmaline

Vincamine

Picrinine
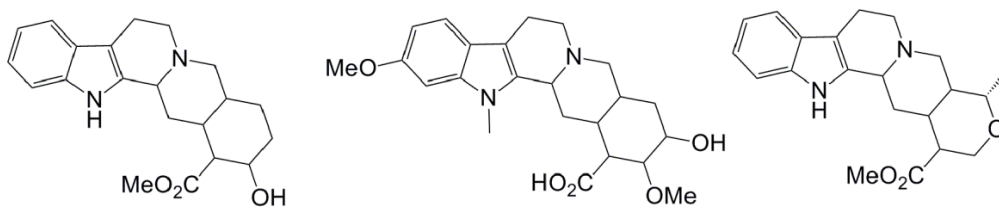
Yohimbine

Reserpinic acid

Ajmalicine
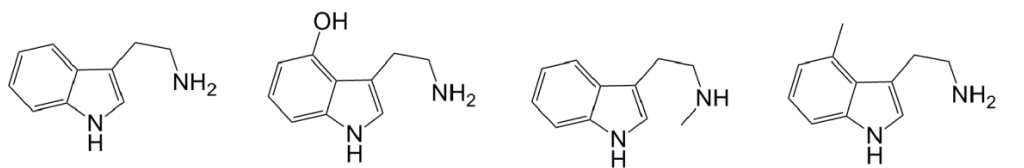
Tryptamine

5-Hydroxytryptamine

N-methyltryptamine

5-methyltryptamine

4.7 – BIBLIOGRAPHY

Bayer, A., Ma, X., Stockigt, J., (2004) Acetyltransfer in natural product biosynthesis--functional cloning and molecular analysis of vinorine synthase, *Bioorg Med Chem*, **12**, 2787-2795.

Broach, J., Thorner, J., (1996) High-throughput screening for drug discovery, *Nature*, **384**, 14-16.

Buckingham, J., Baggaley, K., Roberts, A., Szabo, L., eds. (2010) Dictionary of Alkaloids, CRC Press Taylor & Francis Group, Boca Raton, FL.

Collu, G., Unver, N., Peltenburg-Looman, A.M., van der Heijden, R., Verpoorte, R., Memelink, J., (2001) Geraniol 10-hydroxylase, a cytochrome P450 enzyme involved in terpenoid indole alkaloid biosynthesis, *FEBS Letters*, **508**, 215-220.

De Luca, V., Marineau, C. and Brisson, N., (1989) Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases, *Proc Nat Acad Sci*, **86**, 2582-2586.

De Luca, V., Salim, V., Masada-Atsumi, S., Yu, F., (2012) Mining the Biodiversity of plants: a revolution in the making, *Science*, 336, 1658-1661.

Dethier, M., De Luca, V., (1993) Partial purification of an N-methyltransferase involved in vindoline biosynthesis in Catharanthus roseus, *Phytochemistry*, **32**, 673-678.

Dogru, E., Warzecha, H., Seibel, F., Haebel, S., Lottspeich, F., Stockigt, J., (2000) The gene encoding polyneuridine aldehyde esterase of monoterpenoid indole alkaloid

biosynthesis in plants is an ortholog of the alpha/betahydrolase super family, *Eur J Biochem*, **267**, 1397-1406.

Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Research*, 32, 1792-97.

Facchini, P., Bohlmann, J., Covello, P., De Luca, V., Mehadevan, R., Page, J., Ro, D., Sensen, C., Storms, R., Martin, V., (2012) Synthetic biosystems for the production of high yield metabolites, *Trends in Biotechnol*, **30**, 127-131.

Feher, M., Schmidt, M., (2003) Property distrubtions: differences between drugs, natural products, and molecules from combinatorial chemistry, *J Chem Information and Modeling*, **43**, 218-227.

Felsenstein, J., (1985) Confidence limits on phylogenies: an approach using the bootstrap, *Evolution*, **39**, 783-791.

Glenn, W., Runguphan, W., O'Connor, S., (2012) Recent progress in the metabolic engineering of alkaloids in plant systems, *Current Opin Biotech*, [Epub ahead of print]

Horton, P., Park, K., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. , Nakai, K., (2007) WoLF PSORT: Protein Localization Predictor, *Nucleic Acids Research*, 259, 585-587.

Itoh, A., Kumashiro, T., Yamaguchi, M., Nagakura, N., Mizushina, Y., Nishi, T., Tanahashi, T., (2005) Indole alkaloids and other constituents of Rauwolfia serpentina, *J of Nat Prod*, **68**, 848-852.

Kawai, H., Kuroyanagi, M., Ueno, A., (1988) Iridoid Glucosides from Lonicera Japonica THUNB. *Chem Pharm Bull*, **36**, 3664-3666.

Kumar, D., Kumar, N., Chang, K., Gupta, R., Shah, K., (2011) Synthesis and in-vitro anticancer activity of 3,5-bis(indolyl)-1,2,4-thiadiazoles, *Bioorg and Med Chem Lett*, **21,** 5897-5900.

Kumar, D., Sandaree, S., Johnson, E., Shah, K., (2009) An efficient synthesis and biological study of novel indolyl-1,3,4-oxadiazoles as potent anticancer agents, *Bioorg and Med Chem Lett*, **19**, 4492-4494.

Kutchan, T.M., Hampp, N., Lottspeich, F., Beyreuther, K., Zenk, M.H., (1988) The cDNA clone for strictosidine synthase from Rauvolfia serpentina. DNA sequence determination and expression in Escherichia coli*, FEBS Letters*, **237**, 40-44.

Levac,D., Murata,J., Kim,W.S., De Luca,V., (2008) Application of carborundum abrasion for investigating the leaf epidermis: molecular cloning of Catharanthus roseus 16-hydroxytabersonine-16-O-methyltransferase, *The Plant Journal*, **53**, 225-236.

Li, J., Vederas, J., (2009) Drug discovery and natural products: end of an era or endless frontier, *Science*, **325**, 161-165.

Liscombe, D., O'Connor, S., (2011) A virus induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*, *Phytochemistry*, **72**, 1969-1977.

Liscombe, D.K., Usera, A.R., O'Connor, S.E., (2010) Homolog of tocopherol C methyltransferases catalyzes N-methylation in anticancer alkaloid biosynthesis, *Proc Nat Acad Sci*, **107**, 18793-18798.

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

Malik, N., (2008) Drug discovery: past, present, and future, *Drug Discovery Today*, **13**, 909-912.

Meijer, A.H., Lopes Cardoso, M.I., Voskuilen, J.T., de Waal, A., Verpoorte, R., Hoge, J.H., (1993) Isolation and characterization of a cDNA clone from Catharanthus roseus encoding NADPH:cytochrome P-450 reductase, an enzyme essential for reactions catalysed by cytochrome P-450 mono-oxygenases in plants, *The Plant Journal*, **4**, 47-60.

Murata, J., Roepke, J., Gordon, H., De Luca, V., (2008) The leaf epidermome

Nei, M., Kumar, S., (2000) "Molecular Evolution and Phylogenetics," Oxford University Press, New York.

Newmann, D., Cragg, G., (2007) "Natural products as sources of new drugs over the last 25 years," *J Nat Prod*, **70**, 461-477.

Paul, S., Mytelka, D., Dunwiddle, C., Persinger, C., Munos, B., Lindborg, S., Schacht, A., (2010) How to improve R&D productivity: the pharmaceutical industry's grand challenge, *Nature Reviews Drug Discovery*, **9**, 203-214.

Roepke, J., Salim, V., Wu, M., Thamm, A., Murata, J., Ploss, K., Boland, W., De Luca, V., (2010) Vinca drug components accumulate exclusively in leaf exudates of

Madagascar periwinkle, *Proceedings of the National Academy of Science*, **107**, 15287-15292.

Ruppert, M., Woll, J., Giritch, A., Genady, E., Ma, X., Stockigt, J., (2005) Functional expression of an ajmaline pathway-specific esterase from Rauvolfia in a novel plant-virus expression system, *Planta*, **222**, 888-898.

Saslis-Lagoudakis, C., Savolainen, V., Williamson, E., Forest, F., Wagstaff, S., Baral, S., Watson, M., Pendry, C., Hawkins, J., (2012) Phylogenies reveal predictive power of traditional medicine in biprospecting, *Proceedings of the National Academy of Science*, **109**, 15835-15840.

Sboner, A., Mu, X., Greenbaum, D., Auerbach, R., Gerstein, M., (2011) The real cost of sequencing: higher than you think!, Genome Biology, **12**, 125.

Schreiber, S., (2000) Target-oriented and diversity-oriented organic synthesis in drug discovery, *Science*, **287**, 1964-1969.

Schroeder, G., Unterbusch, E., Kaltenbach, M., Schmidt, J., Strack, D., DeLuca, V., Schroeder, J., (1999) Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase, *FEBS Letters*, **458**, 97-102.

Simoes, A., Endress, M., Conti, E., (2010) Systematics and character evolution of Tabernaemontaneae (Apocynaceae, Rauvolfioideae) based on molecular and morphological evidence, *Taxon*, **59**, 772-790.

Simoes, A., Livshultz, T., Conti, E., Endress, M., (2007) Phylogeny and Systematics of the Rauvolfioideae (Apocynaceae) Based on Molecular and Morphological Evidence, *Ann Missouri Bot Gard*, **94**, 268-297.

St-Pierre, B., Laflamme, P., Alarco, A.M., De Luca,V., (1998) The terminal O-acetyltransferase involved in vindoline biosynthesis defines a new class of proteins responsible for coenzyme A-dependent acyl transfer, *The Plant Journal*, **14,** 703-713

Sun, L., Ruppert, M., Sheludko,Y., Warzecha, H., Zhao,Y., Stockigt, J., (2008) Purification, cloning, functional expression and characterization of perakine reductase: the first example from the AKR enzyme family, extending the alkaloidal network of the plant Rauvolfia, *Plant Mol Biol*, **67**, 455-467.

Svoboda, G., Neuss, N., Gorman, M., (1959) Alkaloids of Vinca rosea Linn. (Catharanthus roseus G. Don.) V. Preparation and characterization of alkaloids, *J Am Pharm Assoc*, **48**, 659-666.

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

Tan, D., (2005) "Diversity-oriented synthesis: exploring the intersection between chemistry and biology, *Nature Chem Biol*, **1**, 74-84.

Tarselli, M., Raehal, K., Brasher, A., Streicher, J., Groer, C., Cameron, M., Bohn, L., Micalizio, G., (2011) Synthesis of conolidine, a potent non-opioid analgesic for tonic and persistent pain, *Nature Chem*, **3**, 449-453.

Vazquez-Flota, F., De Carolis, E., Alarco, A.M., De Luca, V., (1997) Molecular cloning and characterization of desacetoxyvindoline-4-hydroxylase, a 2-oxoglutarate dependent-dioxygenase involved in the biosynthesis of vindoline in Catharanthus roseus (L.) G. Don, *Plant Mol Biol*, **34**, 935-948.

Vetter, H.P., Mangold, U., Schroder, G., Marner, F.J., Werck-Reichhart, D. and Schroder,J., (1992) Molecular Analysis and Heterologous Expression of an Inducible Cytochrome P-450 Protein from Periwinkle (Catharanthus roseus L.), *Plant Physiology*, **100**, 998-1007.

Warzecha, H., Gerasimenko, I., Kutchan, T.M., Stockigt, J., (2000) Molecular cloning and functional bacterial expression of a plant glucosidase specifically involved in alkaloid biosynthesis, *Phytochemistry*, **54**, 657-666.

Weissman, K., Leadlay, P., (2005) Combinatorial biosynthesis of reduced polyketides, *Nature Rev Microbiol*, **3**, 295-236.

Xiao, M, Zhang, Y., Barber, C., Chen, X., Desgagné-Penix, I., Kim,Y.B., Liu, E. Eun-Lee, J., Masada-Atsumi, S., Reed, D., Stout, J.M., Zerbe, P., Zhang, Y., Bohlmann, J., Covello, P.S., De Luca, V., Page, J.E., Ro, D.K., Martin, V.J.J., Facchini P.J., Sensen, C.W. (2013). *J. Biotechnology* (in press)

Zhu, F., Qin, C., Tao, L., Liu, Z., Shi, Z., Ma, X., Jai, J., Tan, Y., Cui, C., Lin, J., Tan, C., Jiang, Y., Chen, Y., (2011) Clustered patterns of species origins of nature-

derived drugs and clues for future bioprospecting, *Proceedings of the National Academy of Science*, **108**, 12943-12948.

CHAPTER 5 – MOLECULAR AND BIOCHEMICAL CHARACTERIZATION OF PICRININE-N-METHYTRALSFERASE, A NOVEL MEMBER OF THE TOCOPERHOL LIKE METHYLTRANSFERASES, FROM *VINCA MINOR* AND *RAUVOLFIA SERPENTINA*

AUTHORS: Dylan Levac, Fang Yu, Vincenzo De Luca

5.1 – INTRODUCTION

The chemical diversity of plants has been the source of a remarkable variety of medicinal products used to treat a range of human diseases. This includes the timely discovery over 50 years ago of the anticancer monoterpenoid indole alkaloid (MIA) dimers, vinblastine and vincristine that are derived by the dimerization of catharanthine and vindoline (Facchini, P., De Luca, V., 2008) and are exclusively found in the leaves of the Madagascar periwinkle (*Catharanthus roseus*). The biosynthesis of *Catharanthus* MIAs has been extensively studied with at least 10 pathway genes for their biosynthesis having been functionally characterized (Facchini, P., De Luca, V., 2008). Recent trancriptomic studies with *C. roseus* have indicated that in addition to their specialization for MIA biosynthesis leaf epidermal cells are also enriched in the pathways for the assembly of triterpenoids, very long chain fatty acids and flavonoids (Murata, J., *et al.* 2008). Other studies have shown that in addition to secretion of antibacterial triterpenes to the leaf surface, *C. roseus* leaves also selectively secrete MIAs such as catharanthine in the waxy cuticle where they may act as herbivory deterrents (Roepke, J., *et al*. 2010).

MIAs that typically occur in the Apocynaceae, Loganiaceae, and Rubiaceae family of plants are derived by the condensation of tryptamine and the 10 carbon iridoid unit secologanin to form the central precursor strictosidine. The conversion of strictosidine to thousands of different MIA, typically involve a series of ring rearrangements and biochemical decorations via the participation of cyclases, hydroxylases, oxidoreductases, glycosyltransferases, acyltransferases, methyltransferases and a few other selected reactions that modify the biological properties of the molecule being produced. For example, ajmaline and ajmalicine from *Rawolfia serpentina* roots are used in the diagnosis of a rare cardiac arrhythmia, brugada syndrome, which primarily affects Asian populations (Wolpert, C., *et al*. 2005) and in the treatment of hypertension respectively. Camptothecin isolated from *Camptotheca acuminata* bark is used to produce the analogues, topotecan and irinotecan to treat various cancers. Vincristine, and vinblastine that are found at very low levels in the *Catharanthus* leaves are key players in combinatorial treatments of childhood lymphomas (Schwartz, C., *et al*. 2009), or individually in refractory anaplastic large-cell lymphomas (Brugières, L., *et al*. 2009). Pierre Fabre and Bristol Meyers have developed semisynthetic derivatives of vinca alkaloids, including the fluorinated molecule vinflunine that is claimed to have improved therapeutic effect against bladder cancer (Bellmunt, J., *et al.* 2009). This particular success has inspired research efforts to identify enzymes capable of accepting halogenated MIAs (Runguphan, W., *et al.* 2010).

In recent years there has been significant progress in identifying new methyltransferase genes and biochemically characterizing their involvement in MIA biosynthesis. Much of this work has focused on identifying early and late biochemical

steps involved in vindoline biosynthesis. Loganic acid *O*-methyltransferase was cloned

based on its amino acid sequence similarity to salicylic acid OMT and biochemical

characterization of the recombinant enzyme confirmed its role in *O*-methylation of the

carboxyl group of loganic acid in the second to last step in secologanin biosynthesis

(Murata, J., *et al*. 2008).   The 16-hydroxytabersonine *O*-methyltransferase involved in

vindoline biosynthesis was biochemically purified, the gene was cloned and its

recombinant gene product was functionally characterized to identify the fifth to last step

in vindoline biosynthesis (Levac, D., *et al*. 2008). Most recently a new class of

methyltransferase, phylogenetically related to γ-tocopherol *C*-methyltransferases was

identified and functionally characterized as a 2,3-dihydrotabersonine *N*-methylransferase

(CrDhtNMT) of (Liscombe, D., *et al*. 2010) (Figure 5-1). Additional *in planta* studies

used VIGS suppressed vindoline accumulation in affected leaves in favor of a novel MIA

with the mass of 16-methoxy-2,3-dihydro-3-hydroxytabersonine (Liscombe, D.,

O'Connor, S., 2011) and made a strong case for the involvement of DhtNMT in the third

to last step in vindoline and vindorosine biosynthesis.

This emergent class of plant tocopherol-like methyltransferases (TLMTs) appears

to preferentially methylate particular MIAs with chemically different nitrogen reaction

centers while maintaining chloroplast targeting features (Dethier, M., De Luca, V., 1993)

typical of tocopherol *C*-methyltransferases. It has been argued that similar but

uncharacterized subcellular targeting, transport and deposition mechanisms, as those

involved in of tocopherol methyltransferases, may also be involved in the transport

TLMTs to chloroplast membranes (Liscombe, D., *et al*. 2010). The present study expands

the biochemical characterization of the *C. roseus* DhtNMT and describes the molecular

cloning and functional characterization of two novel TLMTs involved in the *N*-methylation of picrinine in *Rauvolfia serpentina* and *Vinca minor* (Figure 5-1).

5.2 – MATERIALS AND METHODS

5.2.1 – PLANT MATERIAL

Catharanthus roseus (cv Little Delicata), *Rauwolfia serpentina*, and *Vinca minor* plants were grown in a greenhouse under a long-day (16/8 hour) photoperiod at 30°C. Young leaves (< 1.5 cm in length), whole root, and first leaf pair respectively were harvested for total RNA isolation intended for next generation 454 Roche pyrosequencing. Other tissues to be used for other experiments were also harvested and extracted in the same manner.

5.2.2 – ALKALOID ISOLATION FROM LEAF EXUDATES OF *A. HUBRICHTII*

Leaves from *Amsonia hubrictii* were cultivated in the Niagara Botanical Gardens and were harvested in June, 2010. Fresh leaf material was submerged in chloroform in Erlenmeyer flasks and shaken at 100 RPM for 2 h in order to extract surface MIAs. The chloroform was decanted into a round bottom flask and extracts were taken to dryness by flash evaporation (Buchi Rotavapor R-205, Heating Bath B-490, Brinkmann WLK 230 LAUDA, Thermosavant GP 110 GelPump). The residue was dissolved in methanol and water (85:15 water:methanol) acidified to pH 3 with 10% H_2SO_4. Contaminating small molecules were extracted into ethyl acetate that was discarded and 10 N NaOH was added to to the aqueous phase (pH 12). MIAs were extracted into ethyl acetate and this

fraction was taken to dryness by flash evaporation, dissolved in methanol and stored at -20°C until use.

5.2.3 – TOTAL RNA ISOLATION

Plant tissues were harvested from the greenhouse and homogenized in a mortar and pestle with liquid nitrogen to produce a fine powder. Approximately 0.2 g of powdered homogenate was transferred to 1.5 mL microcentrifuge tubes containing 0.225 mL extraction buffer (0.1 M NaCl, 0.01 M Tris-Cl pH 7.5, 1 mM EDTA, 1% SDS) and 0.15 mL phenol:chloroform:isoamyl alcohol (25:24:1). Samples were then shaken vigorously for 15 minutes at room temperature, and centrifuged (9500 xg, 10 minutes). The aqueous phase was transferred to fresh microcentrifuge tubes and 0.1 volume 3 mM NaOAc pH 5.1, and 2 volumes cold 95% ethanol was added to precipitate nucleic acids over 20 minutes at -80°C. Precipitated nucleic acids were pelleted by centrifugation (9500 g, 15 minutes, 4°C). The supernatant was removed, the nucleic acid pellets were washed with 70% ethanol and after centrifugation (10,000 xg, 10 minutes, 4°C) the ethanol was removed and pellets were dried using a SAVANT DNA120 (Thermo ElectroCompany). Dry pellets were dissolved in 0.5 mL sterile DEPC treated water, and 1 volume 4 M cold LiCl was added with gentle mixing and incubated on ice for 3 hours. Nucleic acids were pelleted by centrifugation (10,000 xg, 10 minutes, 4°C), the supernatant was removed and pellets were dissolved in sterile water. Samples were treated with DNase for 30 minutes at 37°C to remove genomic DNA and purified total RNA was mixed with 0.1 volume 3 mM NaOAc pH 5.1, and 4 volumes 100% ethanol and centrifuged (10,000 xg, 10 minutes, 4°C). The pellet was washed three times with 75% ethanol to remove salts, and stored in 100% ethanol at -80°C until needed.

5.2.4 – DNA SEQUENCING

Databases enriched in sequences encoding genes involved in MIA biosynthesis were generated by extracting RNA isolated from highly active tissues (*C.roseus* young leaves, *R.serpentina* whole root, and *V.minor* 1st leaf pair) and by subjecting them to pyrosequencing on a Genome Sequencer FLX (454 Life Sciences Corp., Bradford, CT) at McGill University Genome Quebec Innovation Centre (Montreal, QC, Canada). Sequence fidelity was verified by traditional Sanger sequencing at Genome Quebec Innovation Center for all PCR products used to establish gene expression profiles or for cloning open reading frames (ORFs) for biochemical characterization of recombinant enzymes.

5.2.5 – DATABASE ASSEMBLY AND TRANSCRIPT ANNOTATION

The data obtained from pyrosequencing of Catharanthus, Rauvolfia and Vinca materials was assembled using to our bioinformatics pipeline as described previously (Facchini, P., *et al.* 2012; Xiao, M., *et al.* 2013). The databases are publicly available and can be found at www.phytometasyn.ca.

5.2.6 – PHYLOGENETIC ANALYSIS OF TLMTs

Amino-Acid alignment of tocopherol and tocopherol-like methyltransferases; Tocopherol and Tocopherol-like peptide amino acid sequences (Table S1) were copied into the Molecular evolutionary Genetics Analysis version 5 application (MEGA 5) integrated tool for phylogenetic analysis (Tamura, K., *et al.* 2011), and peptide sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MuSCLE) (Edgar, R., *et al.* 2004) with the following constraints; Gap Penalties: Open = -2.9,

Extend = 0, Hydrophobicity Multiplier = 1.2. Memory/Iterations: Max Memory in Mb = 4095, Max iterations = 50. Clustering Method UPGMB. Phylogeny construction*;* Tocopherol and tocopherol-like methyltransferase phylogenetic reconstruction was performed using the maximum parsimony statisitical method, and the resulting phylogeny was tested by the Bootstrap method over 10000 independent replications. The substitution model used was Poisson, and rates of substitution were assumed to be uniform. Bootstrap values less than 40 were removed from the resulting phylogeny.

5.2.7 – TLMT cDNA CLONING AND RECOMBINANT PROTEIN EXPRESSION IN *E. COLI*

Full-length cDNA sequences for *CrDhtNMT* (HM584929), *Vm130* (*VmPiNMT;* Accession number KC708450), and *Rs8692* (*RsPiNMT*; Accession number KC708448) were identified in our *C. roseus* cv. Little Delicata, *V. minor*, and *R. serpentina* databases respectively by performing a local BLASTp using the published amino acid sequence of CrDhtNMT (ADP00410) as query.   PCR primers were designed to amplify the known *CrDhtNMT* open reading frame (ORF) (DhtNMT_FW 5' TTCATATGGAAGAGAAGCAGGAG 3', DhtNMT_Rv 5' TTGCGGCCGCATATTGATTTTCGTCCGTAAC 3'), the putative *VmPiNMT* ORF lacking signal peptide, (VmPiNMT_Fw 5' TTCATATGGCGGAAAAGCAAG 3', VmPiNMT_Rv 5' TTGCGGCCGCATTTAGATTTGCGGCATGTAAC 3'), the putative *VmPiNMT* ORF with signal peptide (S_VmPiNMT_Fw 5' TTCATATGTACACTTGTTCAATTATAATATATAT 3', S_VmPiNMT_Rv 5' TTGCGGCCGCATTTAGATTTGCGGCATGTAAC 3') and the putative *RsPiNMT* ORF (RsPiNMT_Fw 5' TTCATATGGCAGAGAAGCAGCAGGC 3', RsPiNMT 5'

TTGCGGCCGCATTTTGATTTCCTGCATGTAATTGCAAC 3'). Note the presence of

5' *NdeI* and 3' *NotI* restriction sites. These restriction sites were incorporated to facilitate

directional cloning into the pET30b (Invitrogen) *E.coli* expression vector. The desired

ORFs were amplified by PCR with pFusion high fidelity DNA polymerase (New England

Biolabs) using the following conditions; 2 minute initial denaturation, 95 $^{\circ}$C; 35 cycles of

(20 second, 95 $^{\circ}$C; 20 seconds, 57 $^{\circ}$C; 60 seconds 72 $^{\circ}$C) followed by a 5 minute final

extension at 72 $^{\circ}$C. PCR products were isolated on 1.2 % agarose gels visualized with

ethidium bromide. PCR products were then excised from the gel and purified using a

liquid nitrogen gel purification technique. Briefly; gel slices were quickly frozen in

liquid nitrogen, and then placed in a 1.5 mL tube which had a hole and glass wool in the

bottom (reservoir tube). The reservoir tubes were then placed on top of a new 1.5 mL

tube (collection tube) and gel slices were thawed at 37 $^{\circ}$C. Once gel slices had thawed

the tubes were centrifuged at max speed for 30 seconds to collect PCR products in the

collection tubes. PCR products were then precipitated and washed with ethanol,

dissolved in MilliQ water, and used directly for A-tailing reactions (The reaction was

conducted with ExTaq DNA polymerase (TaKaRa) under standard conditions in the

presence of dNTPs for 1 hour at 72°C). A-tailed PCR products were used directly for

ligation into pGEM-T Easy T-A cloning vector (Promega). The ligation reaction was

then transformed into XL-10 Gold ultracompotent *E.coli* cells, and plated on LB

ampicillin (100 ug / mL), agar plates with XGAL and IPTG to facilitate blue-white

colony selection after being grown overnight at 37°C. White colonies from each plate

were screened by colony PCR to verify presence of inserts and verified by sequencing

(McGill University Genome Quebec DNA sequencing services, Montreal, Quebec). One

colony was then selected for large scale plasmid purification, *NdeI-NotI* double digestion, and directional cloning into pET30b *E.coli* expression vector. Presence of desired ORF in pET30b, and proper direction was verified by PCR using a combination of gene specific and vector specific PCR primers.

pET30b vectors harboring desired methyltransferase ORFs were then transformed into Bl21 DE3 Codon+ *E.coli* and plated on LB kanamycin (50 μg / mL) agar plates and grown overnight at 37 ^{o}C. The following day single colonies were selected to produce a 5 mL saturated liquid cultures, 500 μl of which was used to inoculate 50 mL liquid LB kanamycin cultures. 50 mL liquid cultures were then grown to OD = 0.6 at 37°C with shaking. At this point recombinant protein expression was induced using 2 mM final concentration IPTG. Proteins were expressed overnight at room temperature (~22°C) with shaking. The following day liquid cultures were transferred to 50 mL conical tubes and centrifuged at 3000 xg to pellet *E.coli* cells and LB media was removed. Cell pellets were then re-suspended in 3 mL 100 mL Tris-HCl pH 7.7, with 14 mM mercaptoethanol. Resuspended cells were lysed by sonication. Lysates were again centrifuged to pellet cell debris, and the supernatant was desalted by PD-10 columns (GE Healthcare) to remove small molecules using. The desalted protein extracts were then used directly for enzyme assays, enzyme kinetics or further protein purification by size exclusion chromatography.

5.2.8 – TLMT EXTRACTION FROM LEAVES

5 grams of first leaf pair *C.roseus, V.minor,* or *R.serpentina* was harvested and homogenized in a cooled mortar and pestle with 15 mL ice cold 100 mM Tris-Cl pH 7.7, 14 mM mercaptoethanol. Leaf homogenates were then filtered into a 15 mL conical tube

with cheese cloth to remove large debris. Filtrates were then centrifuged at 1000 xg to

pellet debris. The resulting supernatant was desalted, and used directly for enzyme

assays or further protein purification by size exclusion chromatography.

5.2.9 – SIZE EXCLUSION CHROMATOGRAPHY

Desalted protein extracts were submitted to gel filtration column chromatography.

Column chromatograpy employed Sephadex G150, 150 ml bed volume, equilibrated in

100 mM Tris–HCl, pH 7.7, 14 mM mercaptoethanol. Flow rate (0.3 ml min^{-1}), back

pressure, fraction collection (5 mL) were all performed using Akta purifier FPLC

controlled by Unicorn 5.0 software package (GE Healthcare). Fractions were assayed for

NMT activity by the standard radioactive enzyme assay.

5.2.10 – *N*-METHYLTRANSFERASE ASSAYS

The standard radioactive enzyme assay (150 µL) contained crude desalted protein

(recombinant or native), 2.5 nCi S-Adenosyl-L-[14C]-methionine (specific activity 58

mCi/mmol; GE Healthcare Canada), and 5 µg substrate (2,3-dihydro-3-

hydroxytabersonine for DhtNMT, or picrinine for PiNMTs), and NMT enzyme assay

buffer (100 mM Tris-HCl, pH 7.7, 14 mM mercaptoethanol). Enzyme assays were

incubated at 37°C for 1 hour. Assays were stopped by basification (10% v/v 10N NaOH),

and reaction products were extracted to 500 µl ethylacetate. 50 µL of the reaction products

was used for quantification by scintillation (Beckman LS 6500 Scintillation Counter).

The standard nonradioactive enzyme assay (200 µL) contained crude desalted

recombinant NMT protein, 2 mM AdoMet, 5 µg substrate, and enzyme assay buffer (100

mM Tris-HCl, pH 7.7, 14 mM mercaptoethanol). Enzyme assays were incubated at 37°C

for 3 hours. Assays were stopped by basification (10% v/v 10N NaOH), and reaction

products were extracted to 500 µl ethylacetate. Organic phases of replicate assays were

combined into a single 2 mL tube, and taken to dryness using a SAVANT Speedvac

(Thermo Scientific). Dried alkaloid products were re-suspended in 200 µL methanol, and

filtered through a 0.22 µm PALL filter (VWR Canada).

Reaction products were analyzed with UPLC (Waters) according to published

methods (Roepke, J., *et al*. 2010). Briefly; The analytes were separated using an Aquity

UPLC BEH C18 column with a particle size of 1.7 µm and column dimensions of 1.0 ×

50 mm. Samples were maintained at 4 °C and 5-µL injections were made into the

column. The analytes were detected by photodiode array and MS. The solvent systems

for alkaloid analysis were as follows: solvent A, methanol: acetonitrile:5-mM ammonium

acetate and 6:14:80; solvent B, methanol: acetonitrile:5-mM ammonium acetate at

25:65:10. The following linear elution gradient was used: 0–0.5 min 99% A, 1% B at 0.3

mL/min; 0.5–0.6 min 99% A, 1% B at 0.4 mL/min; 0.6–7.0 min 1% A, 99% B at 0.4

mL/min; 7.0–8.0 min 1% A, 99% B at 0.4 mL/min; 8.0–8.3 min 99% A, 1% B at 0.4

mL/min; 8.3–8.5 min 99% A, 1% B at 0.3 mL/min; and 8.5–10.0 min 99% A, 1% B at

0.3 mL/min. The mass spectrometer was operated with a capillary voltage of 2.5 kV,

cone voltage of 34 V, cone gas flow of 2 L/h, desolvation gas flow of 460 L/h,

desolvation temperature of 400 °C, and a source temperature of 150 °C.

5.2.11 – SUBSTRATE SPECIFICITY ASSAYS

Desalted crude recombinant protein extracts (rDhtNMT, rPiNMT) were used to

assay pure alkaloid substrates from our collection to identify small molecules accepted by

each NMT. Initial screening used the standard radioactive enzyme assay with an

incubation time of 3 hours. Any substrates that yielded greater than 3x background

radioactive counts were repeated using nonradioactive assay conditions in batches of 3

technical replicates. Replicates of nonradioactive enzyme assay products were extracted

to organic, pooled, dried and prepared for UPLC-MS analysis as described previously.

5.2.12 - SATURATION KINETICS

For *Km* determinations, rVmPiNMT and rRsPiNMT were purified by size

exclusion chromatography. Enzyme assays were performed with 211, 105, 63, 21, 6, 4,

and 2 µM picrinine (AVAChem Scientific, Texas, United States) at constant

concentration of 42 µM (0.025 µCi) (methyl-^{14}C)S-adenosyl-L-methionine, and with 1, 2,

5, 10 and 20 µM (methyl-^{14}C)S-adenosyl-L-methionine at a constant concentration of 0.2

mM picrinine. All assays were performed in triplicate at room temperature for 30 min at

pH 7.7 (100 mM Tris–Cl, 14 mM mercaptoethanol). Kinetic constants were calculated

using GraphPad Prism 5.

5.2.13 – REAL-TIME QUANTITATIVE PCR

Primer design was performed using with the software PRIMER 3 (Whitehead

Institute, MIT, Cambridge, MA, USA) specifying that PCR primers must flank the stop

codon such that the 3' UTR of the transcript assists in increasing the specificity of the

PCR reaction. The primer pairs Actin-F (5'-GGAGCTGAGAGATTCCGTTG-3') and

Actin-R (5'-GAATTCCTGCAGCTTCCATC-3'), CrDhtNMT-F (5'-

CCTTACCCCATCAAGTCGAA-3'), and CrDhtNMT-R (5'-

CCTTACCCCATCAAGTCGAA-3'), VmPiNMT-F (5'- TTCTCTCATGGCTTTGTTCT

-3') and ViPiNMT-R (5'- TTAACGAAATTTTTGGCATT 3'), RsPiNMT-F (5'-

ACATGCAGGAAATCCAAATA -3') and RsPiNMT-R (5'-

ATTCATCAAGAGCTCCACAC -3') (5'-ACAGGGTCAGCCATGGTAAG-3') were used to generate 73 bp, 274 bp and 302 bp length PCR products respectively. The real-time quantitative PCR reaction was carried out in a final 25 μl containing 200 nM of each primer, 12 μl iQ^{TM} SYBR Green PCR Master Mix (Bio-Rad), and 1 μl of cDNA (corresponding to approximately 286 ng cDNA). Real-time PCR conditions were as follow: 95°C for 15 min, then 40 cycles of 95°C for 10 sec, 55°C for 15 sec and 72°C for 30 sec. After all 40 PCR cycles were completed a melting curve was performed to establish if the amplified PCR product was homogeneous. All real-time PCR experiments were run in triplicate for each biological replicate of cDNA produced from total RNA isolated from fresh *C. roseus*, *R. serpentina*, or *V. minor* tissues. PCR was controlled using a C1000 thermocycler (Bio-Rad) and luminescence was detected at the end of each PCR cycle by a CF96 Real-Time System (Bio-Rad). The average threshold cycle (Ct) and relative quantities were calculated using CFX manager version 2.1 (Bio-Rad). b-Actin was used as an internal standard to calculate the relative fold difference based on the comparative Ct method. To determine relative fold differences for each sample in each experiment, the Ct value for the gene was normalized to the Ct value for β-Actin, and was calculated relative to a calibrator (1st leaf pair) using the formula 2^DDCt (Livak, K., Schmittgen, T., 2001).

5.3 – RESULTS

5.3.1 – THE C. *ROSEUS, V. MINOR,* AND *R. SERPENTINA* 454 PYROSEQUENCING DATABASES CONTAIN TOCOPHEROL-LIKE METHYLTRANSFERASE TRANSCRIPTS

Young leaves from *C. roseus* and *V.* mino*r*, as well as roots from *R. serpentina* were extracted for mRNA and samples, were processed for combined 454 and Illumina sequencing and information on these annotated transcriptome databases is available at www.phytometasyn.ca (Facchini, P., *et al.* 2012; Xiao, M., *et al.* 2013). Bioinformatic analysis of each 454 pyrosequencing database revealed 6 tocopherol-like methyltransferases (TLMTs; *Cr91*, *CrDhtNMT*; *Vm130*, *VmPiNMT*; *Rs8692*, *RsPiNMT*) associated with authentic TLMTs and separate from Class I, Class II, and SABATH small molecule MTs (Figure 5-2). In addition to the TLMTs described in Figure 5-2, one, two, and 3 other putative TLMTs could be identified in the *C. roseus*, *V. minor* and *R. serpentina* databases respectively (Table 5-S1).

A Maximum Parsimony phylogeny was constructed from protein sequences of published tocopherol MTs (*AtγTMT*, *HaγTMT*, *BnγTMT*, *Cr1196*) and TLMTs (CrDhtNMT), as well as TLMTs identified in our apocynaceae pyrosequencing databases (*Cr706*, *VmPiNMT*, *Vm265*, *Vm2409*, *Rs820*, *Rs1755*, *RsPiNMT*, *Rs8609*). A Class II methyltransferase (*Cr2551*, Liscombe, D., *et al*. 2010) was used as an outgroup (Table 5-S1). The phylogeny was tested by the bootstrap method over 10000 independent replications, and bootstrap values greater than 40 are identified (Figure 5-2). This phylogeny establishes, with many more TLMT examples, that the TLMT gene family is phylogenetically distinct from γ-tocopopherol C-methyltransferases. Remarkably, *Cr7756*, a previously reported functionally uncharacterized TLMT clone (Liscombe, D., *et al*. 2010), whose sequence was not deposited in NCBI, appears to be phylogenetically closely related to *RsPiNMT* and *VmPiNMT* from our *R. serpentina*, and *V. minor* pyrosequencing databases respectively.

ClustalW pairwise sequence alignments using the *C. roseus* TLMT nucleotide sequences showed that one candidate *Cr91* (365 cluster members) shared 99% sequence identity with the published 16-methoxy-2,3-dihydrotabersonine *N*-methyltransferase (*CrDhtNMT*) gene isolated from *C. roseus* (cv. Little Bright Eyes) (Liscome, D., et al. 2010). *Cr91* has a predicted open reading frame of 870 bp encoding a protein of Mr = 32 kDa, a putative pI = 6, and a single nucleotide difference predicted to result in a substitution mutation (K145 for E145), when compared to the published *CrDhtNMT* sequence. We therefore annotate *Cr91* as *CrDhtNMT*. Other uncharacterized TLMT-like genes (*HM584930*, *Cr7556*; Liscome, D., *et al.* 2010), were not represented in this database. However, Cr7556, which was identified from a *C.roseus* root EST database (Liscome, D., *et al.* 2010) was also present in a previously described *C. roseus* root EST database (Murata, J., *et al.* 2006). The other *Catharanthus* TLMT-like methyltransferase from our pyrosequencing database, *Cr706*, is not represented in the medicinal plant genomic resource database (medicinalplantgenomics.msu.edu).

ClustalW pairwise amino acid sequence alignment (Figure 2) of the VmPiNMT (*Vm130*, 276 cluster members) and RsPiNMT (Rs8692, 339 cluster members) TLMTs revealed that they shared 84% amino acid sequence identity and 92% sequence similarity with each other. The VmPiNMT and RsPiNMT shared only 72% and 71% sequence identity respectively to CrDhtNMT. This greater similarity of VmPiNMT and RsPiNMT to each other raised the possibility that while they were unlikely to possess CrDhtNMT enzyme activity, they might be able to methylate an alternative MIA substrate. The predicted open reading frames of the 969 bp *Vm130* and 873 bp *Rs8692* encode putative proteins of 35 and 32 kDa with pI's of 5.84 and 6.9 respectively.

5.3.2 – ONLY THE *VINCA MINOR* TLMT POSSESSES AN *N*-TERMINAL TRANSIT PEPTIDE

Early investigations into the native CrDhtNMT enzyme employing sucrose density gradient fractionation suggested that it associated with broken and intact chloroplasts, and more specifically, the thylakoid membranes (Dethier, M., De Luca, V., 1993). Release of the CrDhtNMT enzyme from thylakoids using detergents showed that it had an apparent Mr of 60,000 as. CrDhtNMT has a putative subunit Mr of >32,000 that appears to be missing a predicted amino terminal chloroplast transit peptide (Liscombe, D., *et al*. 2010). Together these data suggest that the functional enzyme may be a dimer (Dethier, M., De Luca, V., 1993). Since tocopherol methyltransferases are known to contain *N*-terminal plastid transit peptides, it was suggested that the published sequence for *CrDhtNMT* (Liscombe, D., *et al*. 2010) was incomplete to explain the lack of a predicted amino-terminal transit peptide. While the *Cr91-CrDhtNMT* clone has a virtually identical predicted ORF as the *CrDhtNMT* it has an additional 224 nucleotides upstream of the Kozak consensus sequence that strongly suggest that *CrDhtNMT* does not possess an amino-terminal chloroplast transit peptide. Remarkably, clustalW multiple amino acid sequence alignment of the CrDhtNMT, VmPiNMT, and RsPiNMT revealed highly similar methyltransferase domains absent of gaps (Figure 5-2), WolfP and TargetP 1.1 analysis of each protein predicted that only Vm130 contained a clearly identified 25 amino acid residue secretory transit peptide that may target this enzyme to extracellular spaces (Horton, P., *et al.* 2007).

5.3.3 – THE RECOMBINANT CR91-CrDHTNMT *N*-METHYLATES 2,3-DIHYDRO-3-HYDROXYTABERSONINE AND IS AN AUTHENTIC CrDHTNMT

Biochemical characterization of the native enzyme purified from *C. roseus* leaf chloroplasts identified 2, 3-dihydrotabersonine, and its 3-hydroxy derivative to be substrates for *N*-methylation (Dethier, M., De Luca, V., 1993). Preliminary substrate specificity analysis of the recombinant CrDhtNMT enzyme revealed its preference for 2, 3-dihydro-3-hydroxytabersonine and its pH optimum of 7.5 as previously reported for CrDhtNMT (Liscombe, D., *et al.* 2010) and for the partially purified enzyme from *C. roseus* chloroplasts (Dethier, M., De Luca, V., 1993). While 2,3-dihydrotabersonine does not accumulate in *C. roseus*, 2,3-dihydro-3-hydroxytabersonine is a possible natural intermediate in vindorosine biosynthesis. Enzyme assays using recombinant CrDhtNMT showed that 2,3-dihydrotabersonine, 2,3-dihydro-3-hydroxytabersonine and 2,3,6,7-tetrahydro-3-hydroxytabersonine were also good substrates for this *N*-methylation (Table S2). Analysis of each reaction product by UPLC-MS confirmed that recombinant CrDhtNMT did *N*-methylate each derivative. Previous studies showed that the NMT from chloroplast thylakoids does not accept 2,3,6,7-tetrahydro-3-hydroxytabersonine (Dethier, M., De Luca, V., 1993), whereas detergent solubilized enzyme does accept this substrate for *N*-methylation. An additional 27 MIAs tested (Table 5-S3) were not acceptable substrates for CrDhtNMT (Table 5-S1).

5.3.4 – *R. SERPENTINA* AND *V. MINOR* N-METHYLTRANSFERASES BOTH METHYLATE PICRININE; AN ALKALOID ISOLATED FROM SURFACE EXUDATES OF *A. HUBRICHTII*

To identify substrates for the recombinant VmPiNMT and RsPiNMT enzymes we hypothesized that members of the Apocynaceae whose transcriptomes lack identifiable TLMTs may accumulate MIAs that could serve as substrates for TLMTs. Bioinformatic

analysis of our pyrosequencing databases revealed that *A. hubrichtii* did not appear to express any TLMTs. Alkaloids were harvested from *A. hubrichtii* leaf surface exudates by chloroform dipping and these mixtures were tested using $[^{14}CH_3]$-AdoMet as methyl donor in our standard radioactive methyltransferase assays with recombinant enzyme or with pET 30b empty vector control. At the end of the assay, reactions were stopped by addition of base to the reaction mixture and labeled MIAs were extracted into ethyl acetate. After concentration of MIAs by evaporation, radioactive products were separated by thin layer chromatography to show that both recombinant VmPiNMT (Figure 5-S1, compare lanes 1 and 2) and RsPiNMT (Figure 5-S1, compare lanes 3 and 4) converted a putative unknown MIA from *A. hubrichtii* surface extracts into a major novel methylated product (Rf=5.5). Non-radioactive enzyme assays using AdoMet, *A. hubrichtii* surface alkaloids and recombinant VmPiNMT were repeated on a larger scale to produce sufficient reaction product for UPLC-MS analysis. The appearance of a methylated product eluting at 4.25 min. (Figure 5-S2; VmPiNMT) with the expected mass (M/Z+ 353) coincided with the loss of an MIA eluting at 3.6 min. (M/Z+ 339), while enzyme assays with bacterial extracts expressing the pET 30b empty vector showed no methylation of the 3.6 min. peak (Figure 5-S2; pET30b-EV).

The mass and absorption spectra of the original (Rt 3.6) and novel (Rt 4.25) methylated alkaloid peaks suggested that *Amsonia* surface exudates might contain picrinine that could be converted by these recombinant enzymes to ervincine. Previous phytochemical analyses have identified the presence of picrinine in *Amsonia sinensis* (Liu, H., *et. al.* 1991) and in *Alstonia scholaris* (Chatterjee, A., *et al.* 1965), where it has been documented to have powerful anti-tussive, anti-asthamatic and expectorant

properties (Shangb, J., *et al.* 2010) and in trace amounts in many other MIA producing

species, while ervincine has previously been detected in *Vinca erecta* leaves (Liu, H., *et*

al. 1991). Commercially available picrinine standard was tested to establish the presence

of a picrinine NMT activity in leaf extracts of *Vinca minor* and *Rauvolfia serpentina*

using radioactive enzyme assays conducted as described in Figure 5-S1 (Figure 5-4,

inset). *Vinca minor* leaf protein extracts assayed in the absence picrinine produced 1

major (Rf = 0.98) and several unknown minor radioactive products (with Rfs below 0.3)

(Figure 5-4, inset, Lane 1), while in presence of picrinine two additional products (Rf =

0.55 and 0.45) (Figure 5-4, inset, Lane 2) could be seen. Assays with *Rauvolfia*

serpentina leaf protein extracts displayed similar minor radiolabeled products (with Rfs

below 0.3) in the absence of picrinine (Figure 5-4, inset, Lane 3), while a major product

(Rf = 0.55) was produced in its presence (Figure 5-4, inset, Lane 4). These results

indicate that while these plants may contain internal unidentified substrates (Figure 5-4,

inset, Lanes 1 and 3) that could be methylated by crude leaf protein extracts, both plants

contained NMT activities that could N-methylate picrinine (Figure 5-4, inset, Lanes 2 and

4). When larger scale NMT assays were performed with recombinant VmPiNMT and

RsPiNMT enzymes, they both converted picrinine (Rt = 3.6 min, m/z+ 339) to ervincine

(Rt = 4.25 min, m/z+ 353) (Figure 5-4), compared with extracts expressing empty vector

that produced no reaction product. The *R. serpentina* and *V. minor* TLMTs appear

therfore to be picrinine *N*-methyltransferases (PiNMTs). More detailed substrate

specificity studies using the same 30 MIAs showed that while both recombinant

VmPiNMT and RsPiNMT preferred picrinine as their best MIA substrate, they also

accepted to a lower extent the structurally related 21-hydroxycyclolochnericine (Table 5-

S4) that also features a cyclic ether ring system similar to picrinine in their terpene moieties. Substrate saturation kinetics for both enzymes produced Kms of 6 ± 0.96 µM and 18 ± 4.21 µM for picrinine and 4 ± 0.89 µM and 8 ± 3.86 µM for AdoMet, for RsPiNMT and VmPiNMT respectively (Table 5-S5). To complete the basic biochemical characterizations of the two recombinant PiNMTs, their pH optima were determined to be 7.5 and 7.0, for RsPiNMT and VmPiNMT respectively and the temperature optima was 22°C for both enzymes.

5.3.5 - *N*-METHYLTRANSFERASE ENZYME ACTIVITIES AND mRNAs ARE ENRICHED IN TISSUES ACTIVELY SYNTHESIZING MIAs

Crude whole tissue protein extracts were prepared from different organs of *C. roseus*, *R. serpentina* and *V. minor* as described in methods. Extracts were desalted to remove small molecular weight molecules and were assayed for CrDhtNMT or PiNMT activities using 2,3-dihydrotabersonine or picrinine as substrates, respectively. Inspection of CrDhtNMT reaction products (Figure 5-5 A, C) showed that young *C. roseus* leaf tissues (1st and 2nd leaf pairs) were most active (0.28 - 0.33 $\pm$ 0.01 pmol product mg^{-1}) compared to mature leaves (3rd pair, 0.15 ± 0.009 pmol product/mg/hr), flowers ($0.07 \pm$ 0.030 pmol product/mg/hr), stems (0.14 ± 0.026 pmol product/mg/hr), while no activity could be detected in roots. The RsPiNMT activity was highest in the youngest leaves (0.100 ± 0.013 pmol product/mg/hr) compared to older leaves (2nd leaf pair, 0.06 ± 0.014 pmol product/mg/hr; 3rd leaf pair, 0.04 ± 0.006 pmol product/mg/hr) and to roots ($0.04 \pm$ 0.012 pmol product/mg/hr), while little activity was observed in stems (0.008 ± 0.018 pmol/mg/hr) (Figure 5-5 D, F). Remarkably VmPiNMT activity was also highest in the youngest leaves (1st leaf pair, 0.09 ± 0.012 pmol product/mg/hr) compared to older leaves

(2nd leaf pair, 0.02 ± 0.008 pmol/mg/hr) and to roots (0.01 ± 0.003 pmol/mg/hr), while mature leaves (3rd leaf pair, 0.004 ± 0.02 pmol product/mg/hr) had very low activities (Figure 5-5 G, I).

The levels of *CrDhtNMT*, *RsPiNMT, and VmPiNMT* transcripts were quantified by Real-Time PCR for the same tissue types analyzed for NMT biochemical activities. Total RNA was isolated to prepare cDNA templates from each tissue, and RT-PCR reactions were performed according to experimental procedures. For *C. roseus* the highest *CrDhTNMT* relative transcript abundance (1.96 ± 0.28 DDCt) occurred in the second leaf pair, while first (1.17 ± 0.17 DDCt) and third (0.73 ± 0.29 DDCt) leaf pairs as well as flowers (1.01 ± 0.29 DDCt) also contained relatively lower, but detectable levels of *CrDhTNMT* transcripts. *C. roseus* stems and roots had no detectable *CrDhTNMT* transcript (Figure 5-5 B, C). The highest *RsPiNMT* relative transcript abundance (1.01 ± 0.16 DDCt) occurred in the first leaf pair, followed by second (0.80 ± 0.12 DDCt) and third (0.37 ± 0.20 DDCt) leaf pairs, roots (0.35 ± 0.11 DDCt) and stems (0.11 ± 0.03 DDCt) (Figure 5-5 E, F). Real-time PCR analysis revealed that only the first leaf pair (1.07 ± 0.43 DDCt) had detectable *VmPiNMT* transcripts (Figure 5-5 H, I).

5.3.6 – ONLY THE RECOMBINANT *V. MINOR* PiNMT EXPRESSED WITHOUT A TRANSIT PEPTIDE IS ACTIVE AS A 60 kDa SOLUBLE HOMODIMER WHEN ANALYZED BY SIZE EXCLUSION CHROMATOGRAPHY

Crude enzyme preparations were prepared from young *C. roseus*, *R. serpentina*, *V. minor* leaves, or from bacteria expressing recombinant CrDhtNMT, RsPiNMTand VmPiNMT proteins and each extract was fractionated by gel filtration chromatography

(150 mL bed volume Sephadex G150). The column was calibrated by assaying known *Catharanthus* enzymes [native CrDhtNMT elutes at 50-55 ml in void volume; native 16-hydroxytabersonine O-methyltransferase elutes at 65-75mL (Cr16OMT) as a soluble 80 kDa homodimer; deacetylvindoline acetyltransferase elutes at 80-85mL (CrDAT) as a soluble 50 kDa monomer]. Since proteins larger than 150 kDa appear in the void volume of Sephadex G150 columns, the CrDhtNMT (Figure 6A), RsPiNMT(Figure 6-6B) and VmPiNMT (Figure 5-6C) from each leaf and from recombinant protein extracts all appeared to elute as larger molecular weight aggregates. Approximately 22% of the recombinant VmPiNMT (Figure 5-6C) also eluted as a putative ~60 kDa homodimer which was also observed previously in sucrose gradients with detergent soluble *Catharanthus* chloroplast thylakoid preparations (Dethier, M., De Luca, V., 1993). Remarkably, the proportion of VmPiNMT eluting as a homodimer increased to 47 % of the initial applied activity, when it was expressed without its secretory peptide (Figure 5-6 C). Remarkably, only 7.5 % of the applied activity could be found in the void fraction in these latter preparations.

5.4 - DISCUSSION

A novel class of tocopherol like *N*-methyltransferases involved in MIA biosynthesis that was recently discovered (Liscombe, D., *et al.*, 2010) was used for large scale transcriptome analysis of several MIA producing plant species www.phytometasyn.ca (Facchini, P., *et al*. 2012; Xiao, M., *et al*. 2013). At least 10 new TLMT members of this family were identified whose biochemical functions may be to

catalyze *N*-methylations of different naturally occurring MIAs. The present study

describes the molecular cloning and biochemical identification of 2 TLMTs from *Vinca*

minor (VmPiNMT) and *Rauvolfia serpentina* (RsPiNMT) involved in the *N*-methylation

of picrinine and to a lesser extent 21-cyclolochnericine (Table 5-S4). However,

VmPiNMT and RsPiNMT could not *N*-methylate tabersonine-related substrates that were

acceptable substrates for recombinant *Catharanthus* DhtNMT (Table 5-S2) and they were

unable to use numerous other MIAs that were tested (Table 5-S3).

Previous studies with *Catharanthus* DhTNMT have suggested that this enzyme is

associated with chloroplast membranes and gel filtration experiments with the cell free

enzyme established that in the absence of detergent treatment (Dethier, M., De Luca, V.,

1993) it behaves as a larger molecular weight aggregate rather than as a homodimer. The

three enzymes extracted from leaves, or from bacteria expressing recombinant proteins,

tend to behave as larger molecular weight fractions with a size of at least 150 kDa eluting

in the void volume of Sephadex G150 gel filtration columns (Figure 5-6 A-C). This result

may reflect a tendency of PiNMTs to associate with chloroplast or bacterial membranes

or to form larger multisubunit protein aggregates. Unfortunately, it is not possible to

obtain further information about the size or nature of these PiNMT fractions since they

tend to bind to gel permeation chromatography matrices with larger sieving ranges. The

lack of transit peptides in DhTNMT and in RsPiMT and their tendency to behave as

larger molecular weight complexes (Figure 5-6 A-C) suggest that TLMT sequences

contain features that promote associations with membranes, such as those of chloroplasts,

through hydrophobic interactions or through unknown posttranslational modifications. In

contrast, VmPiMT does have a putative transit peptide and it behaves as a mixed ~60

kDa homodimer, plus higher molecular weight complex, that reverts to a homodimer in the absence of the transit peptide (Figure 5-6C). These results reflect the different properties of DhtNMT and RsPiNMT compared to VmPiNMT that may determine differences in their subcellular location and biological functions in each plant species.

In contrast to DhtNMT, the ability of VmPiNMT and RsPiNMT to *N*-methylate MIAs containing cyclic ether moieties lead to questions of the types of active site amino acid substitutions that would be responsible for this kind of substrate moiety recognition; however there are no suitable published methyltransferase crystal structures for modeling such TLMTs. The production of a soluble rVmPiNMT (Figure 5-6C) could be very useful to produce suitable protein for crystallization and for solving the first three dimensional structure of a TLMT.

The study also presents an interesting strategy for identifying the biochemical functions of different members of the TLMT class of enzymes by harvesting potential MIA substrates from the leaf surfaces of plants that accumulate them and using them for biochemical activity testing (Figure 5-S1). The method depends on sufficient amounts of substrate being available in the extract for conversion into product that can be detected by UPLC MS (Figure 5-S2). While the substrate detected, picrinine, was commercially available for biochemical verification of VmPiNMT and RsPiNMT function, in other cases substrate could be purified from the source tissue for subsequent use in more detailed enzyme characterization of recombinant proteins. This strategy is particularly useful for non-model plants, like *R. serpentina* and *V. minor*, where transient or stable gene knock-down methodologies have not been developed, or where these gene knock-

down strategies do not yield clear target molecules through the backup of precursor due to redirection of carbon flow to other biochemical pathways.

While DhtNMT was previously identified and partially characterized at the molecular level (Liscombe, D., *et al.* 2010), this report provides more complementary biochemical and developmental information about the substrate specificity of DhtNMT (Table 5-S2) and its expression within younger leaves (Figure 5-5A) that is consistent with the stages of development active in MIA biosynthesis and accumulation within *C. roseus* (Facchini, P., De Luca, V., 2008). Noteably, the developmental studies (Figure 5-5) with *C. roseus*, *R. serpentina*, and *V. minor* all suggest that expression of *DhtNMT*, *RsPiNMT* and *VmPiNMT* as well as their biochemical activities are highest in younger leaf tissues. This profile is particularly pronounced in *V. minor* where *VmPiNMT* expression and biochemical activity seems highly confined to the youngest first leaf pair. These profiles suggest that MIA biosynthesis is a very early priority in these plant species and represent a specific investment to protect their young developing tissues against potential disease causing organisms and/or herbivores.

The DhtNMT, RsPiNMT and VmPiNMT enzymes appear to have similar pH optima (pH 7-7.5), while RsPiNMT and VmPiNMT show their highest enzyme activities at 22°C compared that of 30°C for DhtNMT. The RsPiNMT (Km, 4.3 ± 0.9 μM) and VmPiNMT (Km, 8.7 ± 3.8 μM) also show similar affinities, for AdoMet, as those found for DhtNMT (Km, 22 ± 1.8 μM) (Liscombe, D., *et al*. 201). While RsPiNMT and VmPiNMT have a completely different substrate specificity profiles compared those of DhtNMT, their Kms for their respective MIA substrates are very similar. The DhtNMT Km for 2,3-dihydrotabersonine was 8.8 ± 1.0 μM (Liscombe, D., *et al.* 2010), while the

RsPiNMT and VmPiNMT Km for picrinine were 30.7 ± 9.6 µM and 25.7 ± 6.0 µM, respectively. While these similarities suggest that picrinine is likely to be a natural substrate for the RsPiNMT and VmPiNMT enzymes, their biological role in MIA biosynthesis in *R. serpentina* and *V. minor* remains to be established since ervincine has not been detected in alkaloid extracts. It would be of value if transient or stable silencing systems could be developed for *R. serpentina* and for *V. minor* to establish their *in vivo* biochemical roles as shown in *C. roseus* when DhtNMT expression was suppressed by virus induced gene silencing to show its clear involvement in vindoline biosynthesis (Liscombe, D., O'Connor, S., 2011).

While picrinine has been reported to accumulate in trace levels in *R. serpentina* and *V. minor* (Buckingham, J., *et al*. 2010) the only documented biological source of ervincine appears to be from *Vinca erecta* (Buckingham, J., *et al*. 2010). The trace levels of picrinine (Buckingham, J., *et al.* 2010) or the lack of detection of its *N*-methylated product within *R. serpentina* and *V. minor* raises interesting questions since both plants clearly express PiNMT. Perhaps these pathways are silent under the environmental conditions of cultivation and they could be activated with the appropriate environmental cues.

5.5 – ACKNOWLEDGEMENTS

5.6 - ACCESSION NUMBERS

The nucleotide sequences in this paper can be found in the GenBank database under accession numbers KC708448; KC708450; KC708446; KC708452; KC708447; KC708445; KC708453; KC708451

Figure 5-1. Possible biochemical pathway for conversion of Strictosidine to Various *N*-Methylated MIAs in Apocynaceae

The conversion of strictosidine to various *N*-methylated MIAs in *Catharanthus roseus*, *Rauvolfia serpentina* and *Vinca minor* involves members of a newly discovered family of tocopherol-like N-methyltransferase family of enzymes. The reactions yielding ervincine and desacetoxyvindoline are catalyzed by PiNMT and DhtNMT. The reaction

product of the DhtNMT reaction is then converted to vindoline by D4H and DAT. Single

arrows represent single enzymatic steps. Double arrows represent multiple enzymatic

steps. Abbreviations; PiNMT, Picrinine N-methyltransferase, DhtNMT,

Dihydrotabersonine N-methyltransferase, D4H, Desacetoxyvindoline 4-hydroxylase,

DAT, Deacetylvindoline acetyltransferase.

A)

B)

Figure 5-2. Comparative amino acid and phylogenetic analysis of VmPINMT, RsPINMT and CrDhNMT. A) MP phylogeny of tocopherol, and tocopherol-like methyltransferases constructed from peptide sequences listed in Table S1 aligned using MuSCLE. Phylogeny was tested by the bootstrap method over 10000 independent replications, and bootstrap values greater than 40 are identified. Cr2551, a Class II methyltransferase (Liscombe *et al.* 2010) was used as an outgroup. B) ClustalW multiple amino acid sequence alignment of CrDhtNMT, RsPiNMT, and VmPiNMT showing high conservation of the TLMT methyltransferase domain, and a unique N-terminal predicted secretory peptide for only VmPiNMT. Neither RsPiNMT nor CrDhtNMT contain N-terminal targeting peptides.
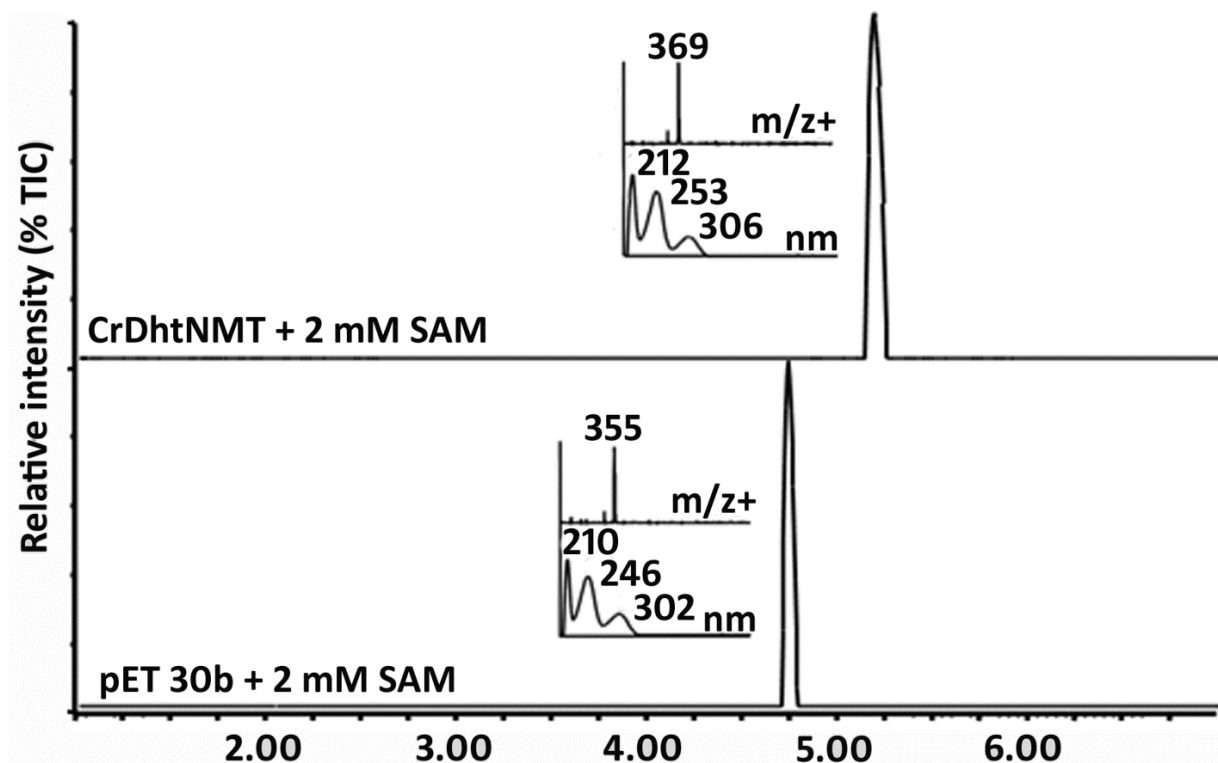
Figure 5-3.**Conversion of dihydro-3-hydroxytabersonine to *N*-methyldihydro-3-hydroxytabersonine by recombinant CrDhtNMT**

Recombinant CrDhtNMT converts 2,3-dihydro-3-hydroxytabersonine (Rt 4.72, m/z+ 355) into 2,3-dihydro-3-hydroxy-N-methyltabersonine (Rt 5.12, m/z+ 369) in the presence of AdoMet. Insets contain the Absorption and MS + spectra of the MIA substrate and product, respectively.
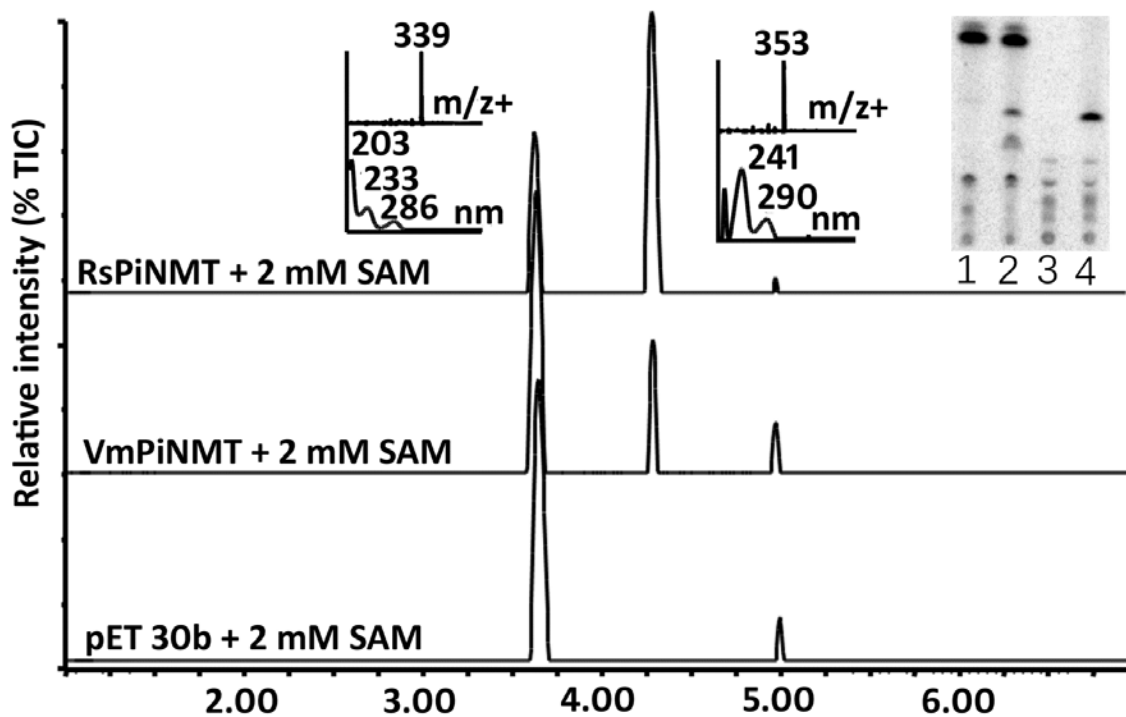
Figure 5-4. Conversion of picrinine to ervincine by recombinant PiNMTs

UPLC chromatograms showing the conversion of picrinine (Rt 3.60, m/z+ 339) into

ervincinine (Rt 4.25, m/z+ 353) by recombinant VmPiNMT or RsPiNMT in the presence

of 2 mM AdoMet. The insets contain the absorption and MS + spectra of MIA substrate

and product as well as an autoradiograms of reaction products separated by thin layer

chromatography in radioactive enzyme assays using plant leaf extracts, picrinine

standard and AdoMet. *Vinca minor* leaf protein extracts were assayed in the absence

(Lane 1) and presence (Lane 2) of picrinine. *Rauvolfia serpentina* leaf protein extracts

were assayed in the absence (Lane 3) and presence (Lane 4) of picrinine.
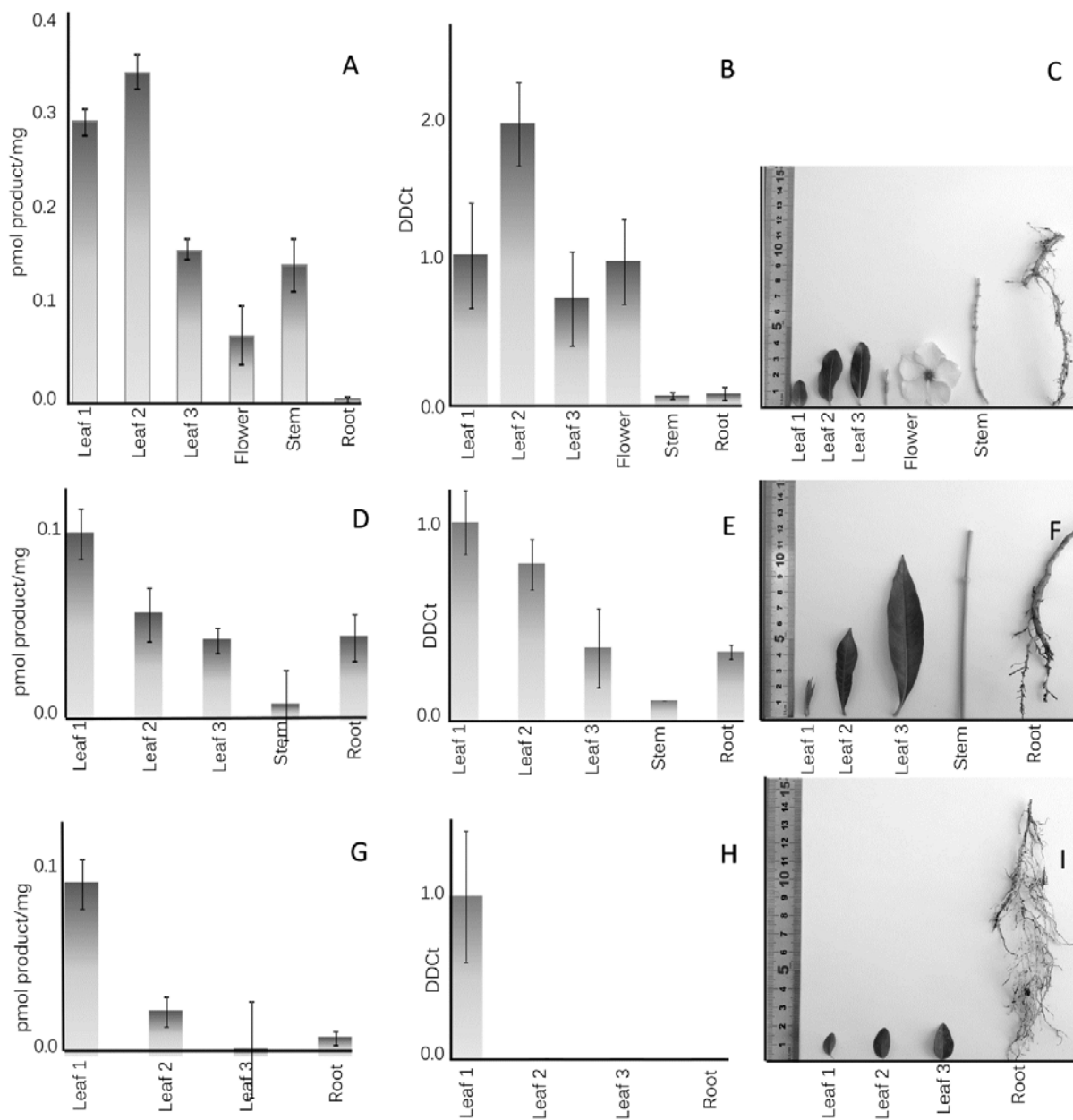
Figure 5-6. TLMT biochemical and transcript levels are coordinated in different plant tissues

CrDhtNMT, VmPiNMT and RsPiNMT enzyme activities are coordinated to different extents with gene expression in different tissues of *C. roseus*, *V. minor* and *R. serpentina*. A) Enzyme activity profile and B) Relative abundance of *CrDhtNMT* transcripts in C) representative tissues of *C. roseus*. D) Enzyme activity profile and E) Relative abundance of *CrDhtNMT* transcripts in F) Representative tissues of *R. serpentina*. G) Enzyme activity profile and H) Relative abundance of *CrDhtNMT* transcripts in I) Representative tissues of *V. minor*.
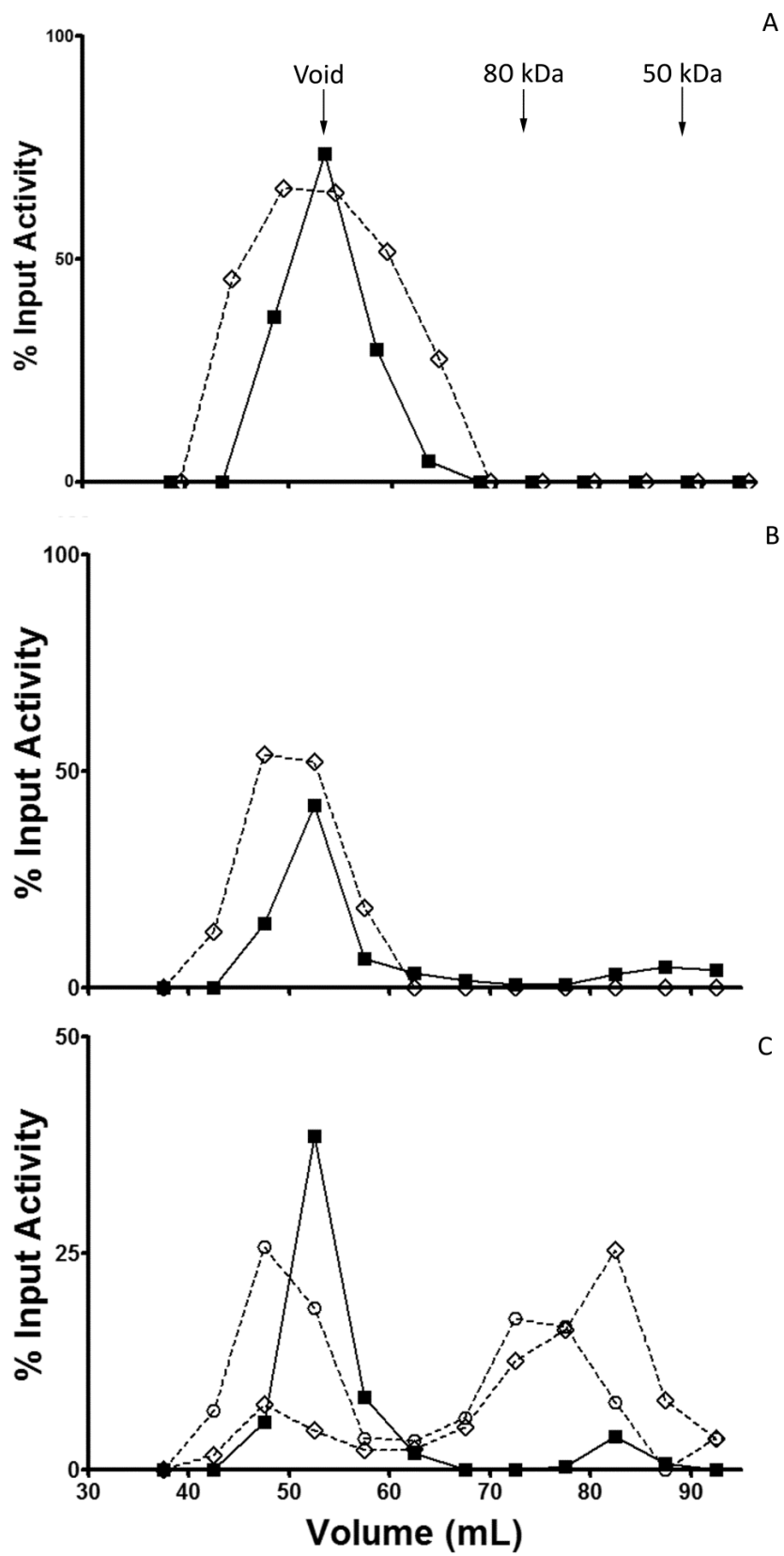
Figure 5-6. Gel filtration of native and recombinant TLMTs

Sephadex G 150 gel permeation chromatography of native and recombinant CrDhtNMT, VmPiNMT and RsPiNMT. Native (■) and recombinant (◊) CrDhtNMT **(A)** and RsPiNMT **(B)** enzyme activities elute in the column void suggesting that functional enzymes are aggregates larger than 150 kDa. Native (■) VmPiNMT **(C)** also elutes in the column void while small amounts of recombinant VmPiNMT also elutes as an apparent 60 kDa soluble homodimer. Expression of recombinant VmPiNMT without its *N*-terminal sectetory peptide (○) allows most of the protein to behave as a homodimer.

1   2   3   4
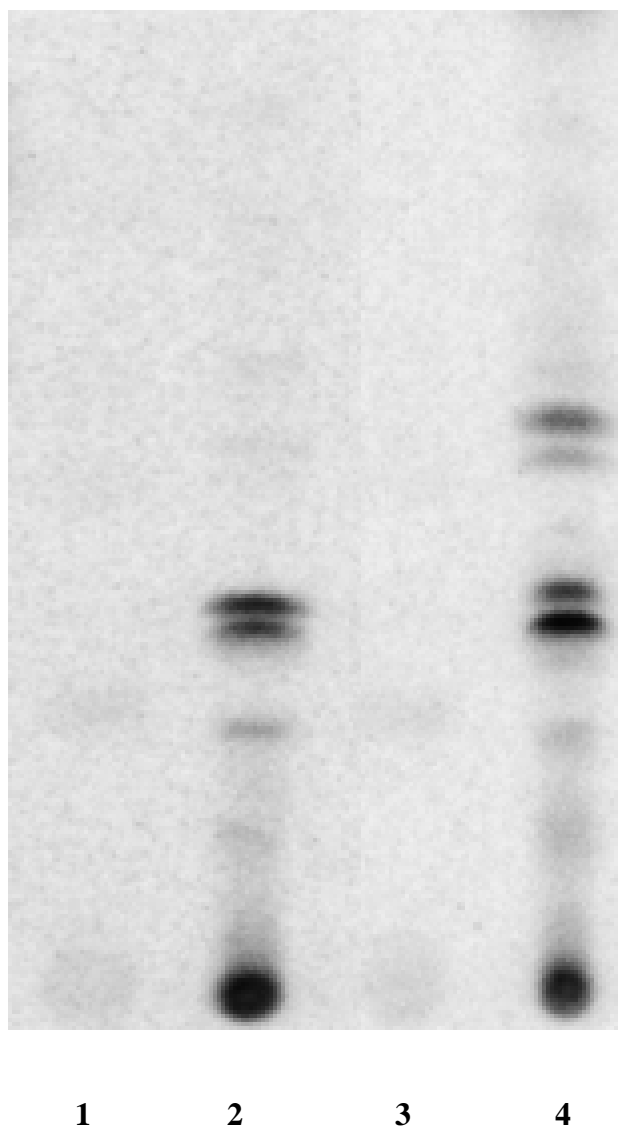
Figure 5-S1. TLC of radioactive reaction products from recombinant enzyme assays using *A. hubrichtii* surface alkaloids as substrate

TLC chromatography and audioradiogrphy of reaction products in enzyme assays using recombinant VmPiNMT and RsPiNMT enzymes incubated with $^{14}CH_3$-AdoMet as methyl donor and MIAs obtained from the leaf surfaces of *Amsonia hubrichtii*. Incubations were conducted with Empty pET 30b vector (Lane 1), VmPiNMT (Lane 2), Empty pET 30b vector (Lane 3), RsPiNMT.Buffer (Lane 4)
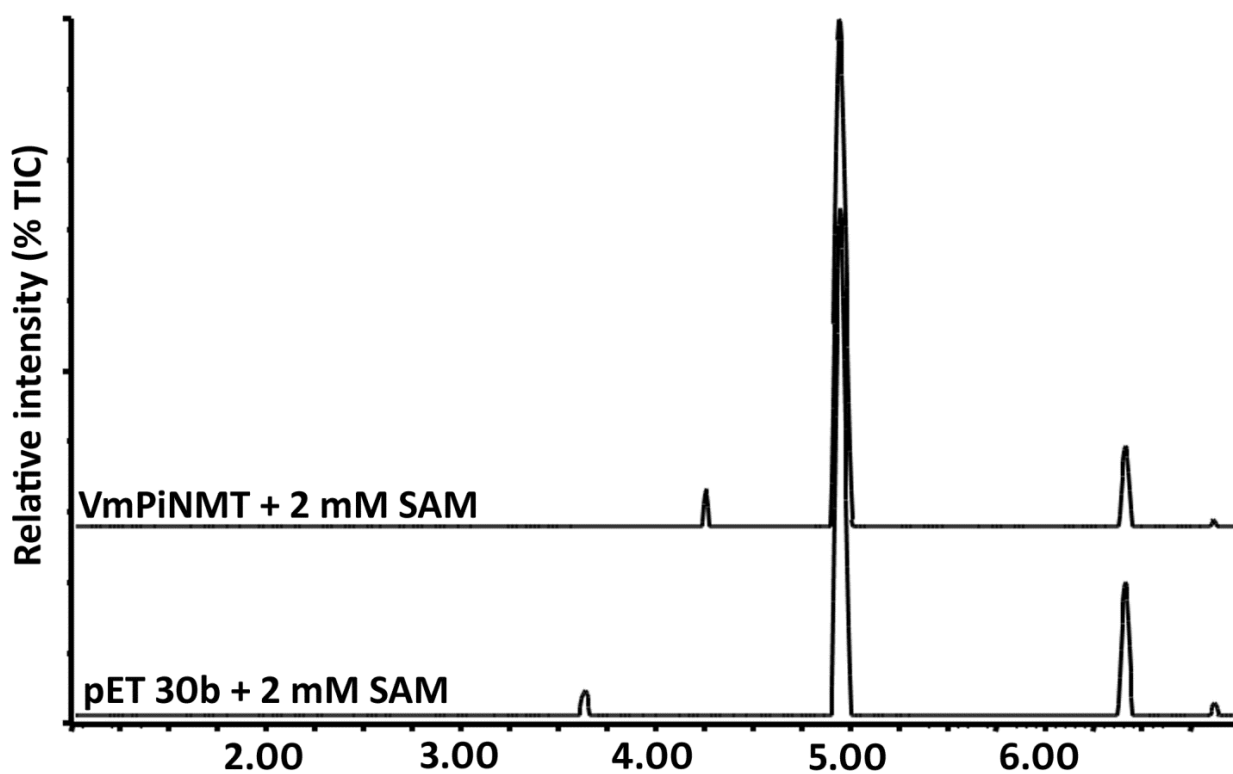
Figure 5-S2. **UPLC analysis of cold assay of Figure 5-S1**

Identification by UPLC-MS of the *A. hubrichtii* leaf surface MIA substrate being *N*-methylated inVmPiNMT assays in Figure S1. Enzyme assays conducted in the presence of pET 30b empty vector display an MIA peak (Rt 3.62 m, m/z+ 339) that shifts to Rt 4.93 m, m/z + 353) in assays conducted with VmPiNMT.

Table 5-S1: Abbreviations, Annotatins and Accession numbers for sequences in MP tree

Abbreviation	Annotation [organism]	Accession #
CrDhtNMT	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	HM584929
Cr7756	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	Liscombe, D., *et al.* 2010
RsPiNMT	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708448
VmPiNMT	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708450
Rs1755	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708446
Vm2409	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708452
Rs8609	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708447
Rs820	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708445
Cr706	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708453
Vm265	16-methoxy-2,3-dihydrotabersonine N-methyltransferase [Catharanthus roseus]	KC708451
Cr1196	gamma-tocopherol methyltransferase-related protein [Catharanthus roseus]	HM584930
HaGTMT	Gamma-tocopherol methyltransferase [*Helianthus annuus*]	ABB52798
AtGTMT	Gamma-tocopherol methyltransferase [*Arabidopsis thaliana*]	AAD02882
BnGTMT	Gamma-tocopherol methyltransferase [*Brassica napus*]	ACJ54674
Cr2551	Putative methyltransferase [*Catharanthus roseus*]	HM584931

Table 5-S2: Substrate specificity of the recombinant CrDhtNMT enzyme. Performed according to the standard radioactive methyltransferase assay. 100% activity = 0.093 pmol product/mg protein/ hour
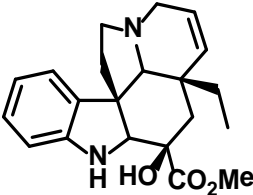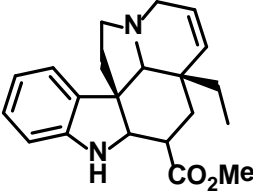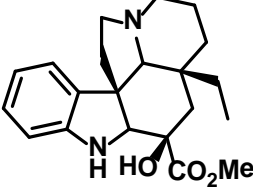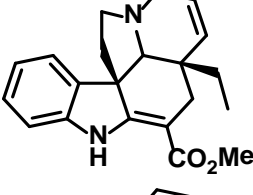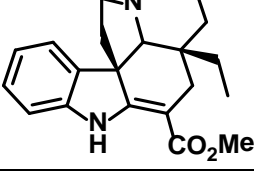
Alkaloid	Structure	Relative activity
2,3-Dihydro-3-hydroxytabersonine		100%
2,3-Dihydrotabersonine		43%
2,3,6,7-Tetrahydro-3-hydroxytabersonine		93%
Tabersonine		0
Vincadifformine		0

Table 5-S3: List of alkaloid substrates used in substrate specificity studies.
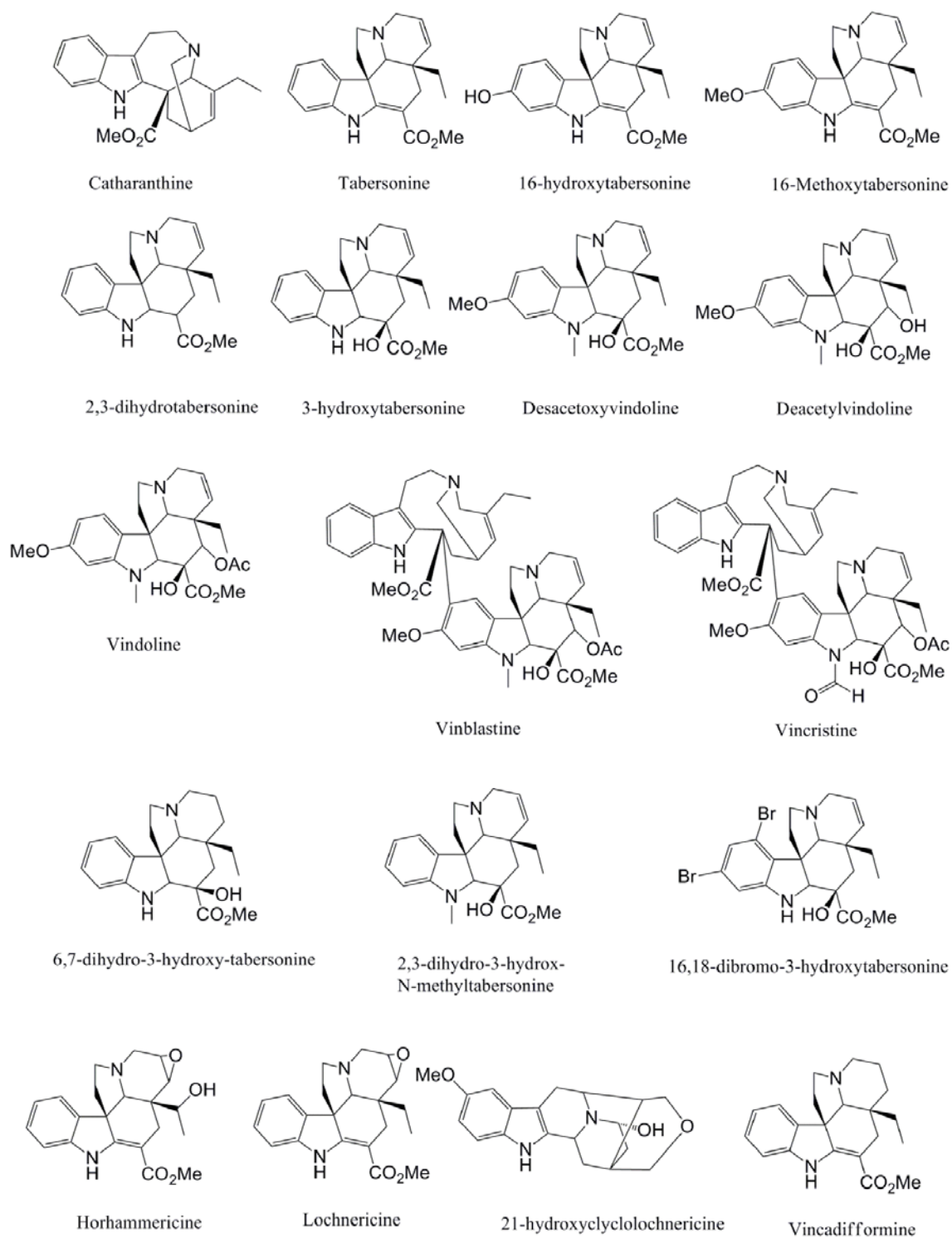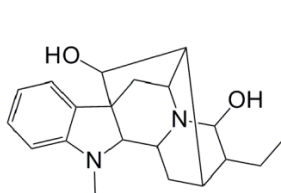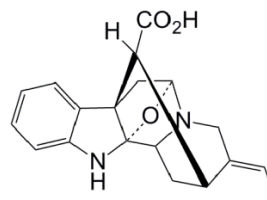
Catharanthine

Tabersonine

16-hydroxytabersonine

16-Methoxytabersonine

2,3-dihydrotabersonine

3-hydroxytabersonine

Desacetoxyvindoline

Deacetylvindoline

Vindoline

Vinblastine

Vincristine

6,7-dihydro-3-hydroxy-tabersonine

2,3-dihydro-3-hydrox-N-methyltabersonine

16,18-dibromo-3-hydroxytabersonine

Horhammericine

Lochnericine

21-hydroxyclyclolochnericine

Vincadifformine

Ajmaline

Vincamine

Picrinine
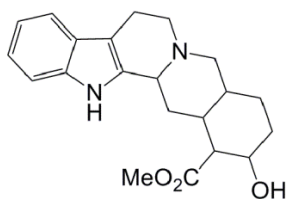
Yohimbine
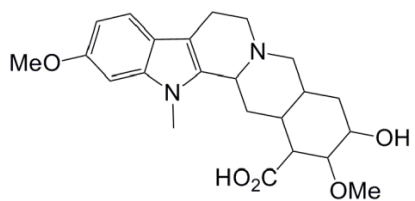
Reserpinic acid
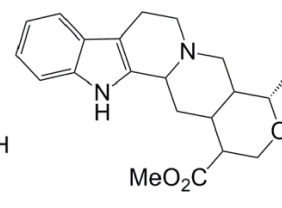
Ajmalicine
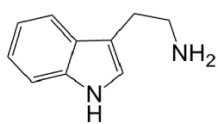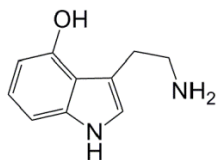
Tryptamine

5-Hydroxytryptamine
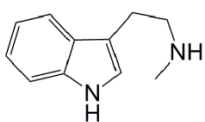
N-methyltryptamine
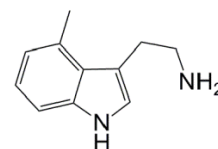
5-methyltryptamine

Table 5-S4: Enzymatic incorporation of C14-AdoMet in *A. hubrichtii* surface alkaloids by recombinant PiNMTs. 10% radiolabel incorporation represents the production of 42 pmol product, 14% radiolabel incorporation represents the production of 61 pmol product.

Surface Extract	Radiolable Incorporation	
	VmPiNMT	RsPiNMT
A. hubrichtii	10%	14%

Table 5-S5: Substrate specificity of recombinant PiNMTs. performed according to the standard radioactive methyltransferase assay. RsPiNMT 100% activity = 0.0018 pmol product/mg protein/hour, VmPiNMT 100% activity = 0.045 pmol product/mg protein/hour
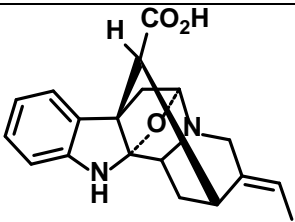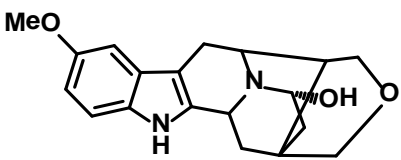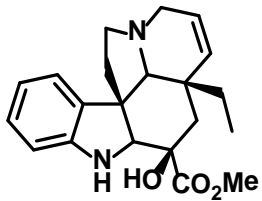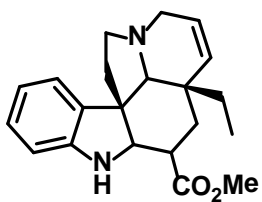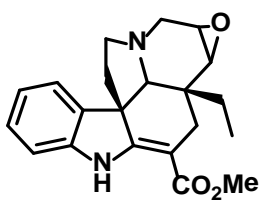
Alkaloid	Structure	Relative activity	
		VmPiNMT	RsPiNMT
Picrinine		100%	100%
21-hydroxylochnericine		0.8%	8%
2,3-dihydro-3-hydroxytabersonine		0%	0%
2,3-dihydrotabersonine		0%	0%
Lochnericine		0%	0%

Table 5-S6: Saturation kinetic parameters for RsPiNMT and VmPiNMT recombinant enzymes at saturating SAM concentrations

Picrinine	Km (μM)	Vmax (fmol/s/mg)	Kcat (s^{-1})	Kcat/Km (S^{-1} M^{-1})
RsPiNMT	30.7 $\pm$ 9.6	490.5 $\pm$ 60	0.11 $\pm$ 0.01	3 548
VmPiNMT	25.7 $\pm$ 6.0	210.0 $\pm$ 19	0.77 $\pm$ 0.07	30 800

Table 5-S7: Saturation kinetic parameters for RsPiNMT and VmPiNMT recombinant enzymes at saturating picrinine concentrations.

AdoMet	Km (μM)	Vmax (fmol/s/mg)	Kcat (s^{-1})	Kcat/Km (S^{-1} M^{-1})
RsPiNMT	4.3 $\pm$ 0.9	58.2 $\pm$ 4.3	0.013 $\pm$ 0.001	3 250
VmPiNMT	8.7 $\pm$ 3.8	116.8 $\pm$ 23.4	0.43 $\pm$ 0.09	53 750

Table 5-S8: Oligonucleotide primers used in this study

Name	5'-3' Sequence	Purpose
NdeI-Cr91-F	TTCATATGGAAGAGAAGCAGGAG	CrDhtNMT ORF cloning
Cr91-NotI-R	TTGCGGCCGCATATTGATTTTCGTCCGTAAC	
NdeI-Rs8692-F	TTCATATGGCAGAGAAGCAGCAGGC	RsPiNMT ORF cloning
Rs8692-NotI-R	TTGCGGCCGCATTTTGATTTCCTGCATGTAATTGCAAC	
NdeI-Vm130-F	TTCATATGGCGGAAAAGCAAG	VmPiNMT methyltransferase domain cloning
Vm130-NotI-R	TTGCGGCCGCATTTAGATTTGCGGCATGTAAC	
NdeI-SVm130-F	TTCATATGTACACTTGTTCAATTATAATATATAT	VmPiNMT ORF cloning
SVm130-NotI-R	TTGCGGCCGCATTTAGATTTGCGGCATGTAAC	
Cr91-RT-F	GGTGCTGTATCACTGTGTTCT	Real-Time PCR
Cr91-RT-R	ATTTTCGTCCCGTAACTACA	
Rs8692RT-F	ACATGCAGGAAATCCAAATA	Real-Time PCR
Rs8692RT-R	ATTCATCAAGAGCTCCACAC	
Vm130RT-F	TTCTCTCATGGCTTTGTTCT	Real-Time PCR
Vm130RT-R	TTAACGAAATTTTTGGCATT	

5.8 – BIBLIOGRAPHY

Bellmunt, J., Théodore, C., Demkov, T., Komyakov, B., Sengelov, L., Daugaard, G., Caty, A., Carles, J., Jagiello-Gruszfeld, A.,Karyakin, O., Delgado, F., Hurteloup, P., Winquist, E., Morsli, N., Salhi, Y., Culine, S., von der Maase, H., (2009) Phase III Trial of Vinflunine Plus Best Supportive Care Compared With Best Supportive Care Alone After a Platinum-Containing Regimen in Patients With Advanced Transitional Cell Carcinoma of the Urothelial Tract, *Journal of Clinical Oncology*, **27**, 4454-4461.

Brugières, L., Pacquement, H., Le Deley, M., Leverger, G., Lutz, P., Paillard, C., Baruchel, A., Frappaz, D., Nelken, B., Lamant, L., Patte, C., (2009) Single-Drug Vinblastine As Salvage Treatment for Refractory or Relapsed Anaplastic Large-Cell Lymphoma: A Report From the French Society of Pediatric Oncology, *Journal of Clinical Oncology*, **27**, 5056-5061

Buckingham, J., Baggaley, K., Roberts, A., Szabo, L., (2010) *Dictionary of Natural Products*, CRC Press Taylor & Francis Group, Boca Raton, FL

Chatterjee, A., Mukherjee, B., Ray, A.B., Das, B., (1965). Alkaloid from leaves of Alstonia scholaris. *Tetrahedron Letters* **41**, 3633–3637.

Dethier, M., De Luca, V., (1993) Partial purification of an N-methyltransferase involved in vindoline biosynthesis in *Catharanthus roseus*, *Phytochemistry,* **32**, 673-678

Edgar, Robert C. (2004), MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Research,* **32**, 1792-97.

Facchini, P., De Luca, V., (2008) Opium poppy and Madagascar periwinkle: model non-model systems to investigate alkaloid biosynthesis in plants, *The Plant Journal*, **54**, 763-784.

Facchini, P.J., Bohlmann, J., Covello, P.S., De Luca, V., Mahadevan, R., Page, J.E., Ro, D.K., Sensen, C.W., Storms, R., Martin, V.J.J. (2012) Synthetic biosystems for the production of high-value plant metabolites. *Trends in Biotechnology*, **30**, 127-131.

Horton, P., Park, K., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J. , Nakai, K., (2007) WoLF PSORT: Protein Localization Predictor, *Nucleic Acids Research*, **259**, 585-587

Levac, D., Murata, J., Kim, W., De Luca, V., (2008) Application of carborundum abrasion for investigation of leaf epidermis: molecular cloning of Catharanthus roseus 16-hydroxytabersonsine-16-O-methyltransferase, *The Plant Journal*, **53**, 225-236

Liscombe, D., O'Connor, S., (2011) A virus induced gene silencing approach to understanding alkaloid metabolism in *Catharanthus roseus*, *Phytochemistry*, **72**, 1969-1977

Liscombe, D., Usera, A., O'Connor, S., (2010) Homolg of tocopherol *C* methyltransferases catalyzes *N* methylation in anticancer alkaloid biosynthesis, *Proceedings of the National Academy of Science*, **107**, 18793-18798

Liu, H., Wu, B., Zheng, Q., Feng, X., (1991) New indole alkaloids from Amsonia sinensis, *Planta Medica*, **57**, 566-568.

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

Murata, J., Bienzle, D, Brandle, J.E., Sensen, C.W. and De Luca, V. (2006) Expressed sequence tags from Madagascar periwinkle (Catharanthus roseus), *FEBS Letters*, **580**, 4501-4507

Murata, J., Ropeke, J., Gordon, H., De Luca, V., (2008) The Leaf Epidermome of *Catharanthus roseus* Reveals Its Biochemical Specialization, *The Plant Cell*, **20**, 524-542

Rakhimov, D.A., Malikov, V.M., Yunusov, S.Y. (1967) Isolation of Kopsinilam and ervincine *Khimiya Prirodnykh Soedinenii*, **3**, 354-355.

Roepke, J., Salim, V., Wu, M., Thamm, A., Murata, J., Ploss, K., Boland, W., De Luca, V., (2010) Vinca drug components accumulate exclusively in leaf exudates of Madagascar perisinkle, *Proceedings of the National Academy of Science*, **107**, 15287-15292.

Runguphan, W., Qu, X., O'Connor, S., (2010) Integrating carbon–halogen bond formation into medicinal plant metabolism, *Nature*, **468,** 461-464

Schwartz, C., Constine, L., Villaluna, D., London, W., Hutchison, R., Sposto, R., Lipshultz, S., Turner, C., deAlarcon, P., Chauvenet, A., (2009) A risk-adapted, response-based approach using ABVE-PC for children and adolescents with intermediate- and high-risk Hodgkin lymphoma: the results of P9425, *Blood,* **11**, 2051-2059

Shangb, J., Caia, X., Fenga, T., Zhaob, Y., Wangb, J., Zhangc, L., Yanc, M., Luoa, D., (2010) Pharmacological evaluation of Alstonia scholaris: Anti-inflammatory and analgesic effects, *Journal of Ethnopharmacology*, **129**, 174–181

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., (2011) MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods, *Molecular Biology and Evolution,* **28**, 2731-2739

Wolpert, C., Echternach, C., Veltmann, C., Antzelevitch, C., Thomas, G.P., Spehl, S., Streitner, F., Kuschyk, J., Schimpf, R., Haase, K.K., Borggrefe, M., (2005) Intravenous drug challenge using flecainide and ajmaline in patients with Brugada syndrome, *Heart Rythm*, **2**, 254-260

Xiao, M, Zhang, Y., Barber, C., Chen, X., Desgagné-Penix, I., Kim,Y.B., Liu, E. Eun-Lee, J., Masada-Atsumi, S., Reed, D., Stout, J.M., Zerbe, P., Zhang, Y., Bohlmann, J., Covello, P.S., De Luca, V., Page, J.E., Ro, D.K., Martin, V.J.J., Facchini P.J., Sensen, C.W. (2013). J. Biotechnology (in press)

CHAPTER 6 – GENERAL DISCUSSION AND CONCLUSIONS

While I believe some doctoral students may attempt to spin their contribution as groundbreaking, I will not, however. This thesis has documented the development of biochemical and molecular tools, as well as phylogenetic approaches, for elucidating biochemical pathways and/or interesting gene families. What has been groundbreaking is the era during which this work was performed. Over the last five years there has been a major paradigm shift in how we approach elucidating biochemical pathways in non-model medicinal plants. Where, in the past, we would purify and sequence proteins, design and clone genes based on consensus sequences, today we do not even have to make cDNA libraries to perform Rapid Amplification of cDNA Ends (RACE). Present day transcriptome databases, for the most part, reveal full open reading frames for nearly every transcript of interest, even relatively low abundance transcripts. This is the real breakthrough and I had no contribution in the development of that. This has come about because sequencing costs are at a point where academic labs can now perform large scale transcriptome sequencing on non-medical research budgets.

My studies have documented, through using a novel class of indole *N*-methyltransferase which happen to be involved in MIA biosynthesis, how these biochemical, molecular and phylogenetic tools can be used in concert to quickly and effectively, move from small molecule of interest, to gene candidates, through the functional characterization of novel enzymes and further characterization of the function of rest of the gene family. I hope that, some day, individuals engaged in combinatorial biosynthesis will use these techniques, or possibly even some of the enzymes that I have

functionally characterized, to generate improved therapeutics. This direct application of

knowledge for the improvement of people's quality of life has always been my passion.

This may explain at least partly why I was, and probably always will be, a medical school

hopeful.

Improving the probability of task success, at every major junction in this program,

has been central to my success as a Ph.D. student. Where, in the past, we would clone a

single gene and try to functionally characterize it before we would move on and clone the

next candidate gene, I decided to clone all TLMTs at once, and functionally characterize

them all at once. Furthermore, I was not focused on a single, particular reaction but

rather a reaction type. This loosening of target parameters opened up opportunities for

the discovery of completely unknown and uncharacterized biochemical reactions, as

demonstrated by the identification of PeNMT, AjNMT, and PiNMTs. Suggestions by Dr.

Peter Facchini and Dr. De Luca to explore closely related, but non-vinceae plant species,

and the surface secretions of these plants, lead to the development of surface extracts

from a number of Apocynaceae as mixed alkaloid substrate samples. In essence, when

these samples are used in enzyme assays they are natural, high throughput, combinatorial

substrate screens. These high throughput substrate screens are really what broke the

traditional bottleneck in the functional characterization of enzymes involved in

specialized metabolism. Where, in the past, we would screen individual compound

standards, and be limited to compound availability in our local libraries, or commercial

sources at significant costs, these mixed substrate samples permitted access to many

small molecules at the cost of solvent and the value of a graduate student's time. I expect

that the approaches documented in this thesis will greatly improve the rate at which Paulo E. Cázares will characterize the remaining TLMTs.

TLMTs, like γ-tocopherol *C*-methyltransferases, appear to associate with membranes, and are difficult to work with in structural studies involving protein crystallization. The TLMT family of *N*-methyltransferases present a remarkable opportunity to study protein structure. It is exceedingly rare to find two homologues like the RsPiNMT, and the VmPiNMT that are seemingly identical, functionally speaking, and yet one can be recombinantly expressed as an active, soluble homodimer, while the other cannot. The VmPiNMT should be a fantastic candidate to generate crystals of TLMTs to understand the topology, and composition of its active site, as well as to generate insights in to the reaction mechanism involved in indole *N*-methylation. Furthermore, one could imagine conducting domain swapping experiments between the RsPiNMT and VmPiNMTS as a strategy to first understand the domains and peptide residues involved in the apparent aggregation/membrane associations of TLMTs. From there this knowledge could be used to further understand the mechanisms of association of not only TLMTs but also tocopherol *C*-methyltransferases, which are far more ubiquitous. Finally, it would be interesting to express RsPiNMT, VmPiNMT with and without its transit peptide and its transit peptide alone, fused with green fluorescent protein in order to understand the subcellular localization, and trafficking of TLMTs, and tocopherol *C*-methyltransferases.

APPENDIX

A.1 - TANDEM MASS SPECTROSCOPY OF ALKALOID SUBSTRATES AND METHYLTRANSFERASE PRODUCTS OF TLMT CELL FREE ASSAYS

METHODS

Cold, cell free enzyme assays were performed as previously described in chapters 4 and 5 to produce methylated ajmaline, perivine, and picrinine products for tandem MS/MS analysis. All samples were run using a Bruker HCT Ultra LC/MS^n system with a high capacity ion trap mass analyzer. Sample infusion was executed via a PEEK capillary from a Cole Parmer Syringe Pump and a Hamilton 250 uL syringe. Appropriate dilutions of samples and standard solutions were prepared in HPLC grade Methanol and filtered through 0.2u PTFE syringe filters. The filtered solutions were infused by syringe pump to the nebulizer of the LC/MS system at a flow rate of 10 uL/min. For positive ion detection, the major parameters were set as follows: Capillary extraction voltage -4000V, Nebulizer (Nitrogen) 16.0 psi, Dry Gas (Nitrogen) 6.0 L/min. Drying Temperature , 300 C. For normal MS spectra, approximately 50 scans of sample data were averaged. For MS2, MS3 or MS4 spectra, usually 5-10 scans of sample data were averaged.
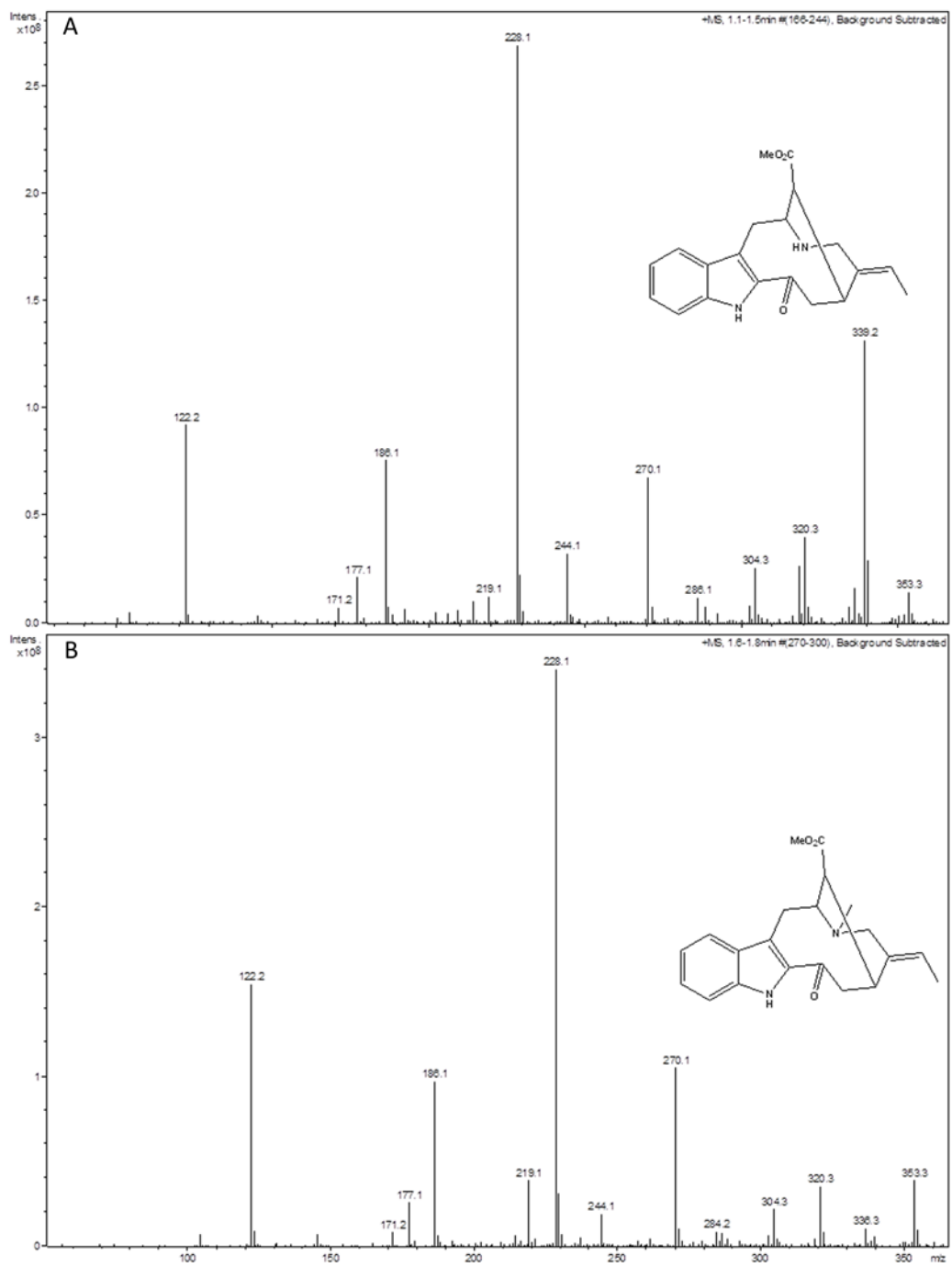
Figure A-1. MS1 analysis of alkaloid products of cell free pET 30b-EV (A) and recombinant CrPeNMT (B) methyltransferase assay productssupplied with perivine standard as substrate.
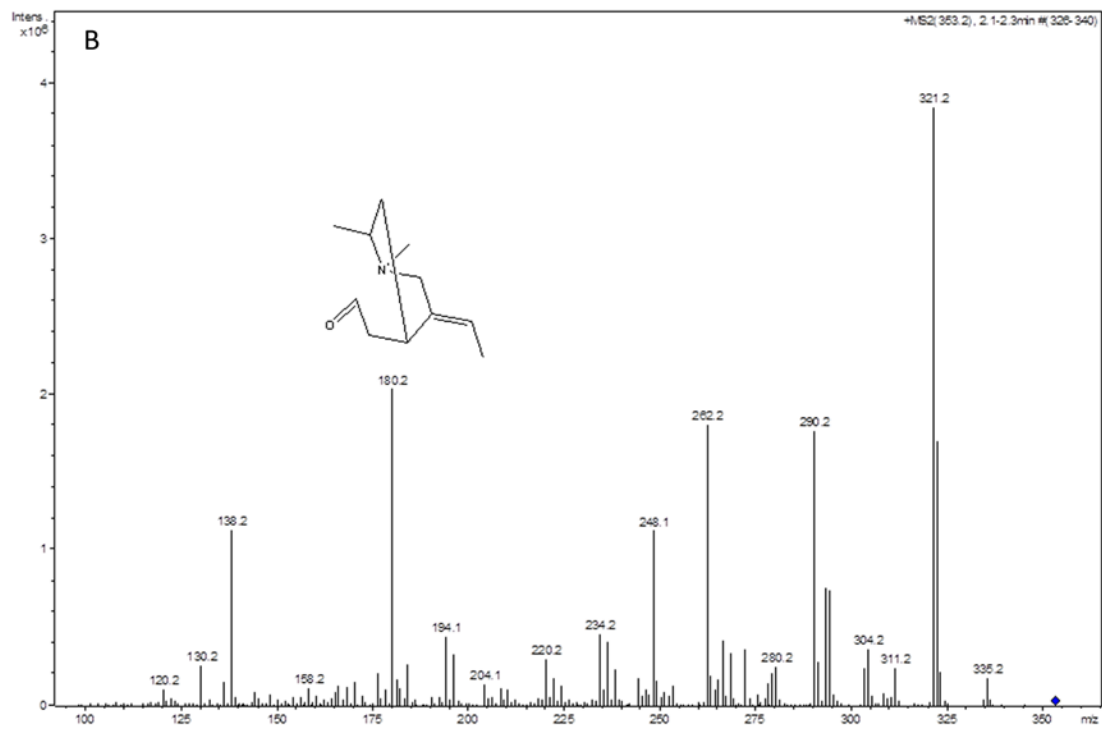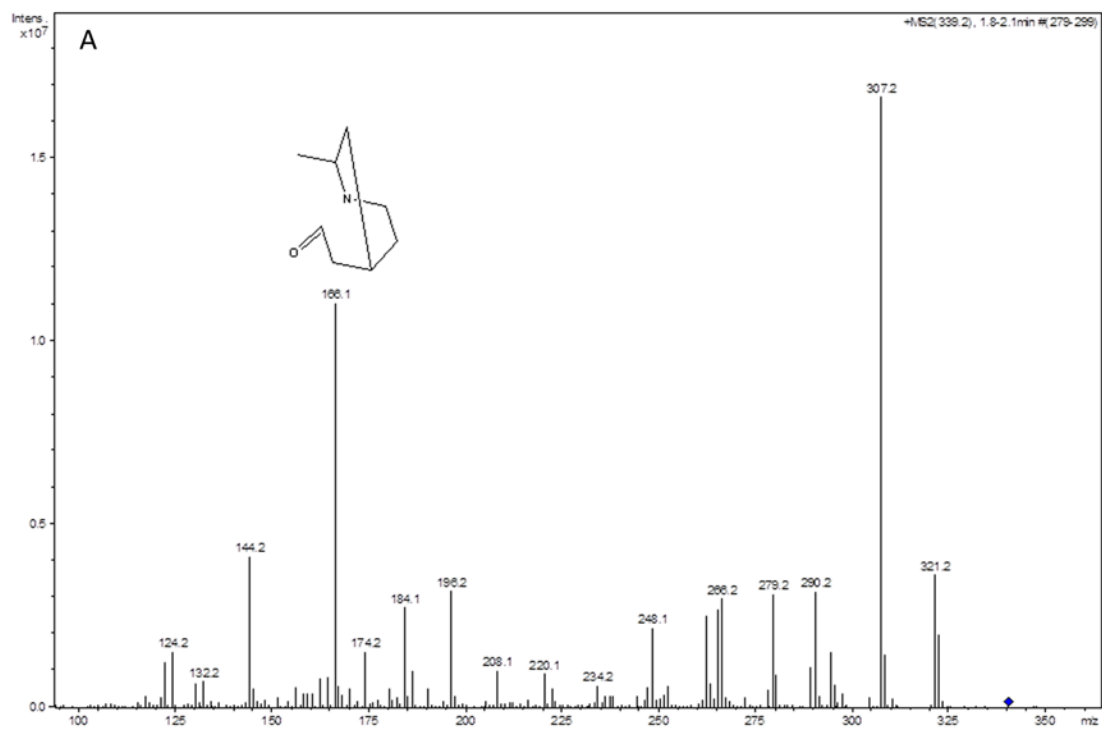
Figure A-2. MS2 analysis of 339 m/z +1 alkaloid of cell free pET 30b-EV (A) and 353

m/z +1 alkaloid of recombinant CrPeNMT (B) methyltransferase assayssupplied with
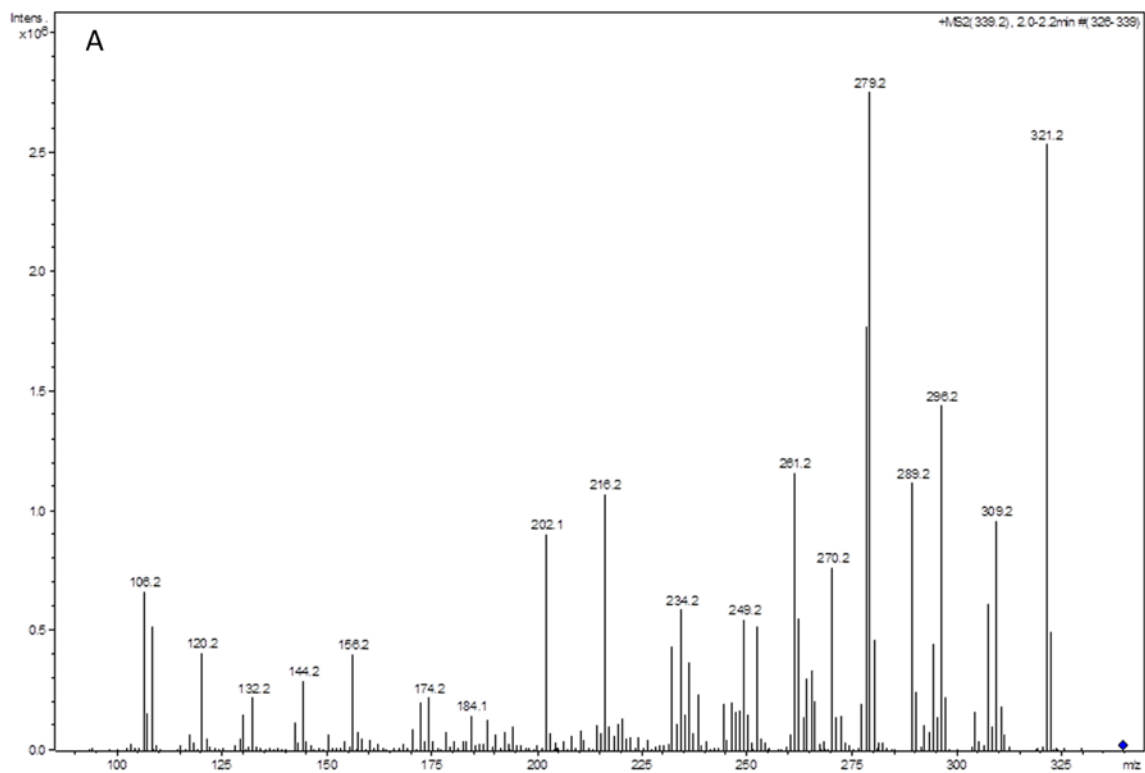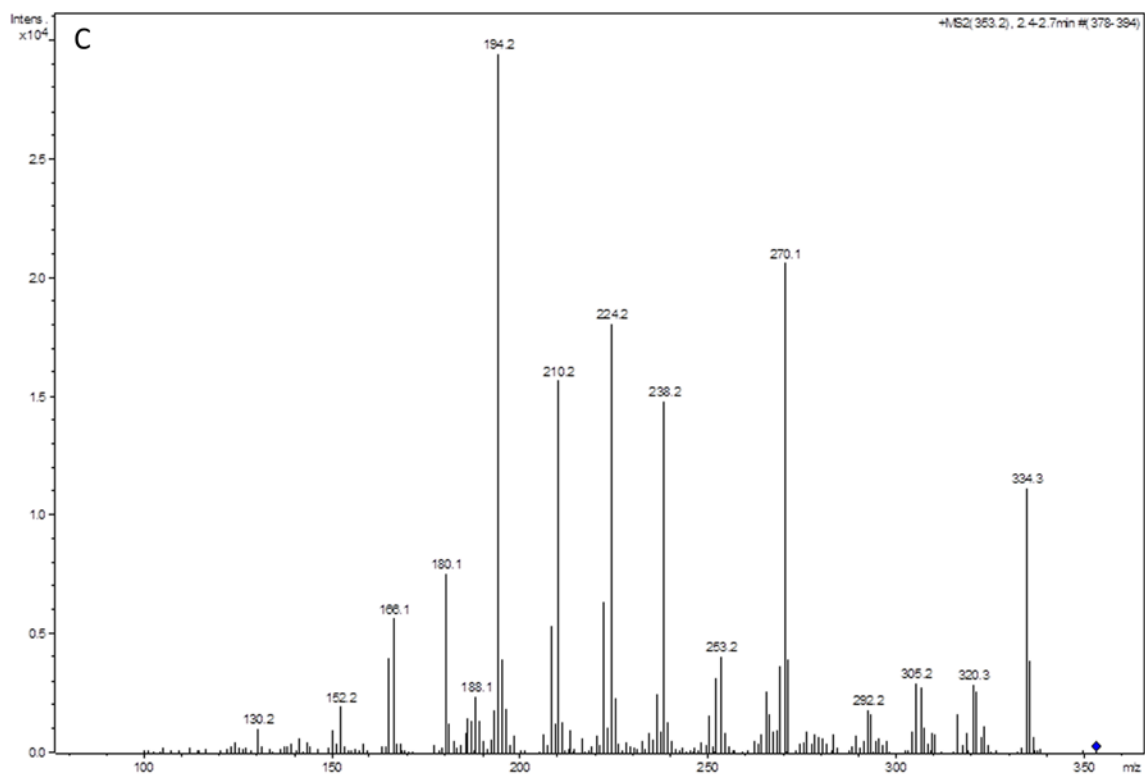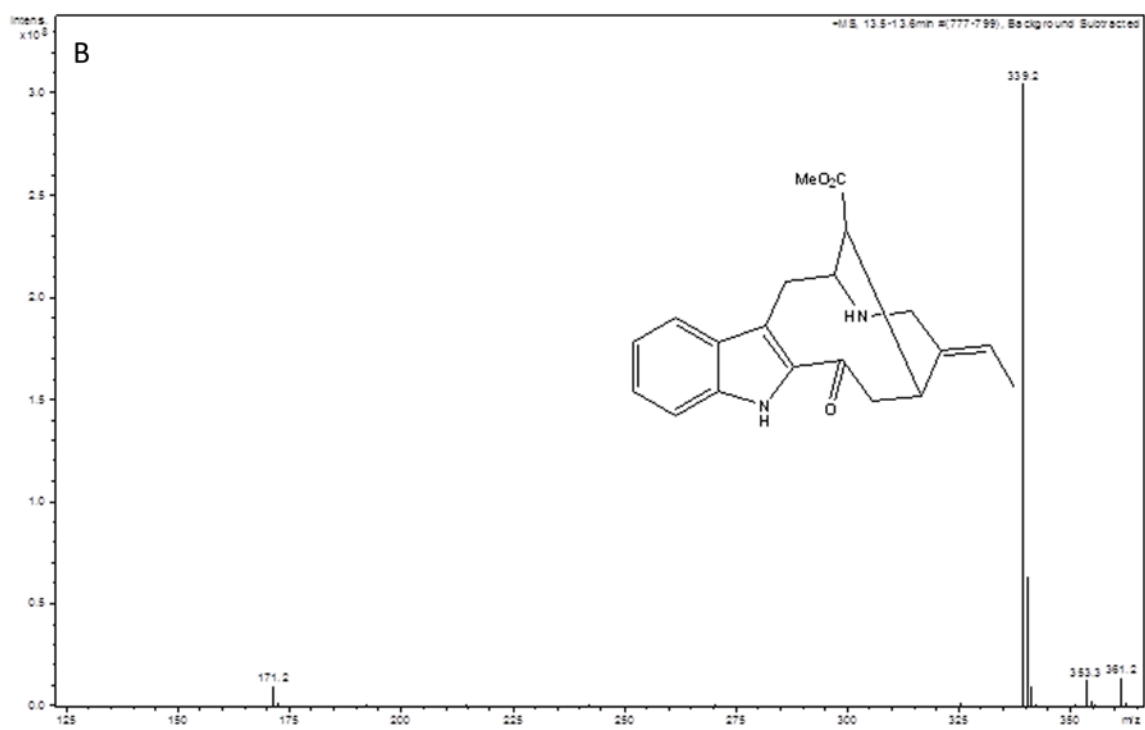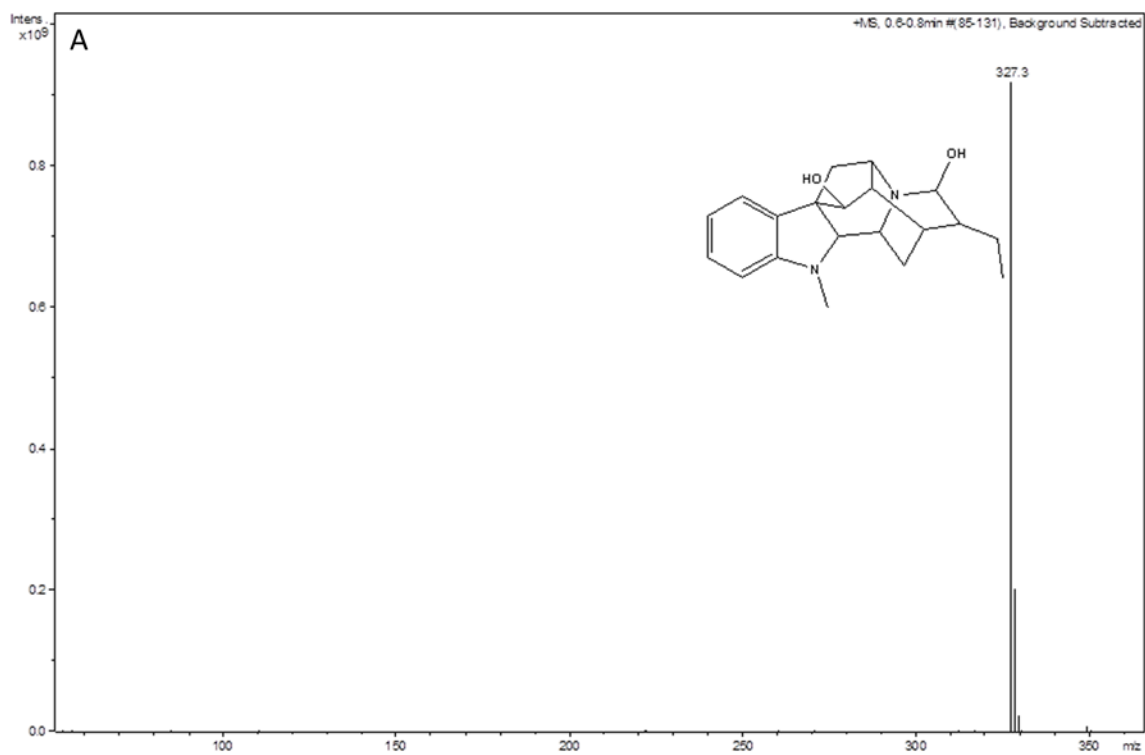
perivine standard.
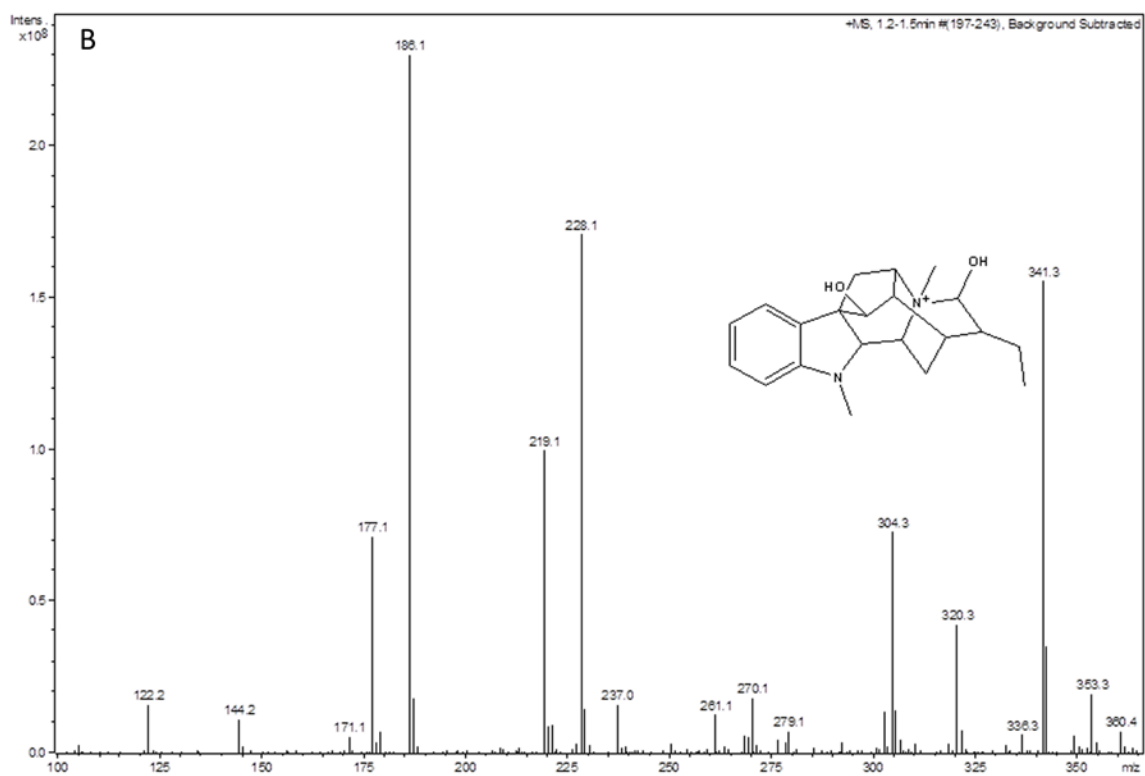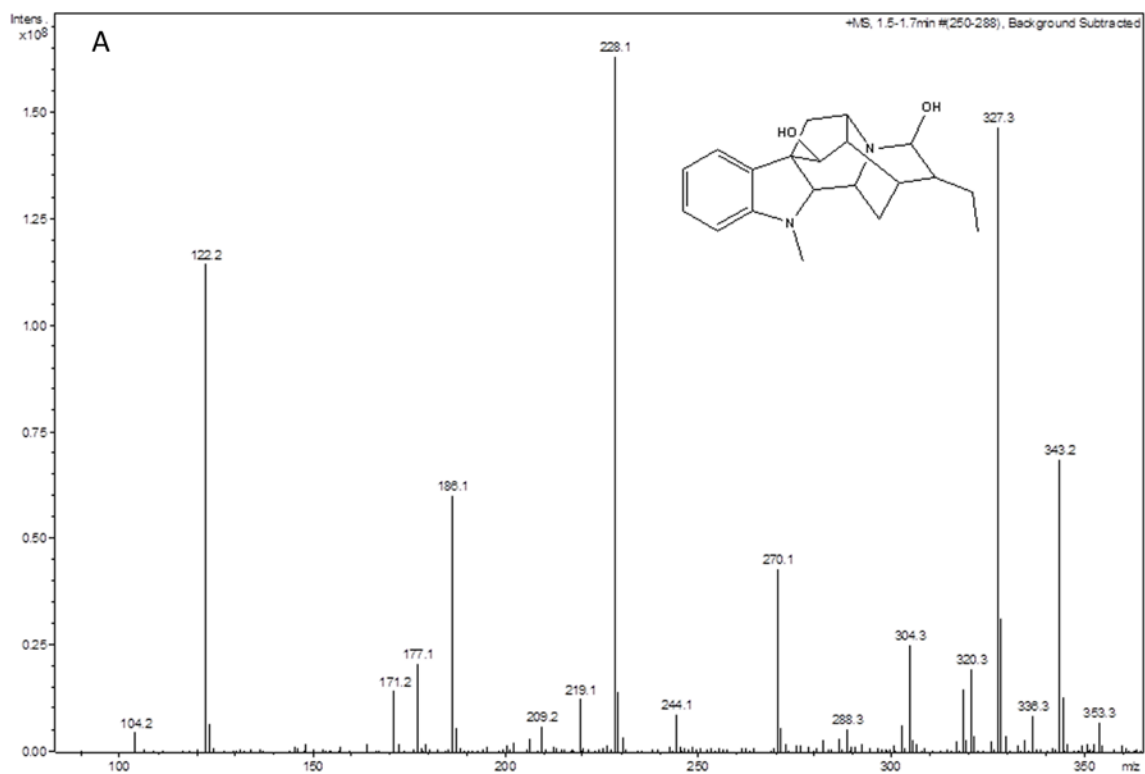
**Figure A-3.** MS1 analysis of alkaloid products of cell free pET 30b-EV (A) and recombinant RsAjNMT (B) methyltransferase assay productssupplied with ajmaline standard as substrate.
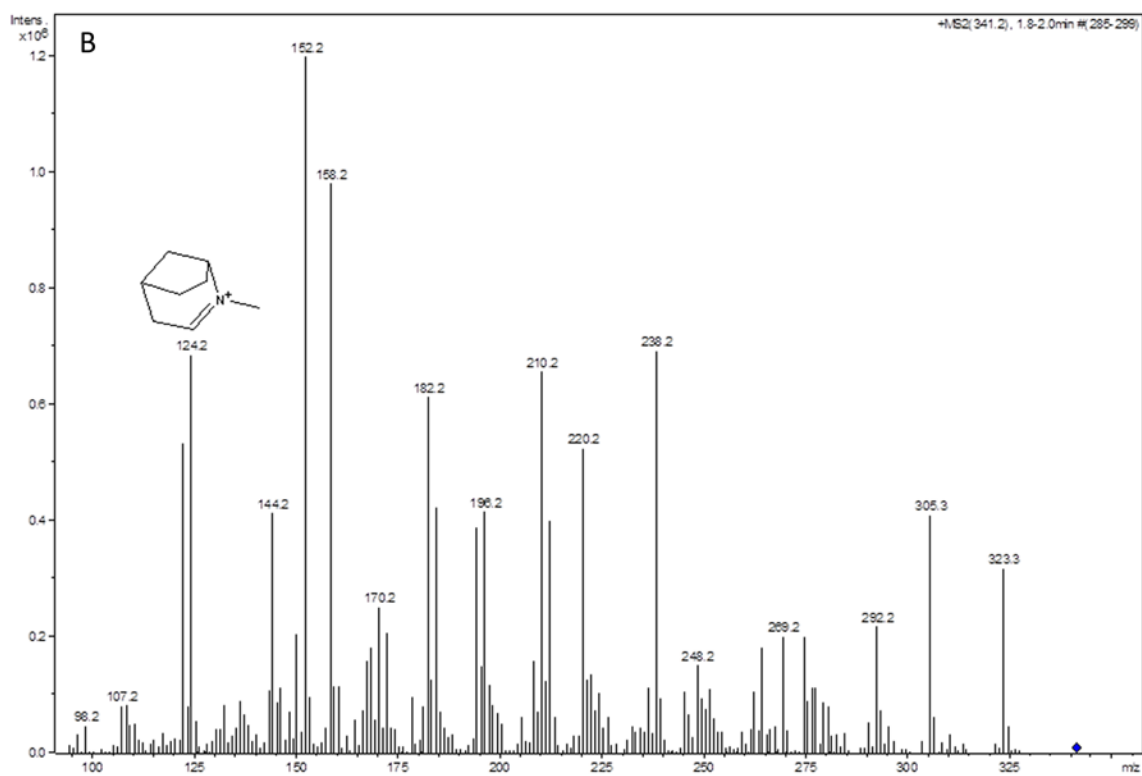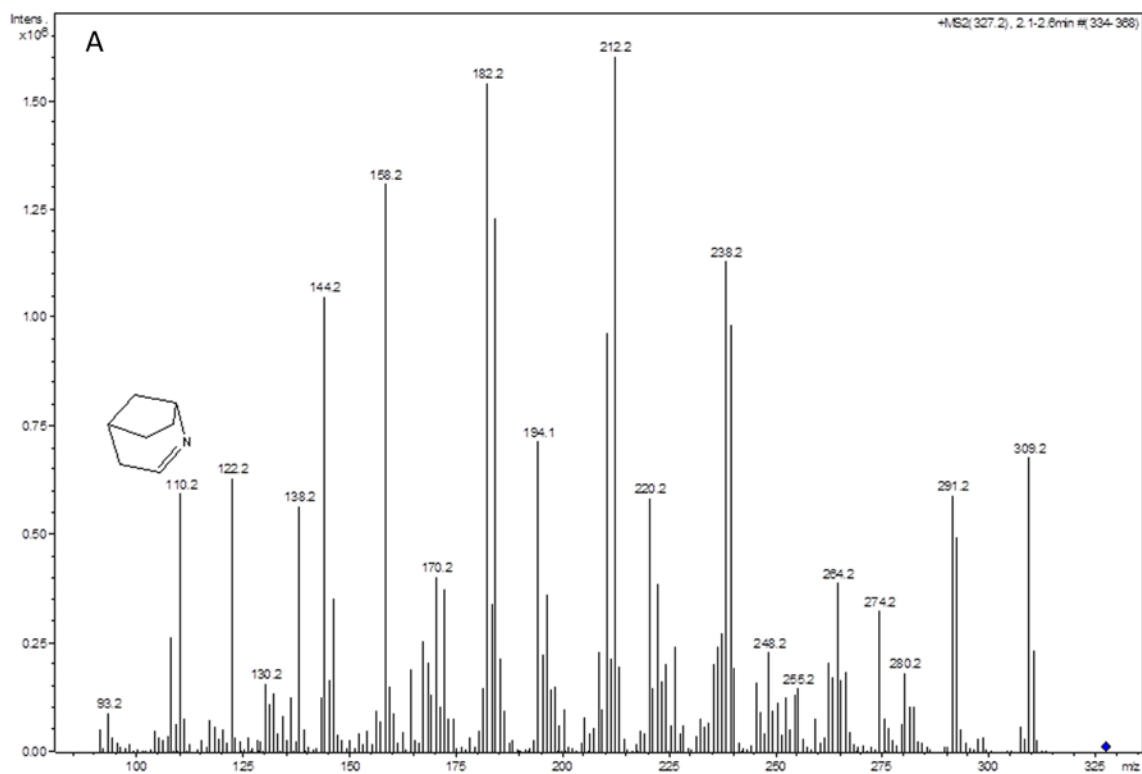
Figure A-4. MS2 analysis of 327 m/z +1 alkaloid of cell free pET 30b-EV (A) and 341

m/z +1 alkaloid of recombinant RsAjNMT (B) methyltransferase assayssupplied with

ajmaline standard.
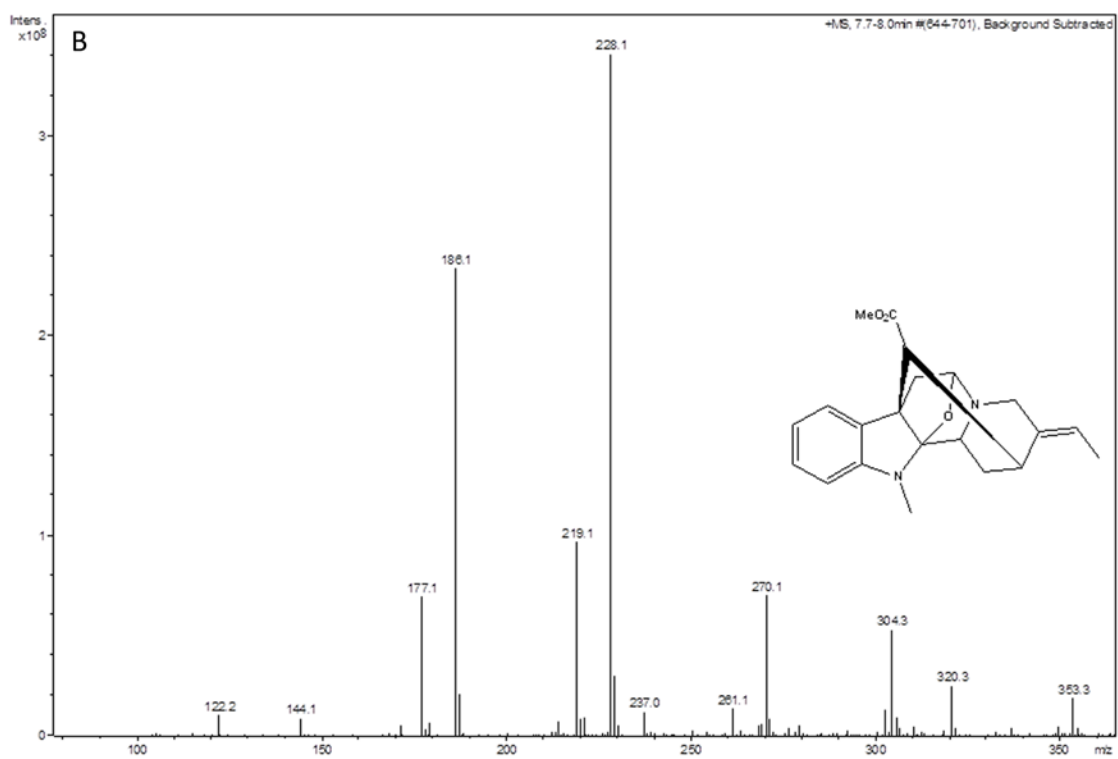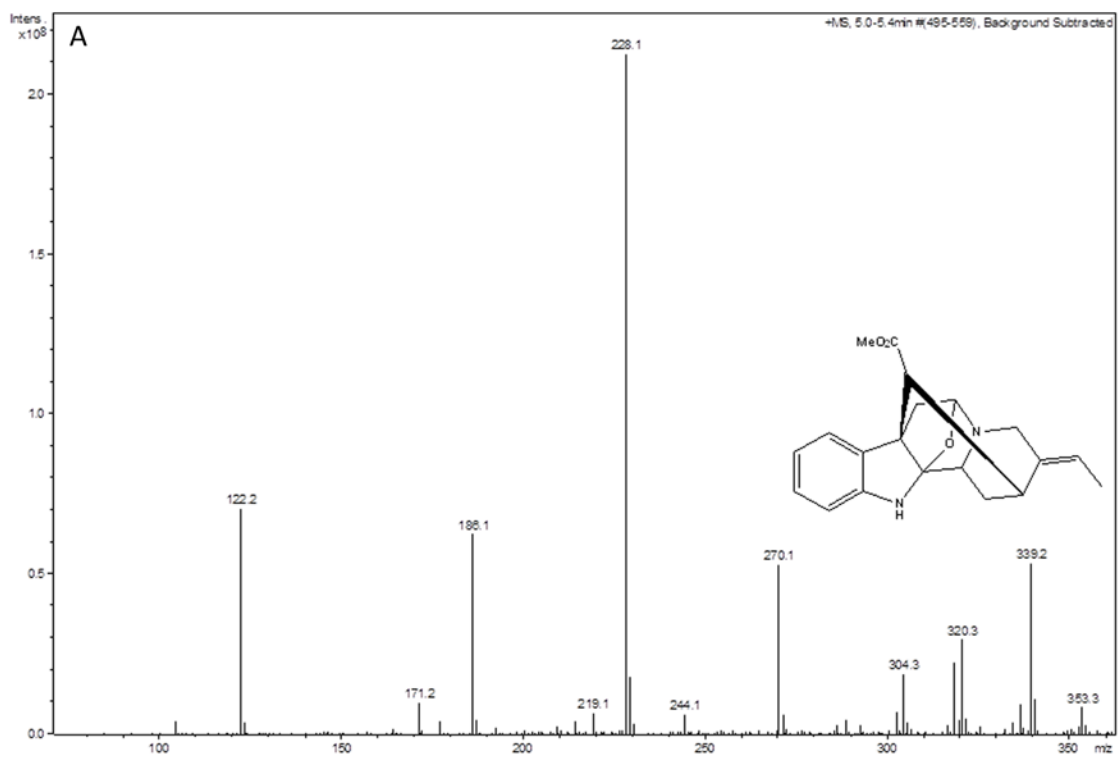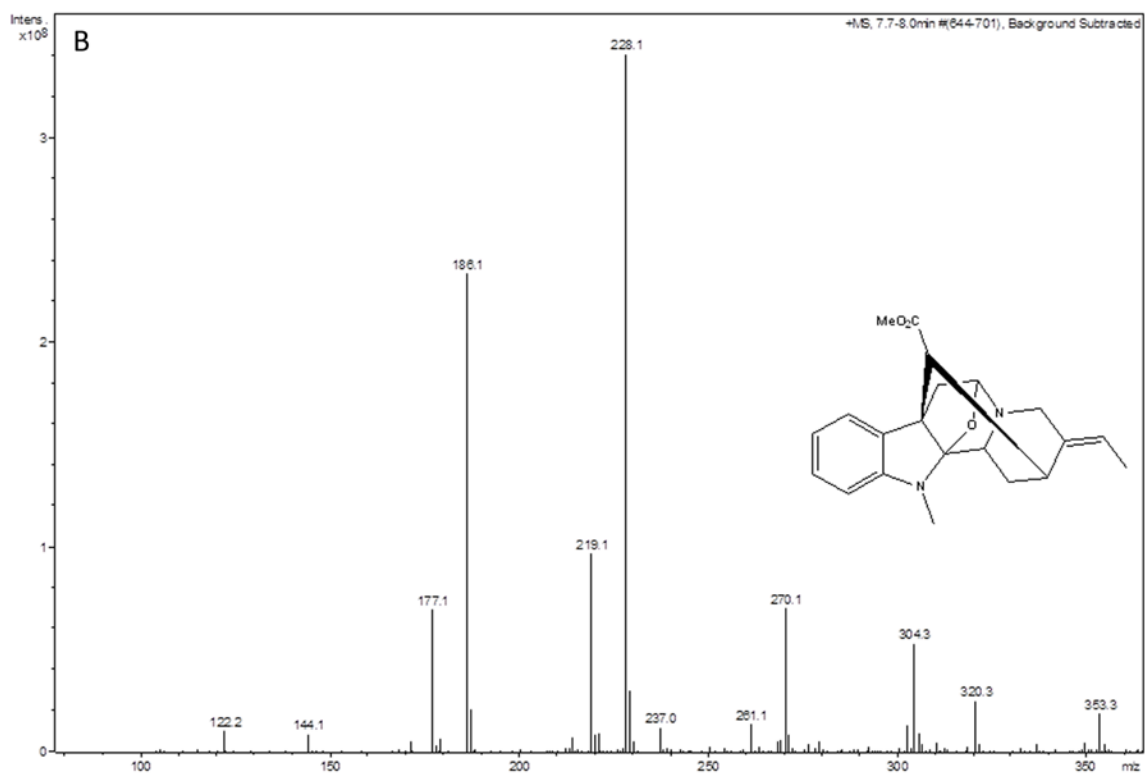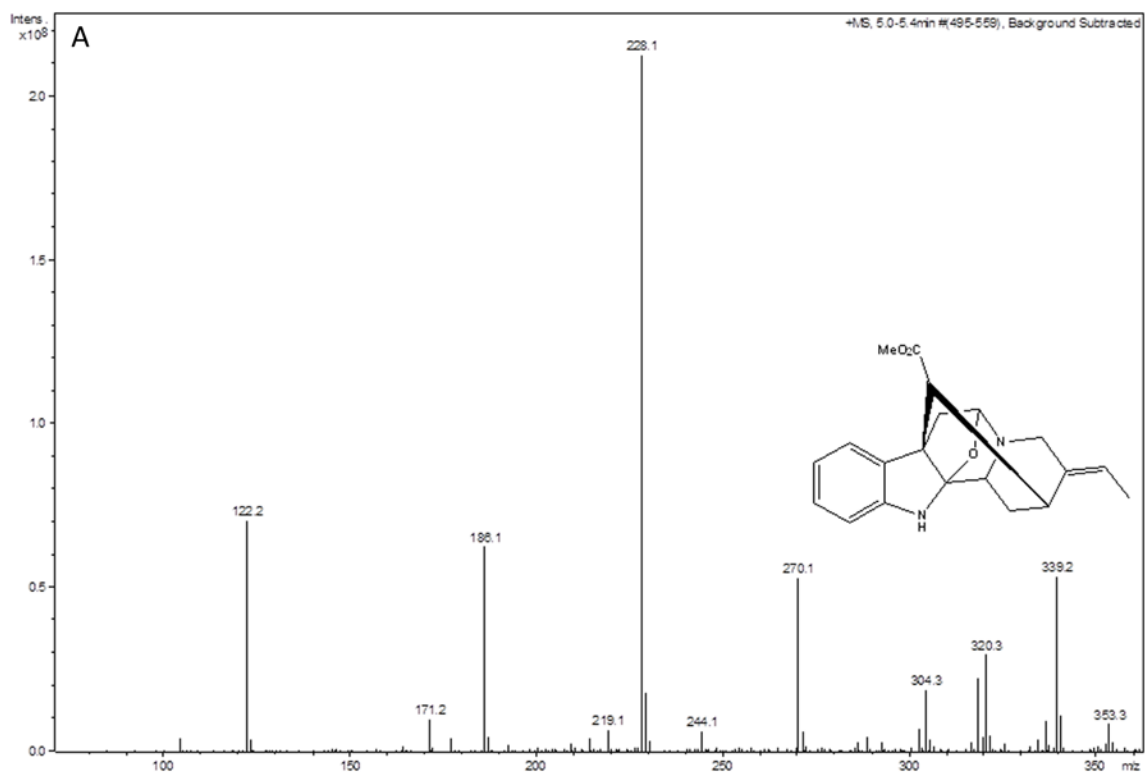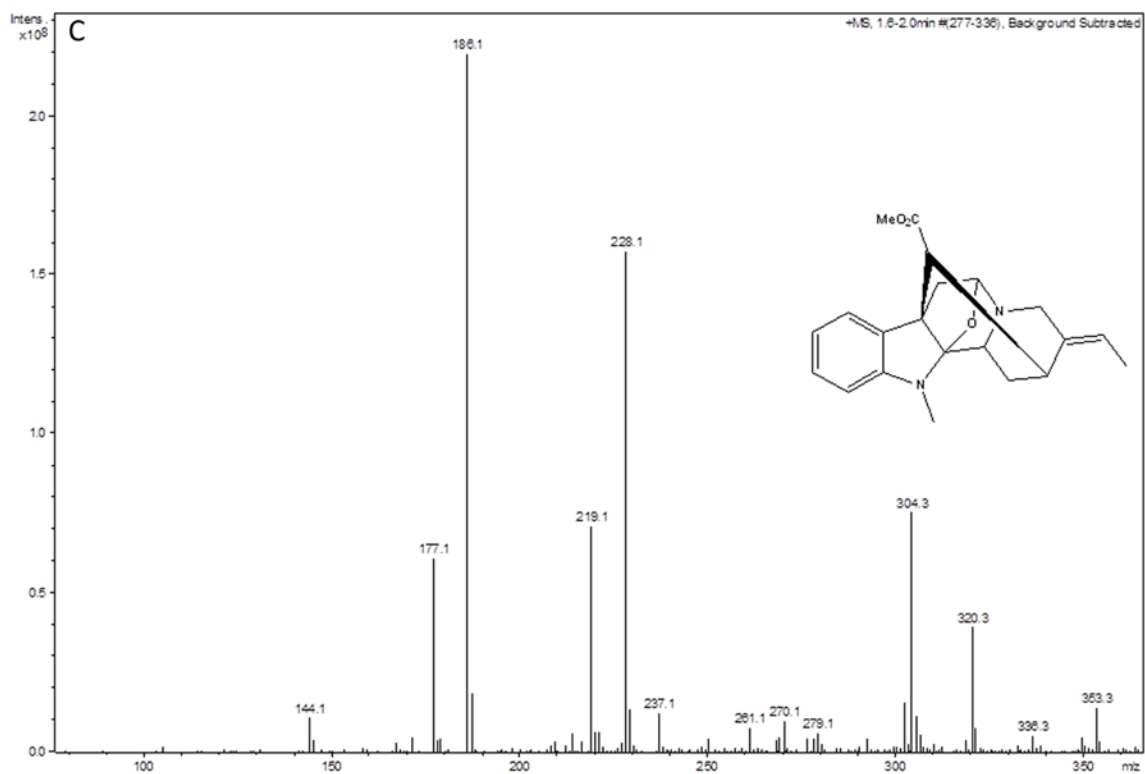
Figure A-5. MS1 analysis of alkaloid products of cell free pET 30b-EV (A), recombinant VmPiNMT (B) and recombinant RsPiNMT (C) methyltransferase assay productssupplied with picrinine standard as substrate.
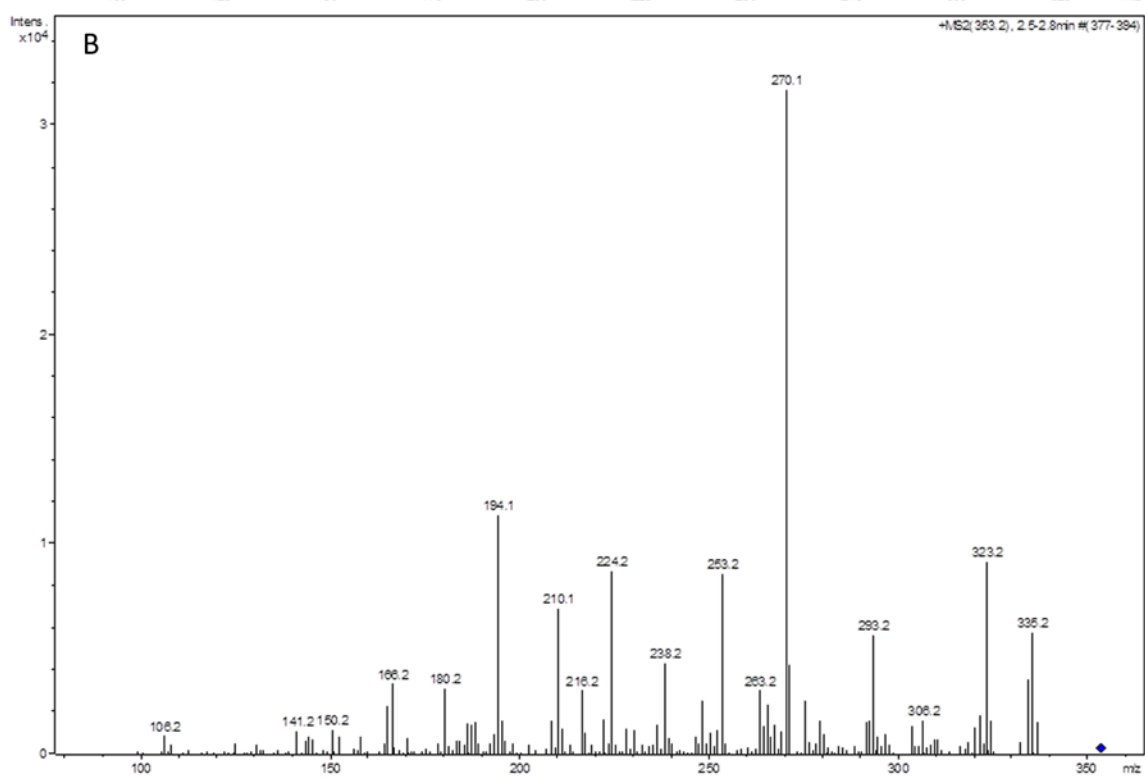
Figure A-6. MS2 analysis of 339 m/z +1 alkaloid of cell free pET 30b-EV (A) and 353

m/z +1 alkaloid of recombinant VmPiNMT (B) and recombinant RsPiNMT (C)

methyltransferase assays supplied with picrinine standard.
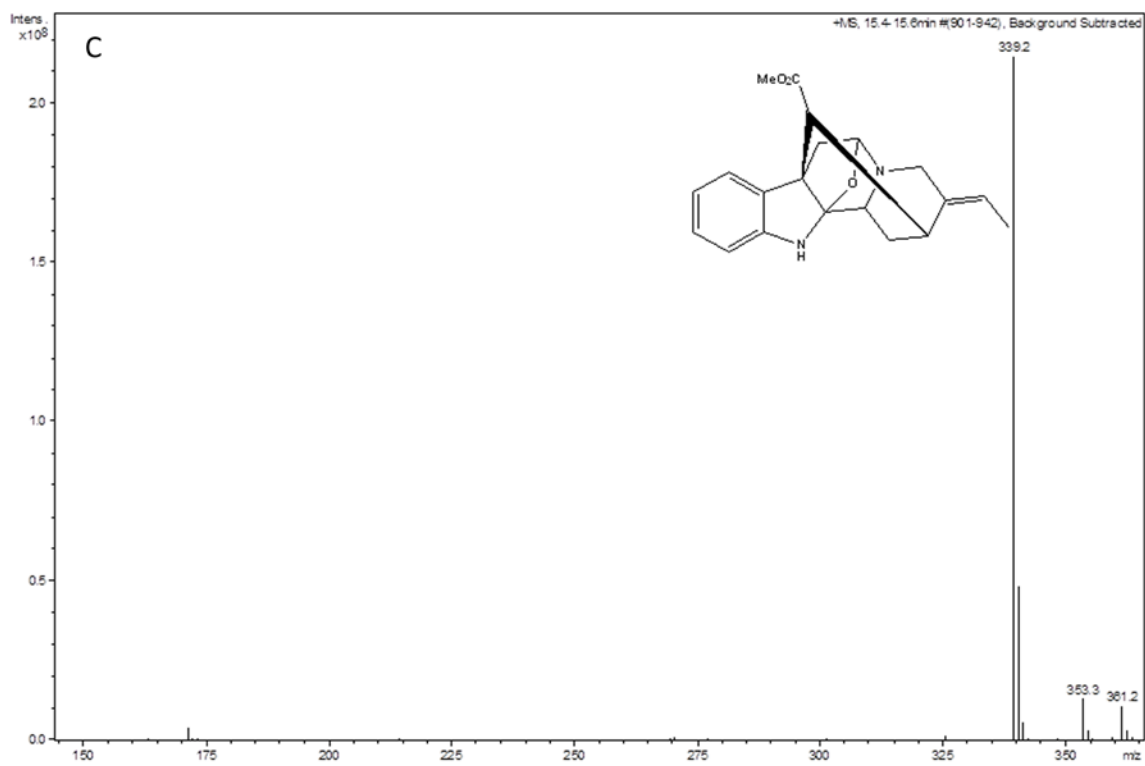
Figure A-7. MS1 analysis of ajmaline (A), perivine (B) and picrinine (C) standards used

in all enzyme assays.
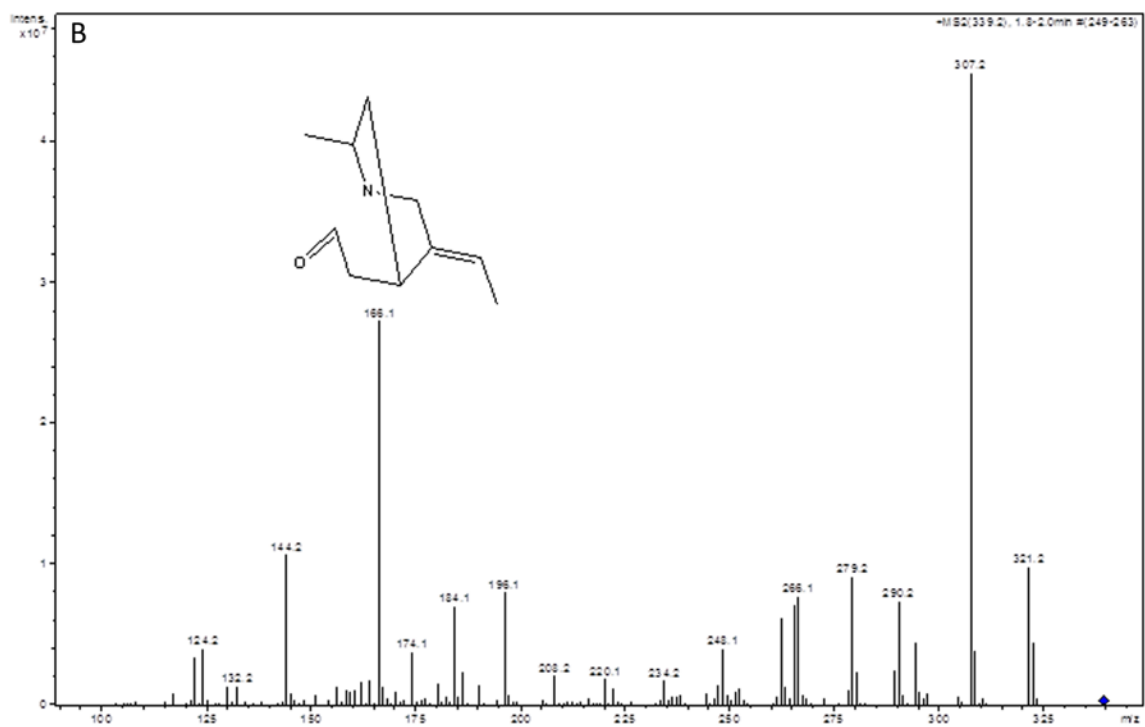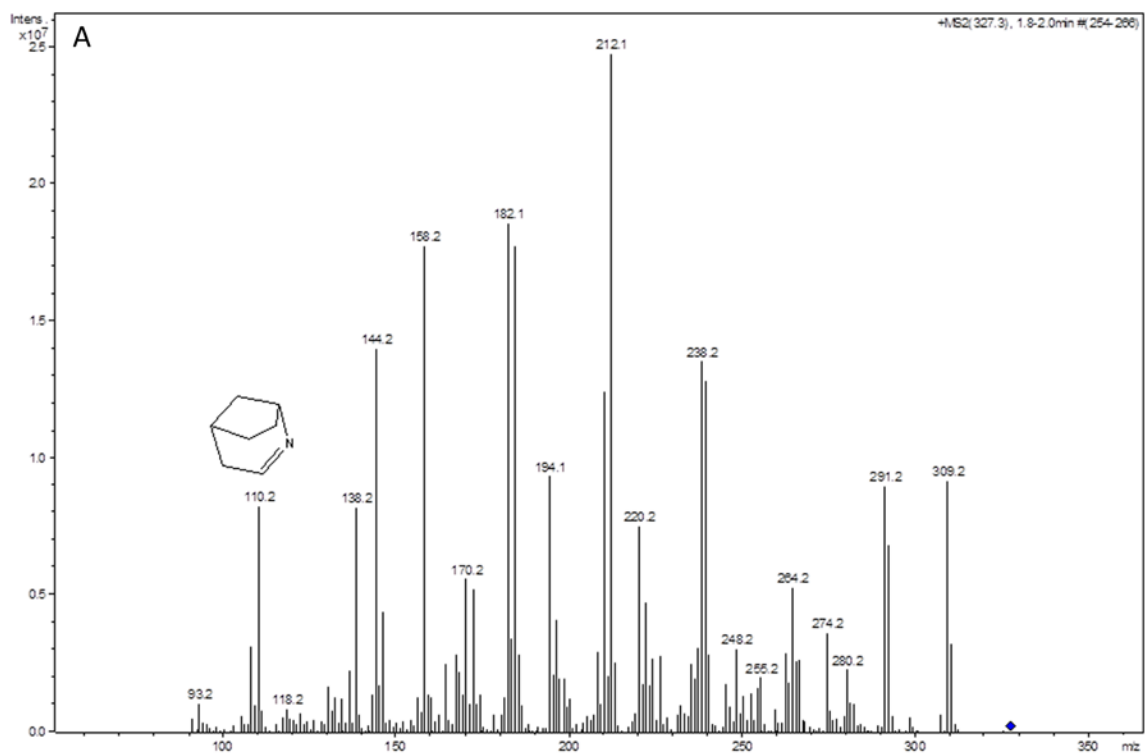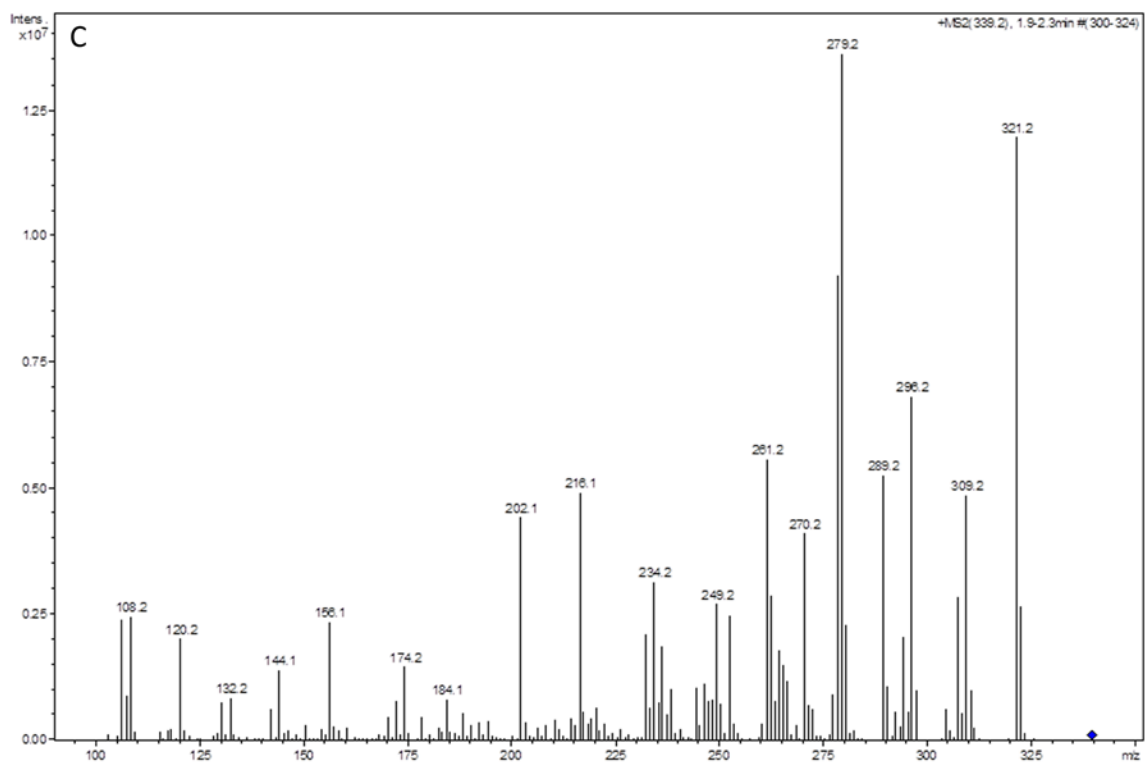
Figure A-8. MS2 analysis of ajmaline, 327 m/z +1 (A), perivine, 339 m/z +1 (B) and

picrinine, 339 m/z +1 (C) standards used in all enzyme assays.

A-2 SATURATION KINETICS OF TLMTS
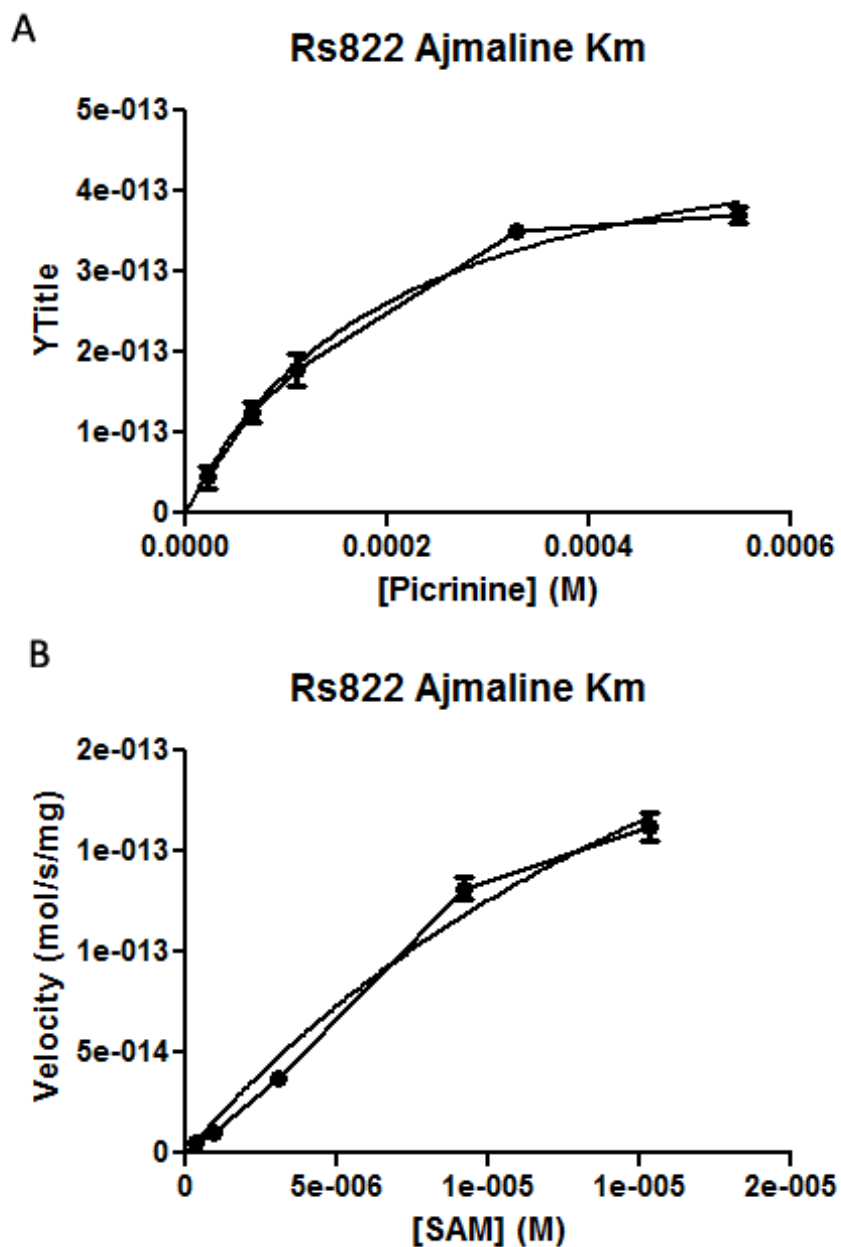
Figure A-9. Michaelis–Menten plots of kinetic data for RsAjNMT. Nonlinear regression

analysis of kinetic data were used to estimate kinetic parameters for RsAjNMT with

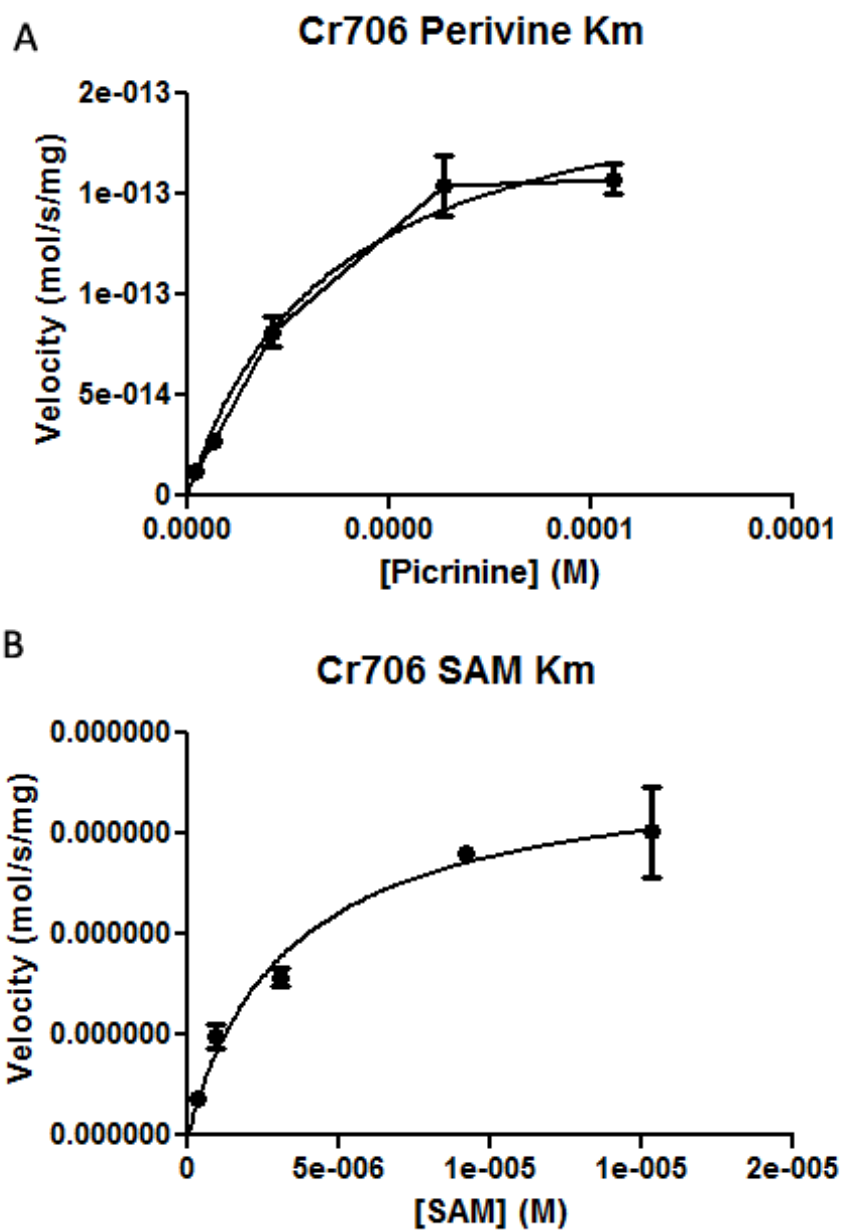substrates (A) ajmaline, and (B) S-adenosyl-L-methionine (SAM).

Figure A-10. Michaelis–Menten plots of kinetic data for CrPeNMT. Nonlinear

regression analysis of kinetic data were used to estimate kinetic parameters for CrPeNMT

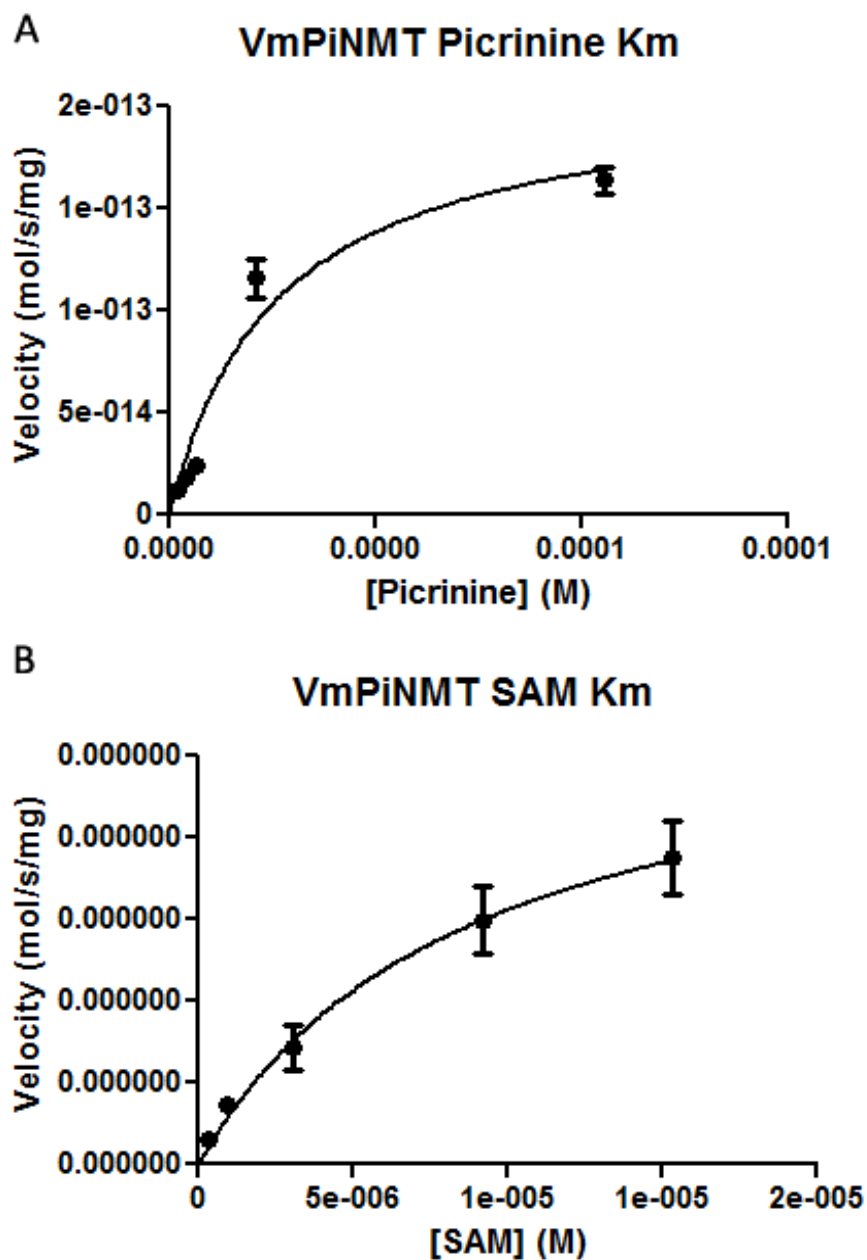with substrates (A) perivine, and (B) S-adenosyl-L-methionine (SAM).

Figure A-11. Michaelis–Menten plots of kinetic data for VmPiNMT. Nonlinear regression analysis of kinetic data were used to estimate kinetic parameters for VmPiNMT with substrates (A) picrinine, and (B) S-adenosyl-L-methionine (SAM).
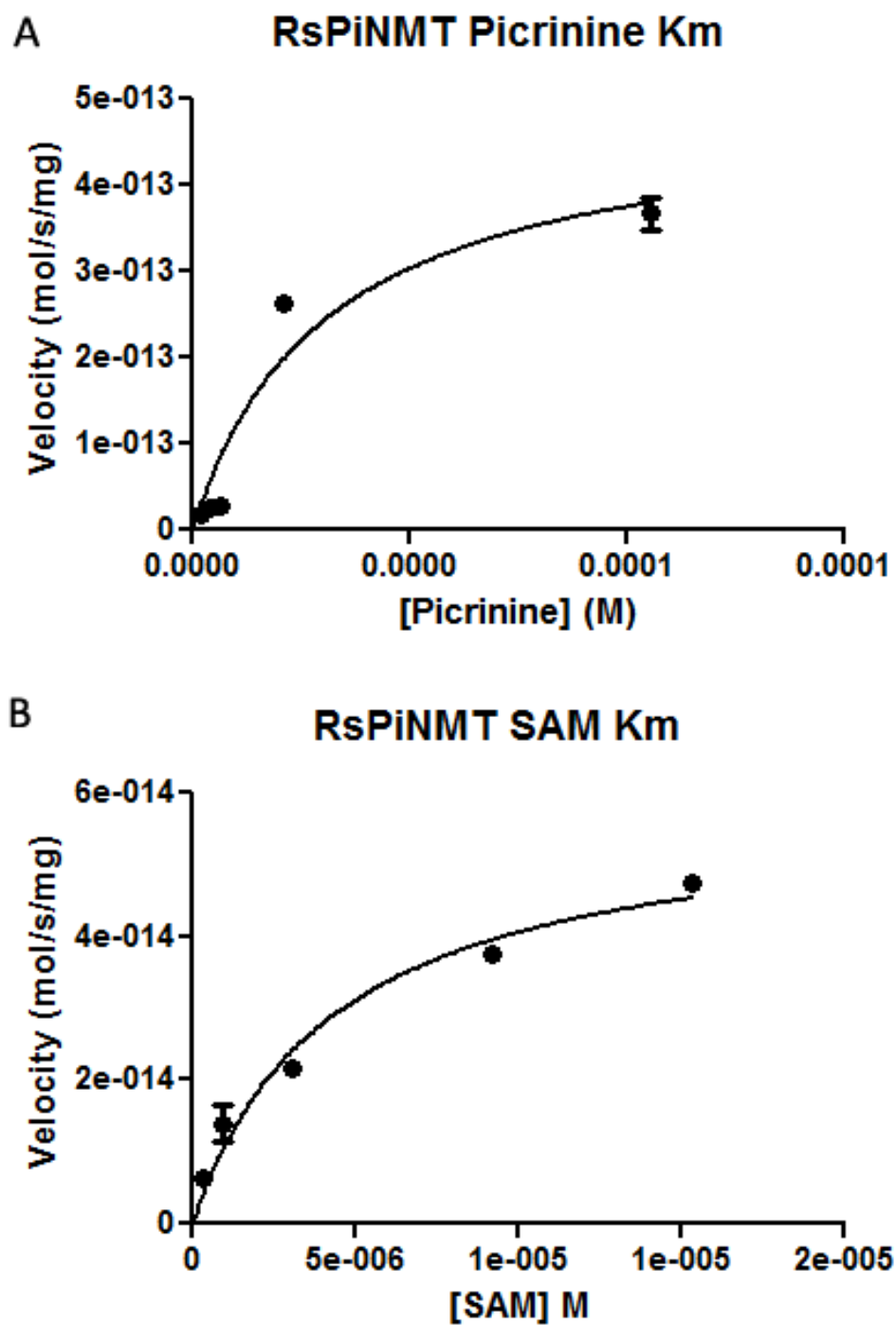
Figure A-12. Michaelis–Menten plots of kinetic data for RsPiNMT. Nonlinear regression analysis of kinetic data were used to estimate kinetic parameters for RsPiNMT with substrates (A) picrinine, and (B) S-adenosyl-L-methionine (SAM).