ANALYTICAL STRATEGIES FOR PROFILING AND STRUCTURE ELUCIDATION OF
INTERMEDIATES IN THE BIOSYNTHETIC PATHWAY OF A MONOTERPENE
INDOLE ALKALOID DRUG

By

Sujana Pradhan

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Chemistry – Doctor of Philosophy

2016

ABSTRACT

ANALYTICAL STRATEGIES FOR PROFILING AND STRUCTURE ELUCIDATION OF INTERMEDIATES IN THE BIOSYNTHETIC PATHWAY OF A MONOTERPENE INDOLE ALKALOID DRUG

By

Sujana Pradhan

Plants are exquisite biochemical factories responsible for synthesizing hundreds of thousands of diverse natural products. These compounds often have remarkable medicinal, pharmaceutical, and other human-health related benefits. Although plant natural products have had a great impact on disease treatment and are prime targets for drug development, full realization of their clinical potential has been hampered due to the target compounds being produced only in low amounts and in a small number of slow growing plant species. Efforts to engineer large-scale production in heterologous systems to combat the low abundance of these metabolites in nature, and to develop fermentation approaches to modify the pathways to produce novel pharmaceutical derivatives not found in nature have not met with great success. Furthermore, the known sources of most plant medicinal compounds have proven to be exceedingly difficult to approach biochemically and hence the genes encoding for the enzymes responsible for transformation of metabolites along the biosynthetic pathway remain elusive.

The research presented in this dissertation will focus on the monoterpene indole alkaloid camptothecin that is currently used to produce semi-synthetic anti-tumor drugs.

Three individual studies were performed to help elucidate the regulatory networks involved in the biosynthesis of camptothecin in *Camptotheca acuminata*. The second chapter in this dissertation focuses on whole-plant labeling of *C. acuminata* in a ^{13}C-enriched atmosphere to generate stable isotope-labeled metabolites and intermediates to track transformation from one

intermediate to the next. Liquid chromatography/mass spectrometry analysis (LC/MS) of leaf tissue extracts from labeled plants showed that the bottleneck in the production of camptothecin lies in the formation of downstream metabolites, but not in the formation of the iridoids in plants that are younger than 10 weeks old. The results showed that it is important to consider the developmental stage of the plant when performing metabolite studies in *C. acuminata.*

The third chapter in this dissertation focuses on the metabolite profiling of wild-type and RNAi-silenced *C. acuminata*. The LC/MS analysis of wild-type tissues showed that *C. acuminata* differs from other MIA-producing (monoterpene indole alkaloid) plants as it does not accumulate the well-known MIA precursor, strictosidine, and instead produces strictosidinic acid. Moreover, nearly all camptothecin pathway intermediates are present as multiple isomers, indicating an alternative seco-iridoid pathway. The analysis of root, stem, shoot apex, young and mature leaf tissues of transformed lines of *C. acuminata* demonstrated that the production of MIAs in tryptophan decarboxylase (TDC) silenced lines were deficient while the cyclase (CYC) silenced lines showed impaired ability to produce MIAs. These results indicate that the CYC knockout was effective in downregulating iridoid synthesis and the TDC knockout was able to successfully eliminate tryptamine synthesis.

The fourth chapter in this dissertation presents the use of a F5 (pentafluorophenylpropyl) column for chromatographic separation for isolation and purification of strictosidinic acid isomers, and 1-D and 2-D NMR (nuclear magnetic resonance) spectroscopy methods for the structure elucidation of the isomers. The results from NMR analysis demonstrated that the two major strictosidinic acid forms are stereoisomers at their glucosylated C21 positions. The knowledge obtained from this study extends our understanding of the multiple isomers that are observed through the pathway.

ACKNOWLEDGEMENTS

The completion of the research presented in this dissertation has been made possible by the help and support of a number of people. I would like to first express my deepest appreciation and gratitude to my Ph.D. advisor and mentor, Dr. A. Daniel Jones. I thank Dr. Jones for giving me the opportunity to learn from him, and for his constant encouragement during my time at MSU. I am very grateful for the late-night over-Skype discussions, as well as, the leadership, communication and life skills imparted to me by Dr. Jones. He has not only prepared me to be the best scientist I can be, but also to be a good human being. His passion for education and his guidance, patience and concern for his students both professionally and personally are qualities I respect and admire the most. His wealth of knowledge accompanied by his humble nature motivates me to follow in his footsteps each and every day. I consider myself very fortunate to have had the chance to work for and with him, and to be part of this rewarding experience.

I would like to thank my collaborators in Dr. Dean DellaPenna's research group at Michigan State University. It has been a wonderful learning opportunity to work with plant biologist. I am grateful for the plants samples that were provided to me for metabolite analysis.

I would also like to show sincere appreciation to my committee members, Dr. Merlin Bruening, Dr. Gary Blanchard, and Dr. Susanne Hoffmann-Benning for their constructive comments and suggestions on my work and in completing this dissertation. I am especially grateful to Dr. Bruening for agreeing to be my second reader, and to Dr. Blanchard for allowing me to unburden myself of graduate school problems in his office, and to Dr.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1. Introduction

1.1. Plant specialized metabolism

The growing population of the world inadvertently puts great strain on the supply of renewable resources including food, energy, and materials. A large part of this burden falls onto photosynthetic organisms that serve as the base of the food supply for humans and animals [1]. In order to sustain these nutritional needs, plants convert sunlight into chemical energy to produce biochemical intermediates that ultimately flow into central metabolites including carbohydrates, proteins, and fats [2]. These primary metabolites are universal across the plant kingdom and are synthesized through essential metabolic pathways [3]. Unlike central metabolites that are directly involved in the growth, development and reproduction of the plant, plants also make another set of metabolites known as secondary metabolites (now more commonly known as specialized metabolites) that are not required for life-sustaining processes but may play important roles in plant survival. These specialized metabolites are synthesized from central metabolites and are often restricted to particular taxonomic groups [4]. Due to the sessile nature of plants, specialized metabolites usually play roles in controlling interactions between the plant and its environment by defending the plant against pests, pathogens and predators. Specialized metabolites often act as toxins, attractants for pollinators, repellents, messengers, or as camouflage [5, 6].

1.2. Importance of plant specialized metabolites in medicine

After several billion years of evolution, nature has developed an enormous diversity in the production of specialized metabolites are produced by at least 500,000 different plant species [7]. A majority of these plant specialized metabolites are derived from a handful of building blocks, namely, the C_5-unit isopentenyl diphosphate, the C_2-unit acetate, the C_9-units phenylalanine/tyrosine, and additional amino acids. These give rise to classes of compounds including alkaloids, terpenes, phenolics, steroids and lignins [8]. Some of these compounds

have great medicinal, pharmaceutical, and human-health related benefits [9]. Notable examples include the strongly addictive and narcotic compound morphine isolated from opium (*Papaver somniferum*), the antipyretic and analgesic compound salicin isolated from willow (*Salix species*) which is used as a precursor for the synthesis of acetylsalicylic acid (aspirin), and the anticancer drug taxol isolated from Pacific yew (*Taxus brevifolia*) [10].

As mortality rates decrease and life expectancy increases, there is a heavier reliance on modern medicine [11]. For example, 51% of all anticancer drugs approved during 1981-2014 were either directly or indirectly derived from natural plant sources [12]. Their production yields considerable pressure on harvesting plant-derived natural products from the plant kingdom. Hence, nature presents an enormous potential for the discovery and development of new pharmacologically active compounds, but more efficient and cost-effective production of plant natural products is needed to satisfy growing needs for these high-value substances.

1.3. Growing need for plant metabolite identification

Even though plant natural products have been in the limelight for their profound impact on human health, attempts to keep up with the global demand by engineering synthetic routes for large-scale synthetic production of compounds are at times not feasible [5]. In addition, due to complexity in the structures of these compounds, chemical synthesis with multiple reaction steps and separation cycles is often deemed inefficient and impractical. Furthermore, the problems with exploiting plants as a large-scale source of medicinal compounds include low abundances of medicinal compounds, slow growth rates of medicinal plants, limited number of species with medicinal properties and limited plant natural resources [10]. Thus, development of alternative sustainable and renewable production systems has growing urgency. Metabolic engineering of plants to produce more of the desired compounds is hampered by lack of versatile genetic tools, as well as long generation times. Hence, metabolic

engineering of plant-derived natural products synthesized in heterologous hosts, either plant or microbial, provides a more attractive solution for such problems [13-16]. Yet, even with the ability to sequence genomes, the greatest challenge in accessing plant metabolic pathways is that most associated gene functions corresponding to the specialized metabolite biosynthesis and regulation remain largely unknown, as are the identities of biosynthetic intermediates. In order to tackle this problem and close the gap between gene annotation and gene function, it is prudent to first define the metabolites involved in the biosynthetic pathways.
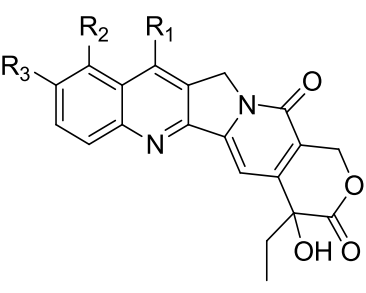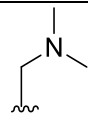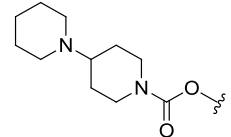
1.4. Introduction to monoterpene indole alkaloids (MIAs)

Many higher plants are proficiently synthesize alkaloids, a large class of complex nitrogen-containing low molecular weight compounds. Among alkaloids, the iridoid-derived monoterpene indole alkaloids, found in thousands of plant species from the Nyssaceae, Apocynaceae, Rubiaceae, Loganiaceae, Icacinaceae and Alangiaceae plant families, have shown promising medicinal properties [17]. Some notable examples include the anti-cancer compound vinblastine, the rat poison strychnine, and the anti-malarial drugs quinine and artemisinin [17] (Figure 1.2). This study will explore the biosynthesis of the high clinical value anti-cancer drug camptothecin will be explored.

Camptothecin was originally isolated from bark and stem of *Camptotheca acuminata* (Table 1.1); a tree native to south China and Tibet [18]. The full potential of camptothecin as an anti-tumor drug was not widely explored until 1985 when Hsiang established its unique mechanism of action [19]. Hsiang concluded that camptothecin inhibits the activity of DNA topoisomerase I by stabilizing the topoisomerase-DNA covalent complex and preventing re-ligation of DNA strands and blocking DNA replication. This causes irreversible DNA damage and subsequent death of neoplastic cells. However, camptothecin exhibited less-than-desirable properties as a drug, and has now become the natural 'lead' compound for semi-synthesis of various analogs that are less toxic and more water-soluble. The most popular examples of FDA-

approved camptothecin analogs are Topotecan (Hycamtin®) and irinotecan (Camptosar®), which are widely used for the treatment of ovarian and small cell lung cancers, and colo-rectal cancers, respectively [20, 21] (Table 1.1).

Table 1.1. Camptothecin and FDA-approved analogs

	R_1	R_2	R_3	Name
	H	H	H	Camptothecin (CPT)
	H	(N,N-dimethylaminomethyl group)	OH	Topotecan
	C_2H_5	H	(piperidinopiperidine carbamate group)	Irinotecan

Though efforts have been made to synthesize camptothecin and its analogs [22], these low yield multi-step syntheses have not been amenable to scale-up for large-scale drug production. Hence, the primary production of camptothecin still involves extraction from *C. acuminata* and *Nothapodytes nimmoniana*. Unfortunately, camptothecin is produced only in tiny amounts in *C. acuminata* and in a handful of other angiosperms [23, 24]. Moreover, *C. acuminata* is a slow growing woody plant. As an alternative, young leaves of *C. acuminata* that can accumulate about 4-5 mg of camptothecin per gram of dry weight [25] have been repeatedly harvested without cutting down the entire tree. Yet, current prices of irinotecan and topotecan are approximately \$40,000/g and \$120,000/g, respectively. These factors mentioned above have led to a need for the development and utilization of plant and microbe metabolic engineering to fashion a more sustainable production of camptothecin in heterologous hosts [26]. In order to be successful in producing medicinally important metabolites, a profound understanding and investigation of the biosynthetic pathway of camptothecin at the molecular level is critical, and such discoveries will require improved analytical approaches to determine

the pathway of camptothecin biosynthesis by identifying metabolic intermediates in the pathway.

These factors have plagued the elucidation of biosynthetic pathways of many medicinally important metabolites, including one class of secondary metabolites – the monoterpene indole alkaloids (MIA). Specifically, this study will focus on structure elucidation of intermediates involved in the biosynthetic pathway that gives rise to camptothecin; a MIA which was initially isolated from *Camptotheca acuminata* in 1966 during a search for anticancer drugs [18]. To accelerate metabolite identification and structural annotation necessary for probing the biochemical genomics of *C. acuminata*, it is crucial to harness the innovations of multiple analytical strategies such as liquid chromatography (LC), mass spectrometry (MS) and nuclear magnetic resonance (NMR) spectroscopy, and to mine DNA sequences and mRNA transcript levels from *C. acuminata* that have recently been made public (http://medicinalplantgenomics.msu.edu/, http://metnetdb.org/mpmr_public/). Furthermore, stable isotope-labeled metabolites and intermediates can be especially advantageous in the use of MS and NMR because these techniques can demonstrate label incorporation and show relationships between precursor and product. In the future, this will not only enable identification of candidate biosynthetic gene activities necessary to generate these compounds in heterologous systems but also allow novel or known bioactive compounds to become natural 'lead' molecules for semi-synthesis which is necessary to decrease the toxicity elicited by camptothecin and its analogs [27]. Even though recent advances in metabolomics, functional genomics, proteomics and bioinformatics have all contributed in the quest for elucidation of the intermediate steps of the biosynthetic pathways, there is much about specialized metabolism in plants yet to be fully understood [28].

1.4.1. Biosynthesis of terpenes and indole alkaloids
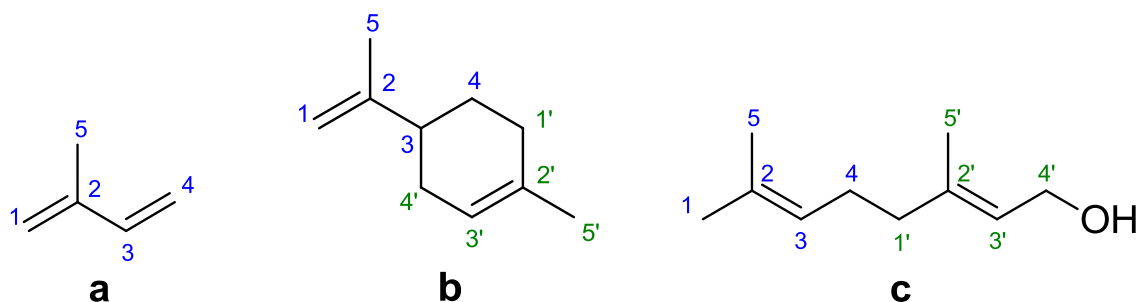
Figure 1.1. Structures of isoprene (a), the monoterpene limonene (b) and the monoterpenoid geraniol (c).

Terpenes and terpenoids are made up of isoprenoid C_5-units (Figure 1.1) joined together in a head-to-tail linkage in linear or ring arrangements [29]. Monoterpene C_{10}-units are classified as having two isoprene units (Figure 1.1b), and "terpenoid" is a term used to describe modified terpenes with oxygen functionalities included or rearrangements (Figure 1.1c).

The biosynthesis of terpenoids begins with the universal isoprenoid precursors, isopentenyldiphosphate (IPP) and dimethylallyldiphosphate (DMAPP). Plants produce both IPP and DMAPP are produced in two distinct pathways; the cytosolic mevalonate (MVA) pathway [30] and the plastidial non-mevalonate 2-C-methyl-D-erythritol-4-phosphate (MEP) pathway [31]. Though these distinct pathways can operate independently, there have been reports of metabolic cross-talk between them [32].

The precursors to the MVA pathway are two molecules of acetyl CoA whereas the MEP pathway begins with pyruvate and glyceraldehyde-3-phosphate. When IPP and DMAPP are finally produced after a series of enzymatic reactions in each pathway, these 5-carbon precursors can undergo several transformations; 1) isoprene synthase can convert DMAPP to isoprene which is usually emitted from the leaves of the plant owing to its volatility, 2) prenyltransferases catalyze the consecutive condensation of IPP with allylic diphosphates to

synthesize linear prenyl diphosphates of various chain lengths. 3) geranyldiphosphate synthase can condense one molecule of IPP and one of DMAPP to yield geranyldiphosphate (C_{10}) which can be further converted into an assortment of cyclic and acyclic monoterpenes (C_{10}) by enzymes identified as terpene synthases and cyclases [33] (Figure 1.3). There are a multitude of possible isomers of terpenes and terpenoids that differ in double bond configurations and cyclizations. In this dissertation, focus will be given to the generation of the monoterpene camptothecin by probing its biosynthetic pathway.
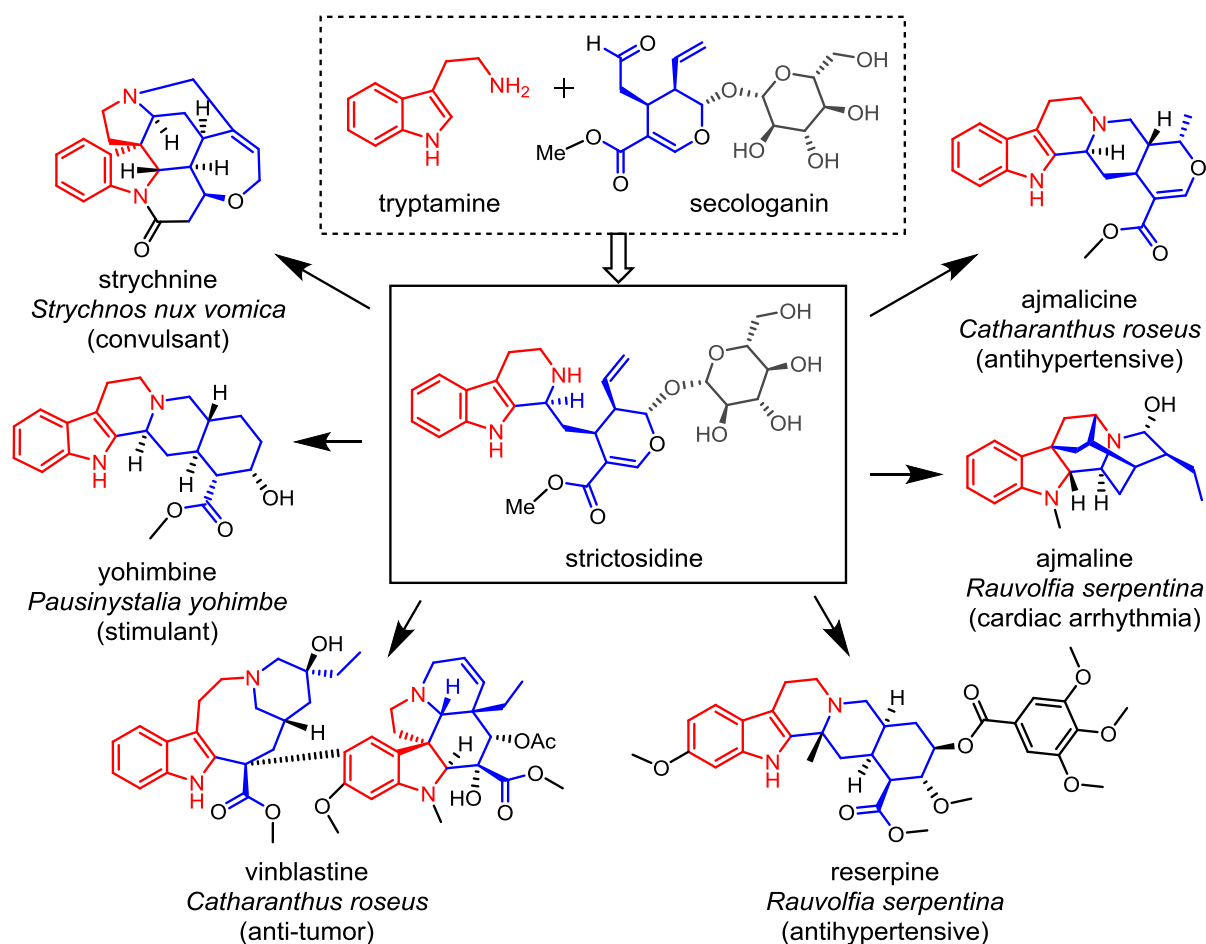
Figure 1.2. Strictosidine as the precursor for all monoterpene indole alkaloids. Structural complexity and chemical diversity can be seen in these plant natural products.
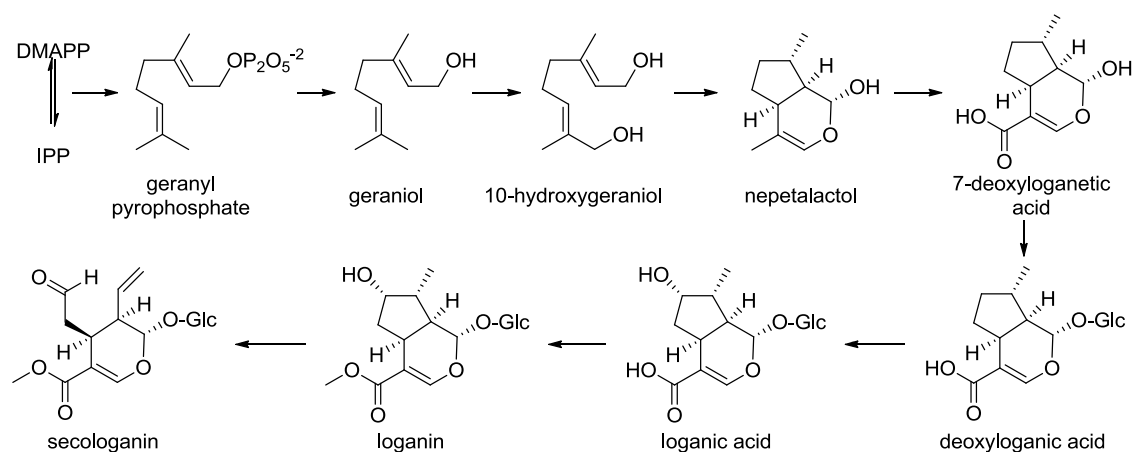
Figure 1.3. Proposed biosynthesis of secologanin beginning with isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP).

1.4.2. *Camptotheca acuminata* as a non-model medicinal plant

The common precursor for the biosynthesis of camptothecin and all other MIAs his proposed to be strictosidine, which is believed to be derived from the condensation reaction between the indole metabolite tryptamine and the iridoid glucoside secologanin, catalyzed by the enzyme strictosidine synthase [34]. Tryptamine is generated through the decarboxylation of tryptophan catalyzed by tryptophan decarboxylase [35] (Figure 1.2). In contrast, secologanin is an example of a monoterpene that has been strongly suggested to have been derived from the non-mevalonate MEP pathway [36]. It is a precursor to many other natural products and occupies a key position in numerous biosynthetic pathways [17]. The biosynthetic pathway that produces secologanin is proposed to occur through the several intermediates (Figure 1.2).

Many of the enzymes involved in the biosynthesis of secologanin have not been clearly defined in terms of their functions. However, secologanin synthase, a P450 oxidase, has been proven to be responsible for the last step in generating secologanin by oxidative cleavage of loganin *in vitro [37]*. This key step generates a metabolite, secologanin, with a reactive aldehyde group capable of condensing with primary amino groups, such as the one in tryptamine.

In *Catharanthus roseus*, strictosidine is converted to strictosamide via intramolecular cyclization. However, in *C. acuminata,* strictosidine is not observed in extracts of any tissues as described in Chapter 3, and instead, two isomers of strictosidinic acid were detected in UHPLC/MS analyses.

1.5. Analytical platforms for metabolomics

Metabolomics is the term given to a comprehensive and quantitative analysis of chemical processes involving all metabolites in a given biological system, and this approach has grown in favor as a more global approach to understanding metabolism and its biological

consequences [38] as well as a strategy for functional genomics. The added challenge in metabolomics compared to other 'omics' is that the complexity, diversity, isomerism and large differences in abundances of metabolites in a single metabolome make it impossible for any single analytical platform to provide efficient identification and quantification of all metabolites. Therefore, it is important to draw on a range of analytical techniques that are both selective and sensitive, and offer rapid characterization of compounds present in crude sample extracts in a high-throughput fashion. Currently, the most powerful analytical methods for metabolite profiling include gas or liquid chromatography (GC/LC), mass spectrometry (MS), and nuclear magnetic resonance (NMR) spectroscopy. Selection of instrumentation is contingent on the aims of the analysis and is usually a compromise among sensitivity, selectivity and speed. For example, multidimensional NMR is a highly selective, non-destructive technique that is generally considered the gold standard for metabolite structure elucidation, but it falls short due to high detection limits (~5 nmol) and the necessity for relatively pure samples when structures have to be determined. In contrast, separation techniques such as LC or GC coupled with MS for orthogonal detection and validation of metabolites offers both high selectivity and sensitivity, with a detection threshold of about 10^{-12} mol [39].

For separation of metabolites, GC is a low cost alternative but requires that analytes be volatile. Although many non-volatile compounds can be chemically derivatized to make them volatile, this comes at the cost of additional sample preparation time and analytical variance. On the other hand, LC separates polar or thermally unstable metabolites using conventional C18 columns with particle sizes of 3-5μm. Other advantages of LC relative to direct introduction into a mass spectrometer include separation of isomeric metabolites, reproducibility of orthogonal data such as retention time, accurate quantification of individual metabolite, and reduction of matrix effects and suppression of ionization. The coupling of GC

or LC to MS has additional benefits including specific chemical information such as accurate mass, isotopic distribution patterns that can be used to derive elemental formulas, and characteristic fragment ions that can be used to pinpoint the class of compounds the metabolite may belong to [40]. Major limitations of GC/LC MS profiling methods are the frequent reliance on existing mass spectrum databases or reference compounds for comparisons and the inability to determine the exact identity of the metabolite or differentiate isomers based only on masses and abundances of fragment ions. Since sensitivity is more important than selectivity when rapidly mining a greater fraction of a chemically complex metabolome, the mass spectrometer becomes a logical precursor to the isolation and NMR analysis of metabolites of interest [41].

1.5.1. Mass spectrometry

Modern mass spectrometers are equipped with an array of technologies that vary in ionization techniques and mass analyzers, resulting in varying mass accuracy, sensitivity, resolving power and acquisition time. Electrospray ionization (ESI) is the most common ionization technique used in analysis of biomolecules. It is an atmospheric pressure soft ionization technique used to generate desolvated gas phase ions by spraying the analyte solution from the tip of capillary that is held at a high potential [42]. The use of ESI in MS began in 1968 when Malcolm Dole employed electrospray as an ionization source for MS to determine the molecular weight distribution of oligomers in synthetic polymers [43]. Roughly twenty years after Dole's implementation of electrospray, John Fenn and co-workers further developed ESI for the analysis of large biological macromolecules. Since then, ESI has been the foremost choice of modern practitioners of MS. The main reason for its prominence over other ionization methods is its suitability with a wide range of analytes with regards to their thermal stability, size and polarity. When compared to electron ionization or chemical ionization, ESI has been used to overcome the propensity of analyte fragmentation because

thermally labile inorganic compounds and biomolecules can now be analyzed without destruction during evaporation [44], allowing for molecular mass determination. This is particularly important when elucidating chemical structures of unknown metabolites from a complex mixture.

1.5.2. Multiplexed collision induced dissociation (CID) mass spectrometry

The generation of fragments from molecular ions is a valuable tool for probing the functional groups comprising the metabolite of interest. Collision induced dissociation (CID) is the most common method for fragmentation with soft ionization methods such as ESI. In CID, precursor ions enter a collision cell and collide with collision gas molecules resulting in a transfer of translational energy to internal energy, most often vibrational excitation. This vibrational energy is rapidly (picosecond to nanosecond time frame) distributed throughout all bonds in the molecule. Eventually, when the vibrational energy of the ion exceeds the activation barrier required for bond cleavages and chemical rearrangements, the molecular ion is converted to smaller fragments. When internal energies are close to the fragmentation threshold, the resulting fragments are often limited to neutral losses that are common to too many biomolecules, for example loss of neutral H_2O. The optimal collision energy for generating fragments may vary among different molecules. Hence, to obtain more characteristic fragments and expedite the large-scale metabolite screening of crude samples, multiplexed CID, as described below, has been employed in these investigations.

Multiplexing of non-selective CID performed using quasi-simultaneous acquisition of mass spectra with multiple collision energies allows for detection of both molecular and fragment ions obtained in a single LC/MS analysis if the duty cycle of ion mass analysis is sufficiently fast. Energy-resolved yields of fragment ions can be obtained by rapidly switching among 4-5 different predetermined CID voltages applied at Aperture 1 on the chromatographic time scale (Figure 1.3). The switching between the different voltages can be accomplished within 10 ms with data for each CID condition stored in separate data acquisition functions within a single data file [45]. Since many modern time-of-flight mass spectrometers acquire spectra at a rate of about 20,000 transient spectra per second, acquisition of signal over 10 ms is still long enough to acquire about 200 mass spectra.

Multiplexed CID offers several advantages over tandem MS with data-dependent acquisition with regards to non-targeted profiling of metabolites. For example, the latter technique often necessitates slow gradient elution because low abundant metabolites that co-elute with more abundant ones may not be selected for MS/MS analysis, resulting in loss of fragmentation information [46]. Furthermore, the MS/MS scan uses a fixed set of rules that control collision energy to induce fragmentation. This may not be ideal if the sample contains a diverse range of metabolites that have a diverse range of fragmentation energy thresholds. Therefore, multiplexed CID becomes a more favorable technique for the analysis of complex mixtures since it is not dependent on a single set of CID parameters and is reproducible from sample to sample. The multiple CID conditions applied will generate fragment ion spectra of all metabolites and eliminate compound-dependent optimization and additional analysis. Moreover, Dynamic Range Enhancement (DRE) function works to extend the linear dynamic range of the instrument when high ion counts saturate the detector or ion detection electronics [47].

1.5.3. Time-of-flight mass detector

The detection of precursor ions and their fragments produced from the combination of HPLC/ESI/CID is commonly coupled with a time-of-flight (TOF) detector owing to inherent characteristics such as rapid acquisition capabilities, high sensitivity, high resolution and nearly unlimited mass range. These factors are especially important when speeding up large scale routine profiling experiments involving crude mixtures of metabolites [48]. In TOF/MS, ions are accelerated into a field free flight tube by a pulsed electric field and a mass spectrum is generated by measuring the signal as a function of time taken for the ions to reach the detector. Since all the ions are accelerated through approximately the same electric potential drop, the lighter ions arrive at the detector earlier than heavier ions that have the same charge. Many

modern TOF/MS are equipped with a reflectron which performs kinetic energy focusing of ions with the same *m/z* but with slight differences in kinetic energies, thereby increasing the resolution and mass accuracy of the instrument. Moreover, since in principal, all ions travelling through the flight tube within the desired mass range are detected, the TOF MS can provide extremely high sensitivity compared to other mass analyzers [49].
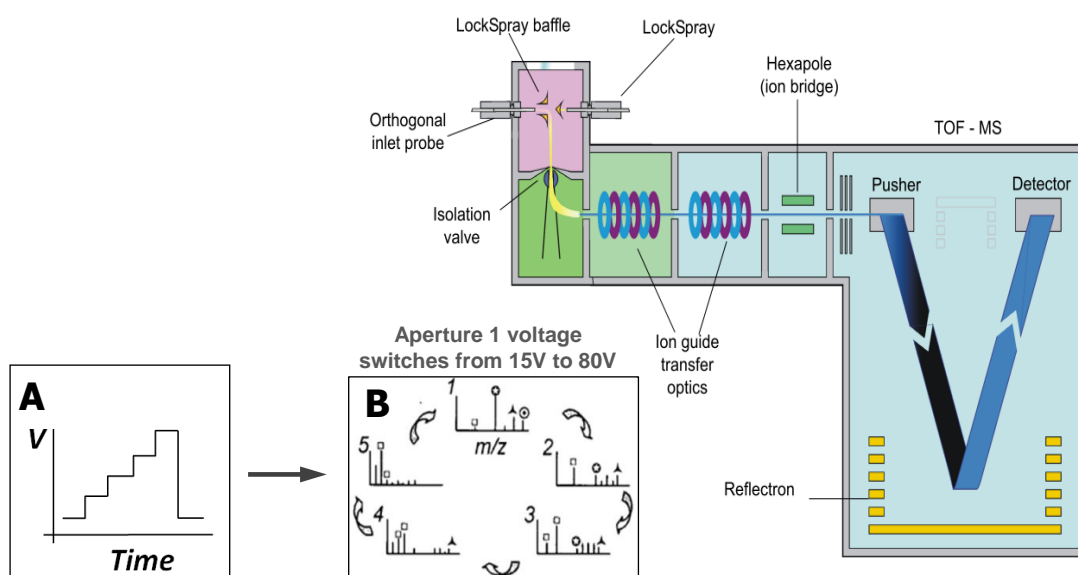
Figure 1.4. Schematic of a Waters LCT Premier™ TOF mass spectrometer used with multiplexed CID showing (A) Aperture 1 voltage stepped across 5 functions and (B) mass spectra acquired using 5 different quasi-simultaneous CID voltages. Figure is adapted from Waters LCT Users' Guide. (Adapted from the Waters LCT Premier user's manual)

1.5.4. Nuclear Magnetic Resonance Spectroscopy
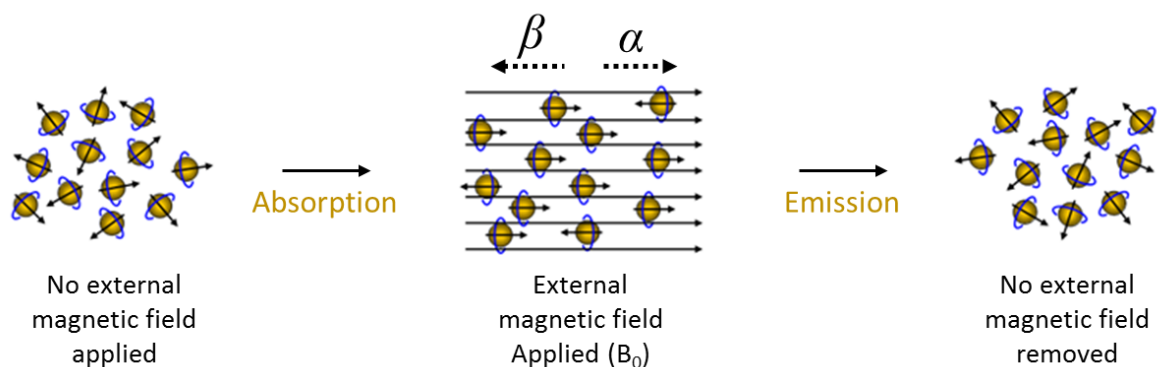
Figure 1.5. Nuclei that have random nuclear spin align to either an alpha-spin state (lower energy) or beta-spin state (higher energy) when placed in an external magnetic field. When the field is removed, the nuclei return to random orientations [50].

Nuclear magnetic resonance (NMR) analyses can provide important structural information about compounds in a given sample, and NMR is widely used in analytical chemistry and biological sciences [51]. NMR exploits the interaction of radio frequency (RF) electromagnetic radiation with elements that have non-zero nuclear spin quantum number in an externally applied magnetic field. Specifically, when a population of magnetic nuclei is exposed to an external magnetic field (B$_0$), the nuclei orient themselves in a very predictable manner. For magnetic isotopes such as ^{1}H, there are two orientations, one of which aligns with the applied external field (*alpha*–spin state), and the other is aligns against the magnetic field (*beta*-spin state) (Figure 1.4). Due to the nuclei acting as very weak atomic magnets, the energy difference between the alpha and beta orientations is small relative to kT, allowing the nuclei to switch between the two orientations at room temperature. At any given moment, there is a slight excess in the population of the alpha (lower energy) versus the beta state (higher energy), and energy from a RF pulse with field perpendicular to B$_0$ can be used to pump these excess alpha state nuclei into the beta state. When the energy supply is paused, the energized nuclei will relax back to their original alpha orientation. This type of relaxation causes a fluctuation

in the magnetic field known as resonance which can be detected and plotted into peaks that are typically seen in an NMR spectrum.

The local environment of a proton can alter the magnetic field strength experienced by the proton. Since the magnetic field strength determines the energy separation of the two spin states, and hence the resonance signal of the RF, all the protons within the same molecule end up with a distribution of resonance signals. Some factors that cause the protons to have different local environments include neighboring electronegative groups, hydrogen bonding and non-uniform magnetic field. Therefore, the NMR spectrum is a plot of the RF input against the absorption. The chemical shift range reported on the spectrum is defined by the frequency of the resonance relative to the frequency of a reference in the magnetic field.   [51, 52].

One- and two-dimensional homo- and heteronuclear NMR methods are the most useful in structural elucidation of novel compounds in natural product research [52, 53]. The 1D-NMR involves detection of proton (^{1}H) and carbon (^{13}C) resonances that are generally used for probing the structures of simpler molecules. Resonances are spread over a range of about 10 ppm for ^{1}H 1-D NMR spectra, but for complex compounds, more resolution of small differences in resonance energy is required to minimize overlap of different resonance signals. Fortunately, 2-D NMR increases the discrimination of resonances by dispersing the signals into a second dimension. On the other hand, due to the low natural abundance (1.1%) of ^{13}C, ^{13}C 1-D NMR suffers from low sensitivity, resulting in the need for higher sample concentration and longer signal averaging times. Hence, 2-D NMR can be used to detect carbon indirectly to improve sensitivity. Some examples of 2-D NMR methods include HSQC (heteronuclear single quantum coherence), cHSQC (coupled HSQC), HMBC (heteronuclear multiple bond correlation), and COSY (homonuclear correlated spectroscopy). The peaks seen in a 2-D HSQC spectra show correlations between two nuclei (^{1}H-^{13}C) that are one bond apart. This enables elucidation of the any protons that are directly attached to a particular carbon. The

cHSQC experiment is similar but is able to provide the J-value for ^{1}H-^{13}C couplings. HMBC spectra show long-range couplings between protons and carbons over 2 to 3 bonds. It is especially useful in deciphering structures of compounds that do not have a proton attached to every carbon. The spectra from a COSY experiment plots frequencies of ^{1}H along both axis, allowing for elucidation of the coupling relationships between protons on each of the two carbons that are separated by a single bond [52]. The use of these methods is discussed further in Chapter 4.

1.5.4.1. Advantages of using NMR spectroscopy in natural product analysis

Although the LC/MS system using multiplexed CID can show the difference in retention times of isomers, it often cannot differentiate isomers unless they fragment to generate fragment ions of different masses or different abundances. Furthermore, CID usually does not break every chemical bond, so mass spectrometry does not reveal the full differences in connectivity between compounds. It is, therefore, useful to employ NMR techniques to determine the connectivity between atoms. NMR techniques provide a universal detection method for all molecules with NMR-active nuclei, unlike MS where detection of analytes is dependent on ionization conditions (e.g. compounds only are detected using ESI if they have acidic or basic functional groups or if they have affinities for anions or cations). Commonly, the fractions collected from a semi-preparative LC system may furnish only microgram to low milligram amounts of purified metabolites. The analysis of such small amounts can be performed using a high field NMR spectrometer such as the Bruker Avance 900 (21.1 T) spectrometer at the Max T. Rogers NMR Facility at Michigan State University. Such high magnetic fields yield best results when samples are prepared in Shigemi microscale NMR tubes [54] which are made of materials with magnetic susceptibility matched to suitable solvents.

1.6. Challenges in metabolite quantification

Aside from detection and purification, quantification of products and intermediates plays an important role in elucidating the biosynthetic pathway of a metabolite. The metabolite amounts present or absent in different tissues of the plants often provide additional clues about the cryptic biological processes that produce these metabolites, particularly when this information is evaluated in the context of global gene expression (transcriptomics) of individual tissues. Unfortunately, metabolite quantification can be challenging because ionization efficiency in mass spectrometric analysis is not only dependent on analyte concentration but also on other factors including chemical structure, and the presence of other analytes in the matrix that compete for charge and suppress analyte ionization. There are two strategies that avoid the aforementioned problems; (1) relative quantification which is often used in non-targeted large scale metabolite profiling where the metabolite signal intensity is normalized to the signal of an internal standard with known concentration and, (2) absolute quantification used mostly in targeted metabolomics and involves the use of external standards or stable isotope-labeled standards [40] when they are available. In the case of many plant specialized metabolites, not only are standards not available, but the identities of the metabolites are often largely unknown.

1.7. Summary of research

The objectives of the research described in this dissertation are driven by the need to understand previously intractable, "non-medicinal" medicinal plants so as to have improved access to the bioactive natural products. The diversity in plant natural products is remarkable, paving the way for discovery of thousands of novel compounds that could advance human health. However, while this diversity holds great potential for treating human diseases, it also presents a challenge for researchers looking to identify and manipulate the enzymes that are part of these specialized metabolic pathways. Furthermore, the study of pharmaceutically

relevant metabolites and their metabolic pathways has been restricted to only a handful of species that are poorly characterized at the molecular and genetic levels. As a result, researchers have not met with great success in dissecting individual specialized metabolic pathways beyond model plants such as *Arabidopsis thaliana*. Many current knowledge gaps are related to the tedious and time consuming process of isolating and purifying compounds of interest from the myriad of phytochemicals present in plant tissues. Recent advances in unraveling the cryptic biochemical pathways that yield important drugs have begun with redesigning the pathway in microbial hosts such as yeast, which allows sufficient quantities of active proteins and substrates to be generated and studied. However, the biggest bottleneck in this process lies in assigning functions to the cDNAs (complementary deoxyribonucleic acids) that encode for specific steps in the biosynthetic pathway, as this often needs *in vivo* evidence (e.g. changes in metabolite profiles that result from silencing a specific gene) and *in vitro* evidence of biochemical functions (e.g. enzyme activity toward a specific substrate). Hence, in order to recreate the entire pathway successfully in a heterologous host, the individual enzymatic steps and the metabolite transformations that occur have to first be deciphered to demonstrate function of a specific gene product. The work presented in this dissertation deals with the analysis of the metabolites that are proposed to be involved in the biosynthetic pathway of camptothecin.

To address the challenges detailed above, Chapter 2 describes the methods for generating stable isotope-labeled metabolites for evidence of metabolite structure and dynamics, including bottlenecks in the progression of the intermediates along the pathway at different stages of plant development. *C. acuminata* seedlings at 5- and 10-weeks post germination were grown under $^{13}CO_2$-enriched atmosphere for 2 weeks. The chapter presents data from chromatographic separation and mass spectrometric profiling methods for the analysis of the crude extracts obtained from tissues harvested every 2 days during the experiment. The

discoveries in this chapter underscore the importance of the developmental stage of the plant when performing metabolomics studies.

The first section of Chapter 3 presents identification and characterization of known and unknown camptothecin pathway intermediates by profiling the root, stem, shoot apex, young leaf and mature leaf tissues of *C. acuminata*. Based on the presence of strictosidinic acid isomers, and absence of strictosidine, it has been concluded that an alternative seco-iridoid pathway is more likely to be taking place in this plant, rather than the previously postulated pathways for the production of camptothecin. The results also demonstrated that perhaps all glycosylated camptothecin pathway intermediates are present as multiple isomers. The second section of Chapter 3 presents HPLC/MS analysis of wild-type and RNAi lines of *C. acuminata* to assess the impact of silencing of two candidate biosynthetic genes. The rationale behind this set of experiments was to find the genes responsible for specific steps in the biosynthetic route. It was observed that the production of indole alkaloids was impaired in plants with knocked out cyclase gene, and indole alkaloids were dramatically reduced in level after knockout of tryptophan decarboxylase.

Chapter 4 includes results from the isolation and purification of two important isomers, strictosidinic acid, involved in the biosynthetic pathway of camptothecin. Since the work in Chapter 3 defined the presence of isomeric metabolites, it was necessary to determine the exact point of difference in the structures of the isomers. By pinpointing the difference in stereochemistry of attachment of the glucose moiety in glycosylated pathway intermediates, it was concluded that the isomerism takes place early in the MIA pathway, and it is likely that the isomers are substrates to the same enzyme or multiple enzymes are present at each step.

Chapter 5 includes concluding remarks and future perspectives for the questions that remain unanswered.

REFERENCES

REFERENCES

1. Oeppen, J. and J.W. Vaupel, *Demography. Broken limits to life expectancy.* Science, 2002. **296**(5570): p. 1029-31.

2. Pichersky, E. and D.R. Gang, *Genetics and biochemistry of secondary metabolites in plants: an evolutionary perspective.* Trends Plant Sci, 2000. **5**(10): p. 439-45.

3. Schwab, W., *Metabolome diversity: too few genes, too many metabolites?* Phytochemistry, 2003. **62**(6): p. 837-49.

4. Pichersky, E., *The validity and utility of classifying genes as 'essential'.* Trends Plant Sci, 2009. **14**(8): p. 412-3; author reply 414-5.

5. Verpoorte, R., *Exploration of nature's chemodiversity: the role of secondary metabolites as leads in drug development.* Drug Discovery Today, 1998. **3**(5): p. 232-238.

6. Iriti, M. and F. Faoro, *Chemical diversity and defence metabolism: how plants cope with pathogens and ozone pollution.* Int J Mol Sci, 2009. **10**(8): p. 3371-99.

7. Zhang, W., et al., *To Stretch the Boundary of Secondary Metabolite Production in Plant Cell-Based Bioprocessing: Anthocyanin as a Case Study.* Journal of Biomedicine and Biotechnology, 2004. **2004**(5): p. 264-271.

8. Hartmann, T., *From waste products to ecochemicals: fifty years research of plant secondary metabolism.* Phytochemistry, 2007. **68**(22-24): p. 2831-46.

9. Hogan Jr, J.C., *Combinatorial chemistry in drug discovery.* Nature Biotechnology, 1997. **15**(4): p. 328-330.

10. Cragg, G.M. and D.J. Newman, *Plants as a source of anti-cancer agents.* Journal of Ethnopharmacology, 2005. **100**(1–2): p. 72-79.

11. Balunas, M.J. and A.D. Kinghorn, *Drug discovery from medicinal plants.* Life Sci, 2005. **78**(5): p. 431-41.

12. Newman, D.J. and G.M. Cragg, *Natural Products as Sources of New Drugs from 1981 to 2014.* J Nat Prod, 2016. **79**(3): p. 629-61.

13. Westfall, P.J., et al., *Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin.* Proc Natl Acad Sci U S A, 2012. **109**(3): p. E111-8.

14. Minami, H., et al., *Microbial production of plant benzylisoquinoline alkaloids.* Proc Natl Acad Sci U S A, 2008. **105**(21): p. 7393-8.

15. Ajikumar, P.K., et al., *Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli.* Science, 2010. **330**(6000): p. 70-74.

16.   Watanabe, K., et al., *Total biosynthesis of antitumor nonribosomal peptides in Escherichia coli.* Nat Chem Biol, 2006. **2**(8): p. 423-8.

17.   O'Connor, S.E. and J.J. Maresh, *Chemistry and biology of monoterpene indole alkaloid biosynthesis.* Nat Prod Rep, 2006. **23**(4): p. 532-47.

18.   Wall, M.E., et al., *Plant Antitumor Agents. I. The Isolation and Structure of Camptothecin, a Novel Alkaloidal Leukemia and Tumor Inhibitor from Camptotheca acuminata1,2.* Journal of the American Chemical Society, 1966. **88**(16): p. 3888-3890.

19.   Hsiang, Y.H., et al., *Camptothecin induces protein-linked DNA breaks via mammalian DNA topoisomerase I.* J Biol Chem, 1985. **260**(27): p. 14873-8.

20.   Creemers, G.J., et al., *Topotecan, an active drug in the second-line treatment of epithelial ovarian cancer: results of a large European phase II study.* Journal of Clinical Oncology, 1996. **14**(12): p. 3056-61.

21.   Fuchs, C., E.P. Mitchell, and P.M. Hoff, *Irinotecan in the treatment of colorectal cancer.* Cancer Treatment Reviews, 2006. **32**(7): p. 491-503.

22.   Lorence, A. and C.L. Nessler, *Camptothecin, over four decades of surprising findings.* Phytochemistry, 2004. **65**(20): p. 2735-2749.

23.   Kulkarni, A.V., et al., *Production of camptothecin in cultures of Chonemorpha grandiflora.* Pharmacognosy Research, 2010. **2**(5): p. 296-299.

24.   Lorence, A., F. Medina-Bolivar, and C.L. Nessler, *Camptothecin and 10-hydroxycamptothecin from Camptotheca acuminata hairy roots.* Plant Cell Rep, 2004. **22**(6): p. 437-41.

25.   Lopez-Meyer, M., C.L. Nessler, and T.D. McKnight, *Sites of accumulation of the antitumor alkaloid camptothecin in Camptotheca acuminata.* Planta Med, 1994. **60**(6): p. 558-60.

26.   Kai, G.-Y., et al., *In vitro plant regeneration from leaf explants of Ophiorrhiza japonica.* Biologia Plantarum, 2008. **52**(3): p. 557-560.

27.   Gottlieb, J.A. and J.K. Luce, *Treatment of malignant melanoma with camptothecin (NSC-100880).* Cancer Chemother Rep, 1972. **56**(1): p. 103-5.

28.   Saito, K. and F. Matsuda, *Metabolomics for functional genomics, systems biology, and biotechnology.* Annu Rev Plant Biol, 2010. **61**: p. 463-89.

29.   Ruzicka, L., *The isoprene rule and the biogenesis of terpenic compounds.* Experientia, 1953. **9**(10): p. 357-67.

30.   Wang, K.C. and S.-i. Ohnuma, *Isoprenyl diphosphate synthases.* Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, 2000. **1529**(1–3): p. 33-48.

31.   Rohmer, M., et al., *Isoprenoid biosynthesis in bacteria: a novel pathway for the early steps leading to isopentenyl diphosphate.* Biochem J, 1993. **295 (Pt 2)**: p. 517-24.

32.   Dudareva, N., et al., *The nonmevalonate pathway supports both monoterpene and sesquiterpene formation in snapdragon flowers.* Proceedings of the National Academy of Sciences of the United States of America, 2005. **102**(3): p. 933-938.

33.   Nagegowda, D.A., *Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation.* FEBS Lett, 2010. **584**(14): p. 2965-73.

34.   Kutchan, T.M., *Alkaloid Biosynthesis[mdash]The Basis for Metabolic Engineering of Medicinal Plants.* The Plant Cell, 1995. **7**(7): p. 1059-1070.

35.   De Luca, V., C. Marineau, and N. Brisson, *Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases.* Proc Natl Acad Sci U S A, 1989. **86**(8): p. 2582-6.

36.   Miettinen, K., et al., *The seco-iridoid pathway from Catharanthus roseus.* Nat Commun, 2014. **5**: p. 3606.

37.   Irmler, S., et al., *Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase.* Plant J, 2000. **24**(6): p. 797-804.

38.   Dixon, R.A. and D. Strack, *Phytochemistry meets genome analysis, and beyond.* Phytochemistry, 2003. **62**(6): p. 815-6.

39.   Sumner, L.W., P. Mendes, and R.A. Dixon, *Plant metabolomics: large-scale phytochemistry in the functional genomics era.* Phytochemistry, 2003. **62**(6): p. 817-836.

40.   Lei, Z., D.V. Huhman, and L.W. Sumner, *Mass spectrometry strategies in metabolomics.* J Biol Chem, 2011. **286**(29): p. 25435-42.

41.   Kruger, N.J., M.A. Troncoso-Ponce, and R.G. Ratcliffe, *1H NMR metabolite fingerprinting and metabolomic analysis of perchloric acid extracts from plant tissues.* Nat. Protocols, 2008. **3**(6): p. 1001-1012.

42.   Gaskell, S.J., *Electrospray: Principles and Practice.* Journal of Mass Spectrometry, 1997. **32**(7): p. 677-688.

43.   Clegg, G.A. and M. Dole, *Molecular beams of macroions. 3. Zein and polyvinylpyrrolidone.* Biopolymers, 1971. **10**(5): p. 821-6.

44.   Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

45.   Gu, L., A.D. Jones, and R.L. Last, *Broad connections in the Arabidopsis seed metabolic network revealed by metabolite profiling of an amino acid catabolism mutant.* Plant J, 2010. **61**(4): p. 579-90.

46. Chakraborty, A.B., S.J. Berger, and J.C. Gebler, *Use of an integrated MS--multiplexed MS/MS data acquisition strategy for high-coverage peptide mapping studies.* Rapid Commun Mass Spectrom, 2007. **21**(5): p. 730-44.

47. Belov, M.E., et al., *Dynamic range expansion applied to mass spectrometry based on data-dependent selective ion ejection in capillary liquid chromatography fourier transform ion cyclotron resonance for enhanced proteome characterization.* Anal Chem, 2001. **73**(21): p. 5052-60.

48. Cotter, R.J., et al., *Tandem Time-of-Flight Mass Spectrometry with a Curved Field Reflectron.* Analytical Chemistry, 2004. **76**(7): p. 1976-1981.

49. Ryan, D. and K. Robards, *Analytical Chemistry Considerations in Plant Metabolomics.* Separation & Purification Reviews, 2006. **35**(4): p. 319-356.

50. Marquez, B.L., W.H. Gerwick, and R. Thomas Williamson, *Survey of NMR experiments for the determination of nJ(C,H) heteronuclear coupling constants in small molecules.* Magnetic Resonance in Chemistry, 2001. **39**(9): p. 499-530.

51. Breton, R.C. and W.F. Reynolds, *Using NMR to identify and characterize natural products.* Natural Product Reports, 2013. **30**(4): p. 501-524.

52. Friebolin, H. and A.M. Kenwright, *Basic one- and two-dimensional NMR spectroscopy. VCH, Weinheim, 1993, 2nd edition, pp. 368, £24, DM58.* Magnetic Resonance in Chemistry, 1994. **32**(7): p. 439-439.

53. Reynolds, W.F. and R.G. Enríquez, *Choosing the Best Pulse Sequences, Acquisition Parameters, Postacquisition Processing Strategies, and Probes for Natural Product Structure Elucidation by NMR Spectroscopy.* Journal of Natural Products, 2002. **65**(2): p. 221-244.

54. Takeda, M., et al., *Construction and performance of an NMR tube with a sample cavity formed within magnetic susceptibility-matched glass.* Journal of magnetic resonance (San Diego, Calif. : 1997), 2011. **209**(2): p. 167-173.

Chapter 2: Deciphering Bottlenecks in the Biosynthesis of Camptothecin in *Camptotheca acuminata* using ¹³C Labeling

2.1. Introduction

Despite increasing demand of the monoterpene indole alkaloid (MIA) camptothecin in today's pharmaceutical market [1], the world's supply of the drug still relies on direct extraction and isolation from intact plants [2]. Knowledge regarding its biosynthesis in *Camptotheca acuminata* remains limited, even though it has been more than half a century since the alkaloid was first isolated and shown to have pharmacological value [3]. The delay in the elucidation of the pathway can be attributed to the complex structure of the final drug product and lack of information about metabolic intermediates in the pathway. Due to the long wait times for slow-growing medicinal plants, scientists have been keen on genetically engineering plants for the production of specialized metabolites [4]. Some examples include exploiting plant cell cultures [5] to aid in higher efficiency in the production of the drug and/or its intermediates.

Plant cell cultures have been a target for the production of pharmaceutically important plant metabolites for several decades. Yet, only a few plant-based bioactive compounds have reached commercial production scale, for example, the prenylated naphthoquinone shikonin [6] and paclitaxel [7], but not camptothecin. Nevertheless, camptothecin has at least been shown to accumulate in transformed hairy root cultures of *C. acuminata* with accumulations levels equal to 2.5 µg/g dry weight as first reported by Sakato [8]. Recent studies by Lorence have yielded 1.0 mg/g dry weight [9]. Unfortunately, whole plant extraction still affords higher metabolite levels (up to 5 mg/g dry weight) than plant cell cultures [10].

Another alternative to isolating drug products from low-yielding medicinal plant tissues involves the transfer of the synthetic machinery for specialized metabolite biosynthesis to heterologous hosts [11-13]. With the convergence of engineering disciplines, biotechnology, life and physical sciences, new solutions have become easier to generate for increasing the productivity

of natural products. For example, attention has been given to implementation of heterologous microbial hosts as a substitute for the production of camptothecin [14]. The degree of gene complexity is significantly lowered since microorganisms have a much smaller genome size; for example, plants have an average of 25,000 genes [15] whereas yeast has fewer than 6000 genes and is the first eukaryotic organism to have its genome completely sequenced [16]. Another significant advantage of microbial hosts is the smaller number of intracellular organelles that make metabolite compartmentalization negligible [17], though sequestering intermediates in subcellular compartments may help the host avoid toxicity of metabolic intermediates. Also, large-scale cultivation of microbial cells is much easier to handle as they are less sensitive to shear stress compared to plants and have shorter doubling times which leads to faster production rates. Moreover, microbes can be grown in inexpensive synthetic media which are easy to make and are the site of metabolite accumulation since cells typically secrete them out into the surrounding medium, unlike the intracellular accumulation of metabolites in different parts of the plant [18].

Therefore, using the wealth of scientific knowledge and tools accumulated over the last century, yeast can be used as a powerful microorganism for heterologous expression of the biosynthetic pathway of camptothecin. However, the exploitation of microbial cells for production of metabolites suffers from the lack of complete knowledge of the biosynthetic enzymes responsible for metabolite transformations. Huge gaps of information about intermediates and enzymatic reactions involved in proposed biosynthetic pathways have limited the large-scale production of camptothecin through the development of a scalable production platform in heterologous hosts. Even though enzymes known to catalyze rate-controlling steps have been overexpressed in the plants, the expected increase in production has not been observed [19]. Hence, it is imperative to foremost identify the specific enzymes involved in the biosynthetic pathway of

camptothecin. In order to hunt for the enzymes responsible for the transformations of intermediates to drug product, it is first necessary to demystify the intermediates involved in the pathway, which can provide clues and help narrow down the class of candidate enzymes that should be targeted.

It is necessary to purify the intermediates and utilize them as substrates for testing the enzymes in question, but recognition of pathway intermediates in crude plant extracts can be challenging owing to the enormous number of plant specialized metabolites and the limited number of metabolites with defined chemical structures. Use of stable isotope-labeled precursors can facilitate structure elucidation and identify metabolic bottlenecks that limit formation of desired products. Therefore, it is essential to identify the important developmental stages where metabolite biosynthetic fluxes could be substantial at specific steps within the biosynthetic pathway.

The rapidly growing field of metabolomics with recent advances in bioanalytical platforms, including gas chromatography (GC), liquid chromatography (LC), mass spectrometry (MS) and nuclear magnetic resonance (NMR), and the development of computational data analysis capabilities and organism-specific metabolite databases that are free for the public, has allowed for accurate interpretation, quantification and characterization of metabolites in whole biological systems [20, 21]. Even though not all metabolites can be detected, an analyst is still able to collect highly informative data using current technologies. However, the shortcoming of many metabolomics methods is the analysis of only a snapshot view of the metabolome depending on the time and stage of development of the plant when it was harvested for experiments [20]. The dynamic metabolic fluxes are not taken into account, hence, for example, the increased concentrations of metabolites cannot definitively indicate whether it is the result of over expression of an enzyme, transportation of the metabolite into the tissue being analyzed or slow degradation of the metabolite at specific growth conditions and time points. In addition, crosstalk between

various metabolic processes within the plant presents an added level of complexity when proposing a hypothesis based on levels of metabolites reflecting steady-state, rather than dynamic fluxes. Therefore, the study of only the steady-state metabolite levels is inadequate for probing the rates of synthesis and turnover reactions of specialized metabolites [22].

Stable isotopic labeling is a remarkable way to trace the transformation of intermediates within the plant. ^{13}C labeling of plants for the study of the central carbon metabolism is most amenable with techniques such as MS[23] and NMR [24]. The techniques provide metabolite characterization and positional information for the isotopically labeled precursors, which in turn, provide significant advantages for tracking the intermediates steps in a biosynthetic pathways. A ^{13}C labeled precursor can easily be distinguished from its unlabeled counterpart in MS analysis based on the mass difference and change in isotopic abundance [25].

In this study, the label is introduced into the plant in the form of labeled CO_2 that is originally added to the experiment as sodium [^{13}C] bicarbonate. The carbon in the carbon dioxide is fixed through biochemical reactions such as photosynthesis and the Calvin cycle, eventually producing carbohydrates. The products of metabolism of these carbohydrates result in the formation of carbon-rich building blocks that go into the biosynthetic pathway leading to the formation of primary and secondary metabolites (Figure 2.1). For tracing the production of camptothecin in *C. acuminata*, focus will be given to the metabolites involved in the production of monoterpene indole alkaloids.
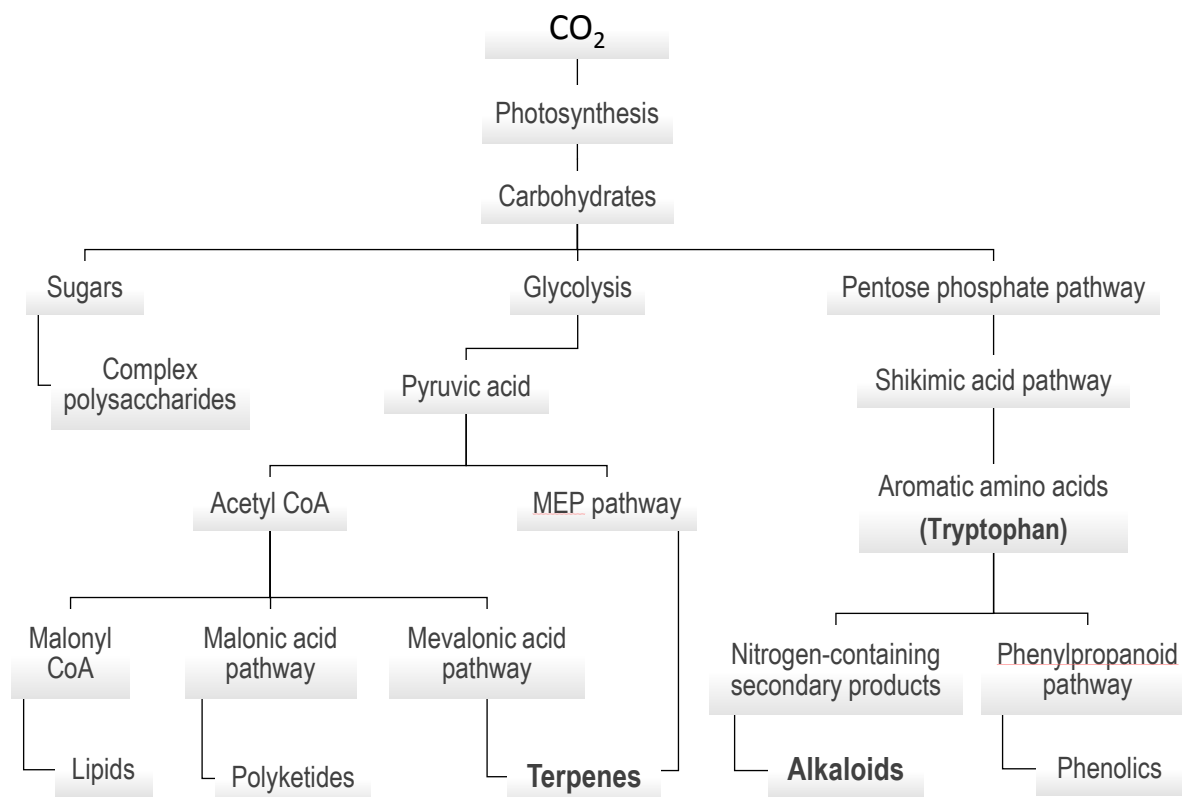
Figure 2.1. Biosynthetic pathway of different specialized metabolites. The chart indicates the intake and transport of carbon from carbon dioxide and into the complex structures of specialized metabolite.

2.2. Materials and methods

2.2.1. Reagents

ACS grade sulfuric acid (95%−98%, J.T. Baker), and HPLC grade methanol and formic acid (88% aqueous solution) were purchased from VWR Scientific. [^{13}C]Sodium bicarbonate (98 atom % ^{13}C) was purchased from Sigma-Aldrich (St. Louis, MO).

2.2.2. Labeling of plants

10-week old and 5-week old seedlings of *Camptotheca acuminata* were obtained from the laboratory of Professor Dean DellaPenna at Michigan State University. Seedlings were divided into four groups and placed into glass chambers, including one set of unlabeled control group. The

7.5 L glass chambers were placed inside a larger growth chamber at 26°C and provided with 24 hours/day light at 100-150 $\mu E\text{-}m^{-2}\text{-}s^{-1}$. They were grown in autoclaved vermiculite containing ½ strength Hoagland solution that was refreshed every 2 days for a 16-day period in total. H_2SO_4 and $NaH^{13}CO_3$ were placed separately into each of the glass chambers and only allowed to come into contact with each other after the lids had been sealed. This allowed for the introduction of $^{13}CO_2$ into the airtight glass chamber to achieve a final concentration of 400 ppm. Unlabeled control plants were treated similarly by growing them in the same growth chamber under ambient atmosphere but without adding labeled $NaH^{13}CO_3$. A packet of ethylene scavenger based on potassium permanganate were added to each glass chamber.

2.2.3. Sample preparation

Freshly-harvested leaf tissue (10-50 mg) from each plant was collected every 48 hours for 2 weeks from the second set of leaves from the apex of the plant. Tissue specimens were immediately frozen in liquid nitrogen after harvest, and metabolites were extracted using 400 µl of 70% acetonitrile containing 5 µM telmisartan as internal standard. After a 16-hour incubation period, extracts were centrifuged at 5000g for 15 min at 4 ˚C, and supernatants were transferred to autosampler vials and analyzed without further processing.

2.2.4. LC/MS metabolite profiling

Liquid chromatography/time-of-flight (LC/TOF) mass spectrometry and electrospray ionization is used to identify metabolites and measure isotopic enrichments using a gradient that was selected based on separation of key metabolites. Metabolites from leaf tissue were resolved using HPLC separation using an Ascentis Express C18 column (10 cm×2.1 mm; 2.7 µm, Supelco)

involving gradient elution based upon 10 mM ammonium formate (solvent A) and methanol (solvent B) over 52-min and 31-min periods at a flow rate of 0.3 mL/min. For both gradients, the column temperature used was 45 °C. Figure 2.2 C shows the workflow that was used to perform the LC/MS analysis. The various symbols within the test tube of crude mixture represents the different metabolites of interest and how they are individually separated and analyzed for isotopic enrichments using a Shimadzu LC-20AD ternary pump coupled to a SIL-5000 autosampler, column oven, and Waters LCT Premier mass spectrometer equipped with an electrospray ionization source for the 52-min method. UHPLC separation was performed on a Xevo G2-S QTOF mass spectrometer (Waters, Milford, MA) for the 31-min method. A 10-µL volume of each extract was analyzed using either a 52-min or a 31-min gradient elution method on an Ascentis Express C18 column (10 cm×2.1 mm; 2.7 µm) with mobile phases consisting of 10 mM ammonium formate in water, adjusted to pH 2.85 with formic acid (solvent A) and methanol (solvent B). The 52-min method gradient was as follows: 2% B at 0.00-2.00 min, linear gradient to 20% B at 20.00 min, linear gradient to 55% B at 43.00 min, then a step to 99% B at 43.01 min, then return to 2% B over 47.01-52.00 min. The 31-min method employed an initial solvent composition of 0% B followed by a linear gradient to 25% B at 10.00 min, 99% B at 24.00-27.00 min, then step to 99% B at 28.00 min and held until 31.00 min. For both methods, the flow rate was 0.3 mL/min and the column temperature was 45 °C. The LCT Premier mass spectrometer was operated using V optics in positive-ion mode with a typical mass resolution (M/ΔM) of ~ 4000 at full width at half maximum. Source parameters were as follows: capillary voltage 3200 V, sample cone voltage 10 V, desolvation temperature 350 °C, source temperature 100 °C, cone gas flow 40 L/h and desolvation gas flow 350 L/h. Mass spectrum acquisition was performed over m/z 50 to 1,500 with scan time of 0.1 s for each acquisition function, using dynamic range extension. Mass

spectra containing fragment ions were generated by rapid switching of aperture 1 voltage over four parallel data acquisition functions (20, 40, 60 and 80 V) [26]. The Xevo G2-S QTOF was operated using sensitivity mode in positive-ion mode, and mass spectra were acquired over m/z 50-1500 using centroid peak acquisition and extended dynamic range in resolution mode which gave resolution ~ 23000 at full width half maximum. Source parameters were as follows: capillary voltage 2.14 kV, sample cone voltage 35 V, source temperature at 90 °C, desolvation gas flow 800 L/h, desolvation gas temperature 280 °C, cone gas flow 0 L/h. Mass spectra containing fragment ions were generated using 0, 20, 35, 55, and 80 V collision cell potentials using argon as collision gas.

2.3. Results and discussion

2.3.1. Design of whole-plant labeling with isotopically labeled carbon dioxide
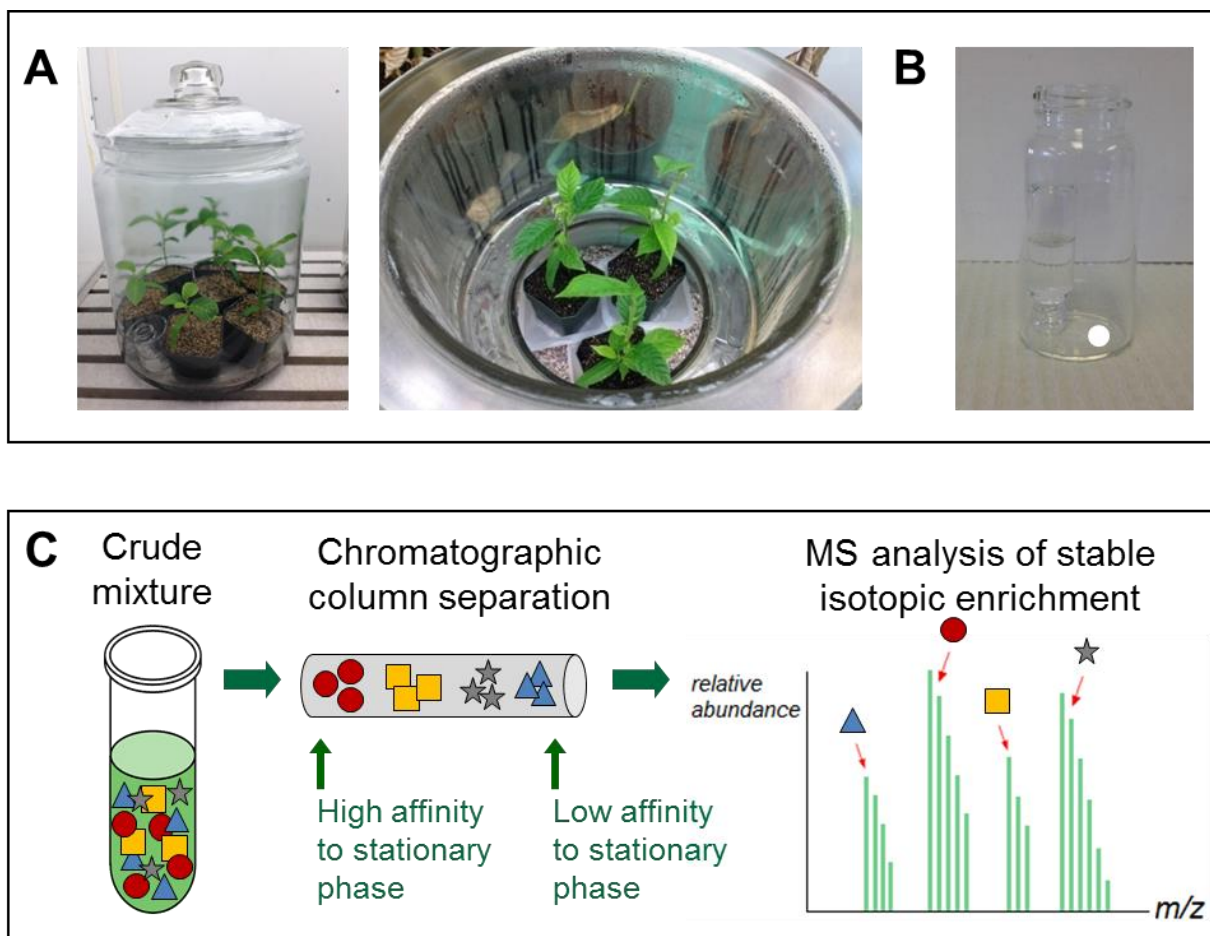
Figure 2.2. Workflow of the stable isotope whole plant labeling process. (A) *Camptotheca acuminata* seedlings were housed inside airtight glass chambers throughout the labeling experiment. (B) For each glass chamber, 1 mL of 97% H_2SO_4 was added into a 2 mL glass vial and placed upside down into a 20 mL glass vial containing 20.8 mg of sodium [^{13}C] bicarbonate (as noted by the white dot), and gently inserted into the glass chamber and allowed to mix by gently shaking the apparatus only after the lid on the glass chamber had been closed. (C) Crude extracts obtained from collected leaf tissue were subjected to a C18 column for metabolite separation, and then subjected to mass spectrometric analysis for quantification of isotopic enrichment.

The following details the efforts made to maximize the mol % labeling of metabolites extracted

from plant tissues in this study. *Camptotheca acuminata* seedlings were grown in autoclaved

vermiculite instead of soil to reduce the intrinsic pool of unlabeled carbon dioxide by limiting bacteria from converting any soil carbon into unlabeled carbon dioxide. The seedlings were introduced into the glass chamber at 5 and 10 week old post-germination to limit the pools of unlabeled carbon in each plant. Leaflets were harvested from the second set of leaves from the apex of the plant to minimize dilution of the newly incorporated ^{13}C into leaf tissue by pools of unlabeled metabolites. In addition, the young plants were still small enough until 12 weeks post-germination to fit into the air-tight glass chambers that were used for these experiments (Figure 2.2 A). In order to avoid losing $^{13}CO_2$ while opening the glass chambers to introduce more $^{13}CO_2$, 1 mL of 97% H_2SO_4 was added into a smaller 2 mL vial and placed upside-down into a larger 20 mL vial containing 20.8 mg of $[^{13}C]NaHCO_3$ that had been added into a corner of the vial. After the glass chamber was made air-tight with silicone and a heavy glass lid, the jar was gently shaken to induce the 2 mL vial to topple over and release the H_2SO_4 so that it would come in contact with the $[^{13}C]NaHCO_3$ (Figure 2.2 B). The reaction produced $^{13}CO_2$ as the main carbon source for growth of the seedlings.

2.3.2. Metabolite identification in *C. acuminata* using multiplexed CID

The structure elucidation process begins with the accurate measurement of ion masses, which are used to narrow the possible number of elemental formulas for a specific metabolite. For example, for *m/z* 349.2 ($[M+H]^+$ of camptothecin), the odd nominal mass acquired in positive ion mode is indicative of an even number (including zero) of nitrogen atoms present in the molecule. Indeed, camptothecin has two nitrogen atoms in its structure. To approximate the number of carbons, isotopic patterns are used by rendering the abundance of the $[A+1]^+$ peak (1 Da heavier than the monoisotopic $[A]^+$ ion) as a percentage of the monoisotopic ion abundance and dividing

it by the fractional isotopic abundance of ^{13}C (1.1%). Similarly, the number of hydrogen atoms can be approximated by dividing the mass defect of the molecular ion (the portion of the ion mass that follows the decimal place) by the mass defect of the hydrogen atom (0.0078 Da). The *m/z* values of fragment ions generated during multiplexed CID experiments are used to piece together the functional and characteristic substructures of the metabolite.

To demonstrate how metabolites were characterized, Figure 2.3A displays positive mode ESI mass spectra of camptothecin with increasing number of fragments as collision potentials increase from gentle (15 V) to harsh (80 V) collision potentials. The higher collision energy deposits more internal energy for breaking bonds within the metabolite. The assignment of the protonated molecule (*m/z* 349) was supported by the presence of ions at *m/z* 371.14 and 387.11 annotated as [M+Na]$^+$ and [M+K]$^+$ respectively. As the collision energy was increased to 30V, a fragment ion at *m/z* 305.15 begins to appear, 44 Da lighter than [M+H]$^+$. Raising the collision potential to 45 V increases the abundance of *m/z* 305.15 while decreasing the level of *m/z* 349.1. This is consistent with more of the protonated molecule at *m/z* 349.1 dissociating to produce the fragment ion at *m/z* 305.15 at the elevated collision energy. Increasing the collision potential further yields ions at *m/z* 277.15 and 249.12. Structures are proposed in Figure 2.3B to represent fragment ions observed at *m/z* 305.15, 277.15, and 249.11 and describes their corresponding neutral losses. Similarly, the assignment of structures to the fragment ions of secologanic acid, a proposed early-pathway precursor of camptothecin at collision potential of 15 V is shown in Figure 2.4. The protonated molecule (*m/z* 375) fragments into smaller ions at *m/z* 213.1 (corresponding to loss of $C_6H_{10}O_5$), 195.1, 177.1, 163.1 and 147.1. Sodium and potassium adducts were also observed at *m/z* 397.1 and 413.1, respectively. Using this method, profiling of some proposed

metabolites involved in the biosynthetic pathway of camptothecin in *C. acuminata* was achieved in both labeled and unlabeled leaf tissues.
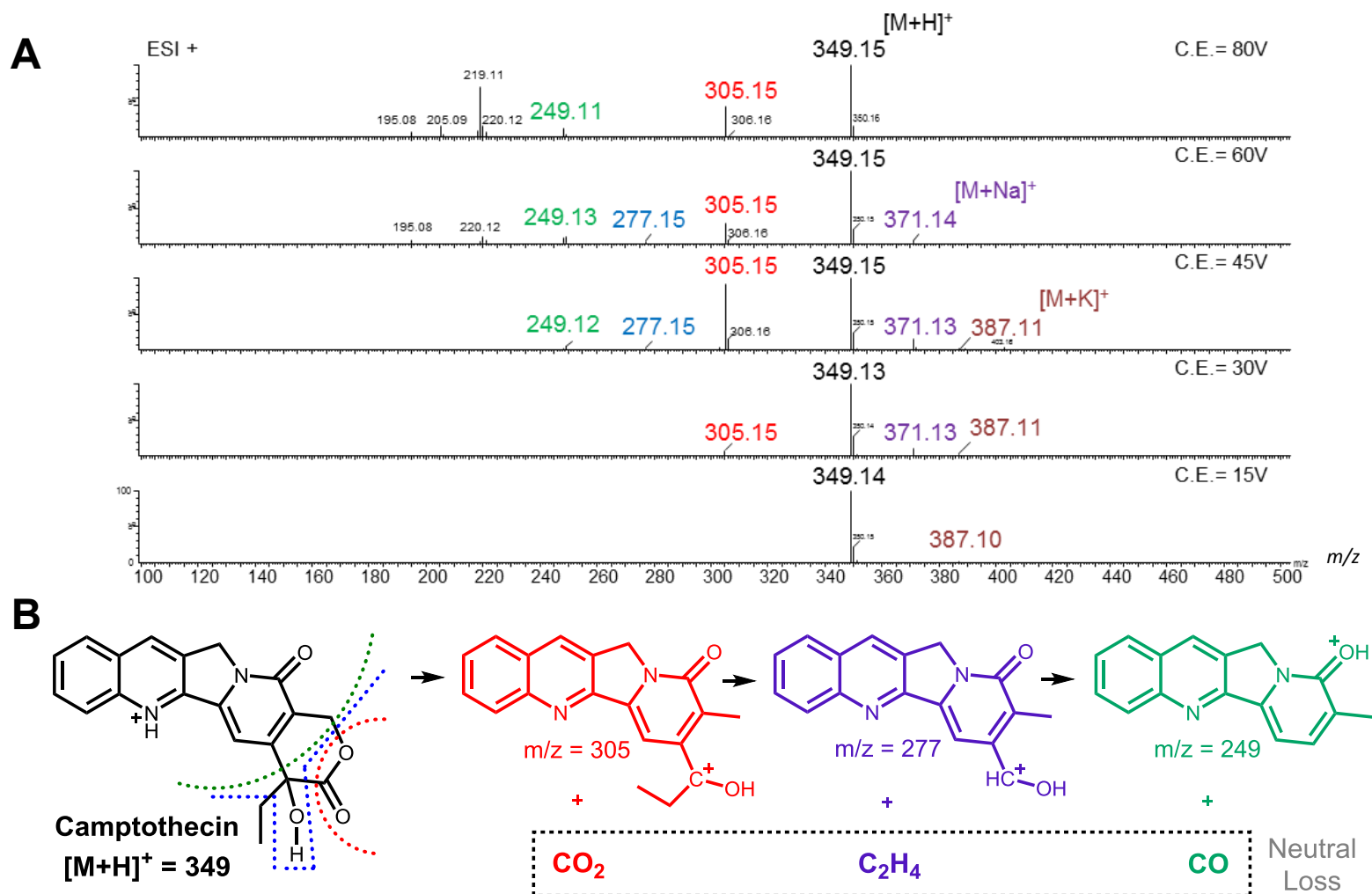
Figure 2.3. Proposed fragmentation pattern for camptothecin revealed by multiplexed CID. (A) Positive-ion mode ESI MS spectra showing annotation of fragments with structures (*m/z* 349.15) acquired at increasing aperture 1 potentials of 15, 30, 45, 60 and 80 V. (B) Proposed structures of fragment ions at *m/z* 305.15, 277.15 and 249.11.
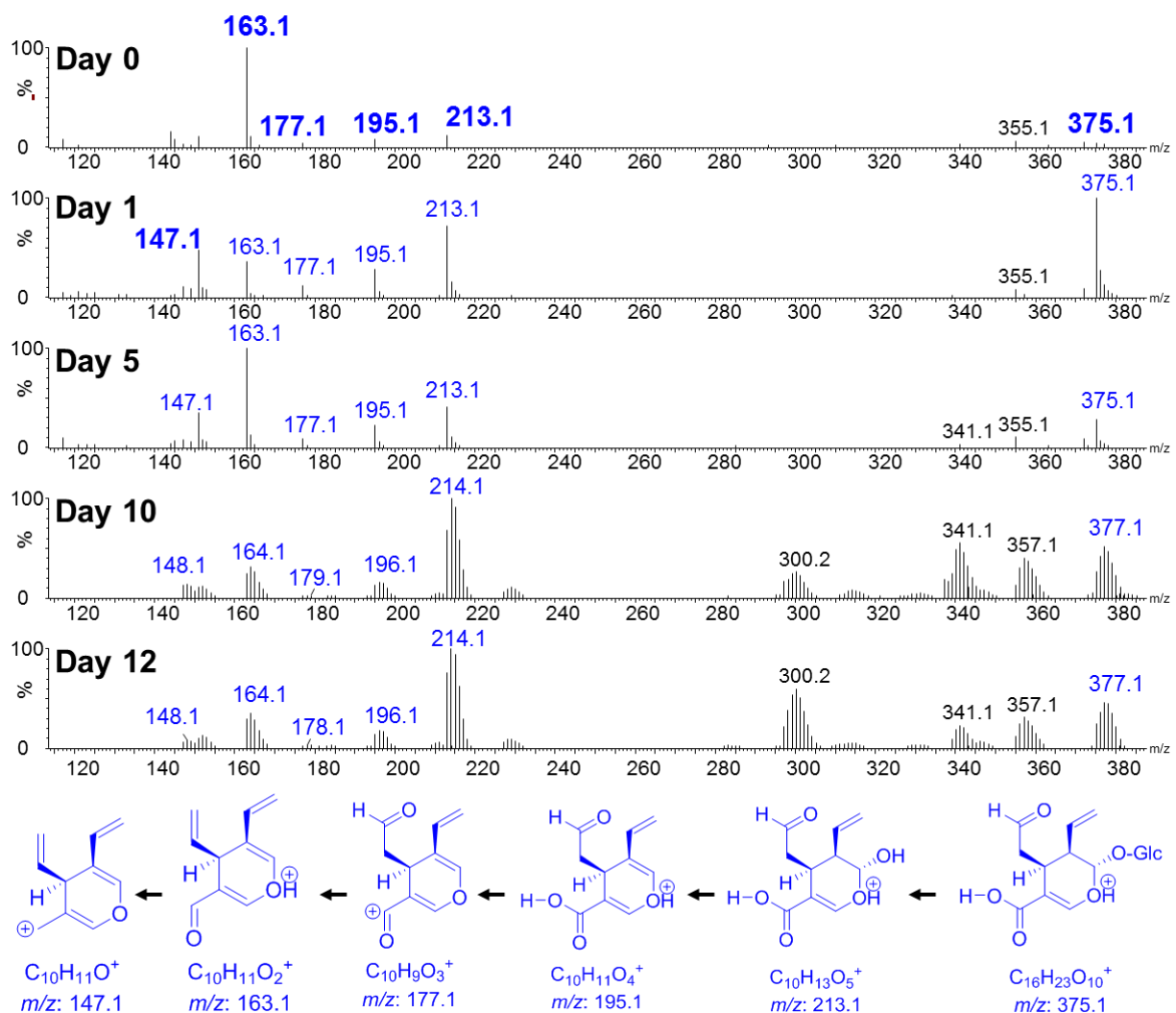
Figure 2.4. Proposed structures for fragments of secologanic acid induced by Multiplexed CID. Positive-ion mode ESI MS spectra showing annotation of pseudomolecular and fragment ions of secologanic acid (*m/z* 375) in Day 0, 1, 5, 10 and 12 of labeling experiment of the 10-week old seedlings. The fragment ions of secologanic acid demonstrate that the isotope label is present in both the glucose and the terpenoid moieties.

2.3.3. Calculation of stable isotope enrichment

Measurement of stable isotope incorporation into specific metabolites and their substructures was achieved using a multiplexed nonselective collision induced dissociation protocol that minimizes mass bias during collisional activation [27]. The absolute abundance of each isotopologue ion within a group was divided by total signal from all isotopologues to obtain a percent signal for each m/z value. The percent signal was then multiplied by the corresponding m/z value, and these individual values summed up to get a signal-weighted average m/z value for each metabolite ion. The difference between the average mass of unenriched metabolite and the signal-weighted mass was divided by the difference between the masses of ^{12}C and ^{13}C to calculate the number of isotopically enriched carbon atoms within the metabolite. Finally, the number of labeled carbons was divided by the number of unlabeled carbons to obtain a mol % enrichment value for each metabolite.

2.3.4. Analysis of labeled intermediates

LC/MS analyses of ^{13}C labeled leaf tissue confirmed successful incorporation of ^{13}C isotope across all structural moieties in *Camptotheca acuminata* metabolites. The Multiplexed-CID method allowed for a nontargeted analysis of the metabolites, allowing for fragmentation of not only the abundant metabolites, but also the metabolites that are low in abundance. The extent of ^{13}C incorporation into the building blocks involved in the biosynthetic pathway of camptothecin, which includes indole, carbohydrate and terpenoid moieties, reflects metabolic fluxes of carbon from $^{13}CO_2$ into the various proposed intermediates that lead to camptothecin (Figure 2.5). For example, for camptothecin, which has 20 carbons, the predicted natural $[A+1]^+$ isotopologue abundance is 21.6% of the monoisotopic ions for m/z 349.15. The $[A+1]^+$ isotopologue abundance

was 23.8% of the monoisotopic ions for *m/z* 349.15 in unlabeled camptothecin. The results indicated that the unlabeled metabolites from the control group produced isotopologue abundances that matched within 5 % of the predicted natural isotopic abundances. All metabolites in the labeled groups were then compared against the results obtained from the control groups.
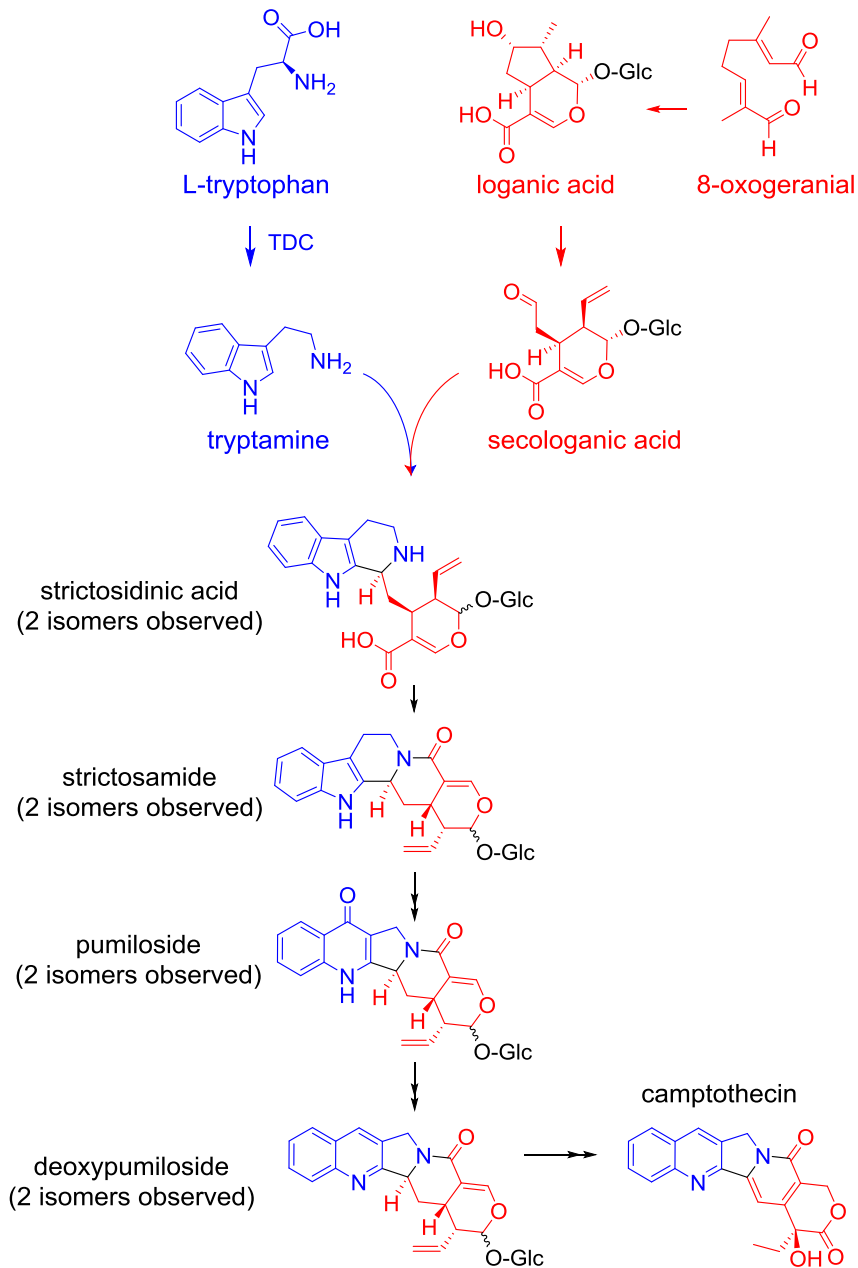
Figure 2.5. Proposed metabolites in the biosynthetic pathway of camptothecin in *Camptotheca acuminata***.** Blue portions of structure represent the moiety that was derived from the amino acid tryptophan while the red portion of the structure represents the moiety derived from iridoid formation.

45

2.3.5 Profiling of key metabolites in the biosynthetic pathway of camptothecin

Using reverse phase chromatographic separation, twelve metabolites from *C. acuminata* leaf tissue grown in unlabeled and labeled atmospheres were annotated based on accurate mass measurements of positive ions and fragment ions observed in LC/MS analyses of extracts (Table 3.1). Based on the comparisons of *m/z* values and retention times, the metabolites proposed to be part of the biosynthetic pathway of camptothecin were present in both labeled and unlabeled plants. These results indicate that minimal alterations were made to the specialized metabolic biochemistry involved in camptothecin biosynthesis by the use of labeled carbon dioxide and the glass chambers.

Metabolite profiling of both 5-week and 10-week old plants of *C. acuminata* revealed that tissues from the younger plants lacked loganic acid but accumulated sucrose, tryptamine, the iridoid secologanic acid, and two isomers each of the MIA strictosidinic acid. Post strictosidinic acid metabolites present in both sets of plants included two isomers of strictosamide and one isomer each of deoxypumiloside, camptothecin and methoxycamptothecin. The retention times of the deoxypumiloside observed in both 5-week old and 10-week old samples matched indicating the absence of the other isomer in younger plants since but it has been previously detected in mature plants as reported in Chapter 3. With regard to proposed metabolites in the biosynthetic pathway, pumiloside is another metabolite that was not observed in the LC/MS analyses in any samples. The absence of pumiloside can be attributed to the focus of the analysis on leaf tissues whereas pumiloside is much more abundant in root tissues as reported in Chapter 3.

Table 2.1. Summary of major metabolites observed in leaf tissues of Camptotheca acuminata. Metabolites in 5-week old seedlings annotated with retention time and m/z values assigned from measurements of pseudomolecular and fragment ions generated by multiplexed non-selective CID.

Annotated metabolite	Molecular Formula	Retention time (min)	Calculated m/z for $[M+H]^+$	Experimental m/z
Sucrose	$C_{12}H_{22}O_{11}$	0.8	343.1235	381.12 ($\{M+K]^+$)
Tryptamine	$C_{10}H_{12}N_2$	7.8	161.1079	144.10 ($[M+H-H_2O]^+$)
Secologanic acid	$C_{16}H_{22}O_{10}$	16.4	375.1291	375.15
Strictosidinic acid isomer 1	$C_{26}H_{32}N_2O_9$	26.1	517.2186	517.27
Strictosidinic acid isomer 2	$C_{26}H_{32}N_2O_9$	27.8	517.2186	517.27
Strictosamide isomer 1	$C_{26}H_{30}N_2O_8$	39.5	499.2080	499.25
Strictosamide isomer 2	$C_{26}H_{30}N_2O_8$	43.5	499.2080	499.25
Deoxypumiloside isomer 1	$C_{26}H_{30}N_2O_8$	38.2	497.1924	497.23
Deoxypumiloside isomer 2	$C_{26}H_{30}N_2O_8$	40.3	497.1924	497.23
Camptothecin	$C_{20}H_{16}N_2O_4$	36.3	349.1188	349.15
Hydroxycamptothecin	$C_{20}H_{16}N_2O_5$	12.7	365.1132	365.07
Methoxycamptothecin	$C_{21}H_{18}N_2O_5$	39.2	379.1288	379.16

2.3.6. Evaluation of isotopic enrichment
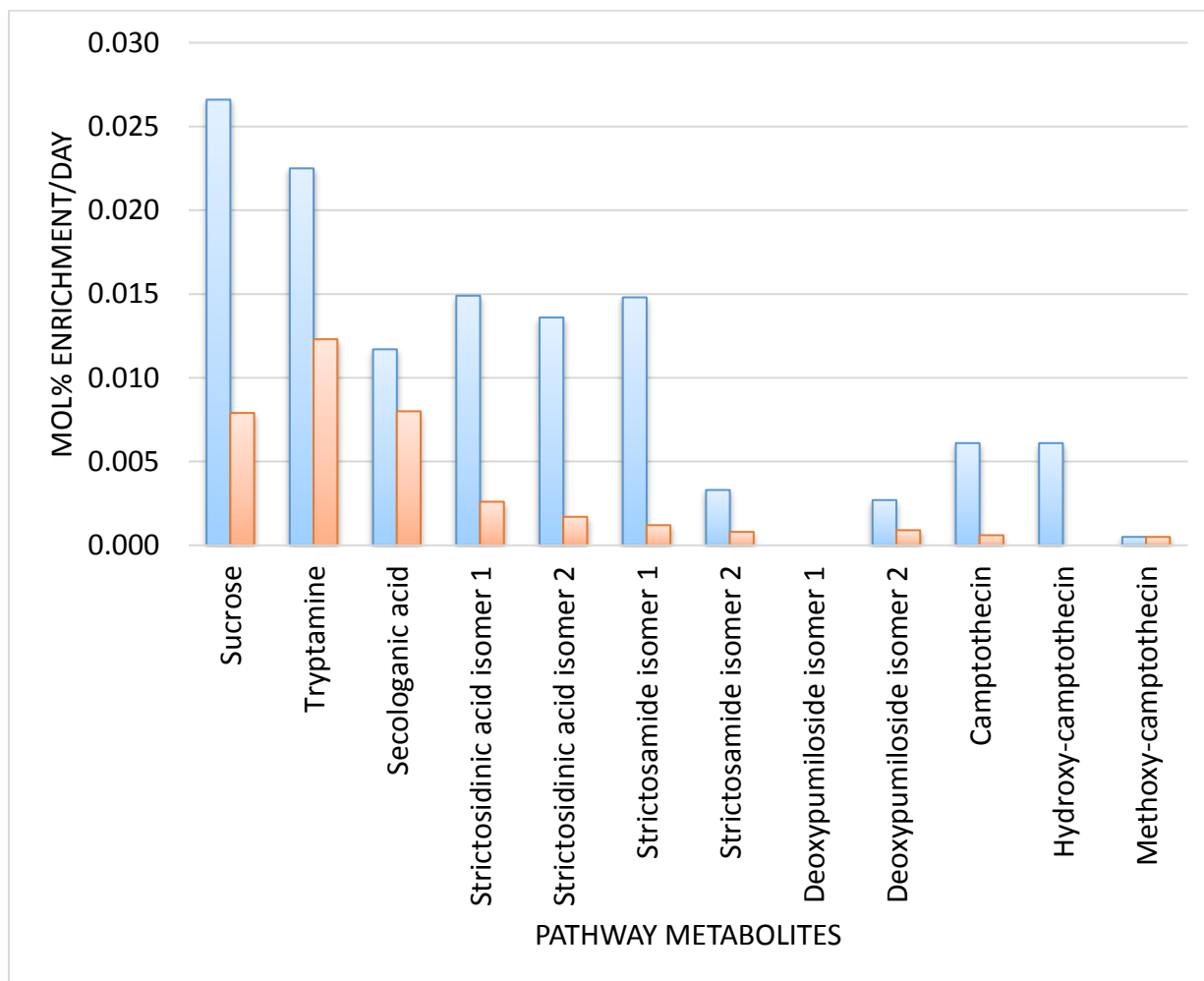
Figure 2.6. **^{13}C enrichments (mol %) of the major proposed camptothecin precursor metabolites observed in leaf tissues of *Camptotheca acuminata*.** Five-week old seedlings (orange bars) are metabolically inert in the production of intermediates in the biosynthetic pathway of camptothecin compared to 10-week old seedlings (blue bars).

To assess differences in metabolic fluxes in leaf tissues between 10-week and 5-week post-germination seedlings, enrichment levels for both sets of samples were calculated based on isotopologue abundances. Incorporation of label in upstream pathway intermediates showed higher levels of labeling than downstream pathway intermediates at all time points. Furthermore, significant incorporation of ^{13}C across all functional groups was observed after 2 weeks of labeling in 10-week old seedlings, while 5-week old seedlings showed lower isotopic enrichments (Figure

2.6). For example, after 16 days of labeling, the 10-week old seedlings demonstrated more than 10 mol % ^{13}C enrichment in camptothecin, while camptothecin from 5-week old seedlings showed no measurable label incorporation. Other major intermediates including strictosidinic acid isomers showed 13-15 mol % labeling in 10-week old seedlings, but only 2.5-4.5 mol % labeling in 5-week old seedlings. Strictosamide isomers showed 13-14.7 mol % labeling in 10-week old seedlings, while only 1.5-2.7 mol % labeling was observed in 5-week old seedlings. In contrast, ^{13}C enrichment in early pathway intermediates including sucrose and tryptamine showed significant labeling in both sets of seedlings (Figure 2.6). For example, sucrose in 10-week old plants exhibited 2.6 mol % enrichment/day and tryptamine 2.3 mol % enrichment/day. This translated into approximately 20 mol% labeling of sucrose and 18 mol% labeling of tryptamine at the end of 16 days of labeling. In 5-week old seedlings, sucrose exhibited 15 mol % enrichment but a shallow slope of 0.8 mol % enrichment/day (Figure 2.7). For example, on Day 5, 10-week old seedlings were already at 14.1 mol % labeling in sucrose while it took up to Day 16 for 15 mol % labeling in 5-week old plants. The average the mol % enrichment/day for all metabolites was lower for 5-week old plants when compared to 10-week old plants. For example, tryptamine exhibited 22 mol % enrichment, also with a shallow slope of 1.2 mol% enrichment/day for 5-week old plants. This indicated that 5-week old seedlings are about 2-fold slower to incorporate ^{13}C than 10-week old plants. Similarly, for sucrose, the 5-week old seedlings are more than 3-fold slower than 10-week old seedlings (Figure 2.7). The larger leaf area of the plants in a 10-week old seedlings may result in higher expression of the biosynthetic genes, which may explain why labeled sucrose is incorporated into the plant at a faster rate than downstream intermediates. Five-week old plants may still procure some of their carbon from the seed and roots, and as a result, do not fix carbon as much as mature plants [28].

A

y = 0.0266x + 0.0124
R² = 0.9609

y = 0.0079x + 0.0203
R² = 0.9418

mol % enrichment of sucrose

Day (s)

B

y = 0.0225x + 0.0262
R² = 0.6567

y = 0.0123x + 0.0334
R² = 0.8978

mol % enrichment of tryptamine
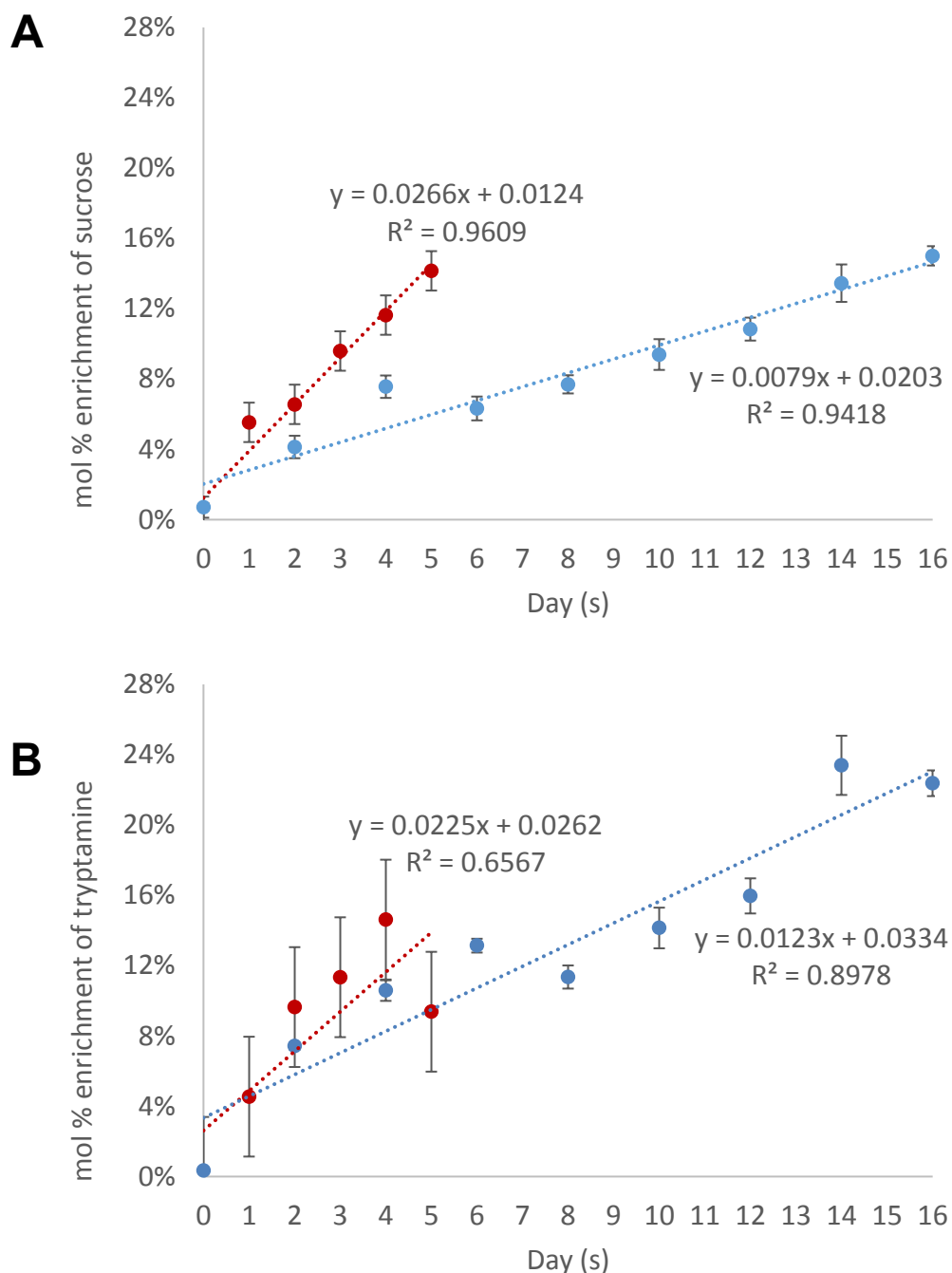
Day (s)

Figure 2.7. Mol % enrichment of (A) sucrose and (B) tryptamine in 10-week (red) and 5-week (blue) old seedlings over a labeling period of 16 days. The slope for each best-fit line represents mol % enrichment per day for the metabolite. The error bars for the 10-week old plants are calculated based on the Excel LINEST function, and the error bars for the 5-week old plants are calculated based on the standard deviation from n=3.
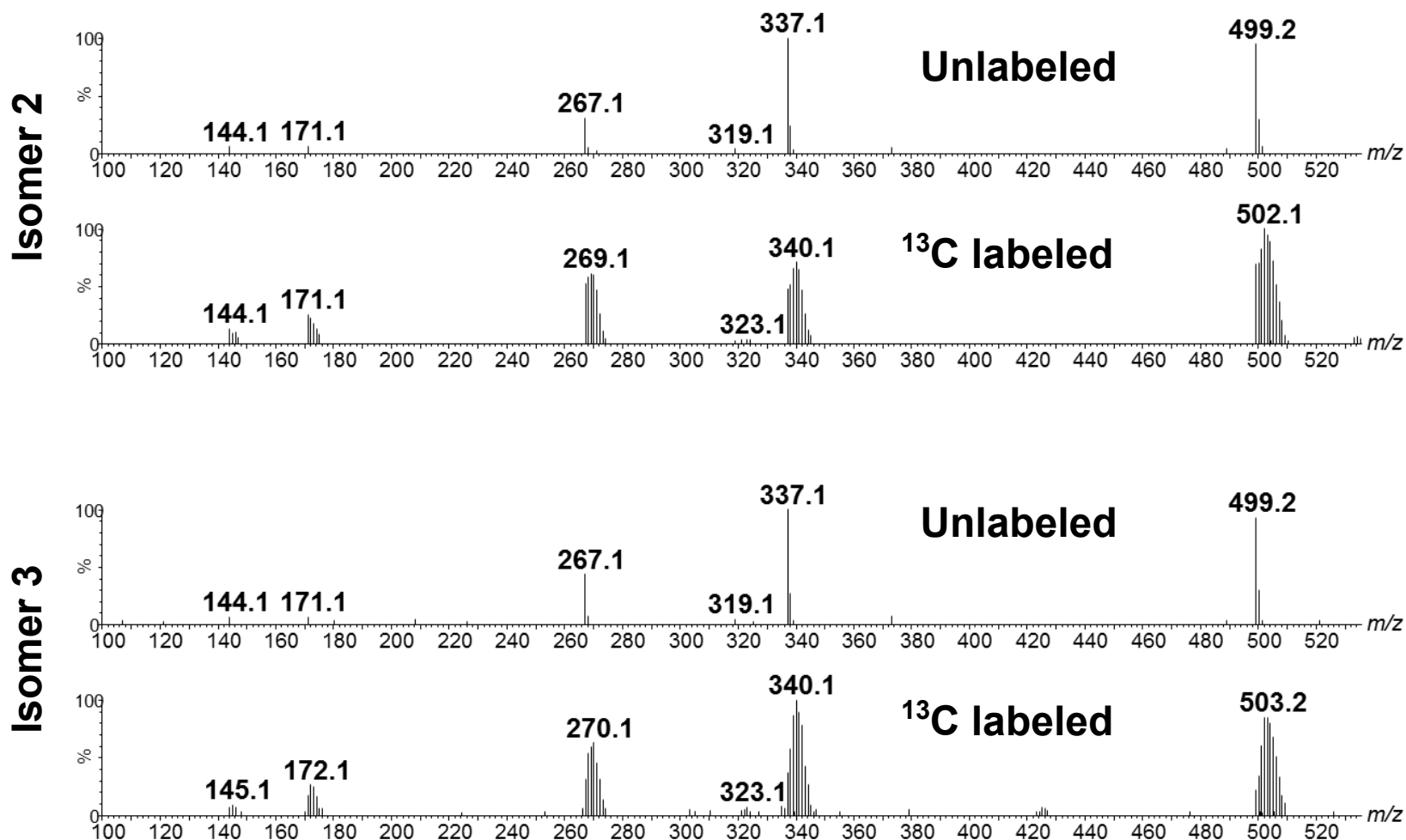
Figure 2.8. Positive-ion mode ESI MS spectra of strictosamide isomers 2 and 3 from labeled and unlabeled (control) *Camptotheca acuminata* **at Day 16 from 10-week old seedlings.** Positive ion mode mass spectra generated for strictosamide (*m/z* 499.2) isomers at collision potentials of 40 V.
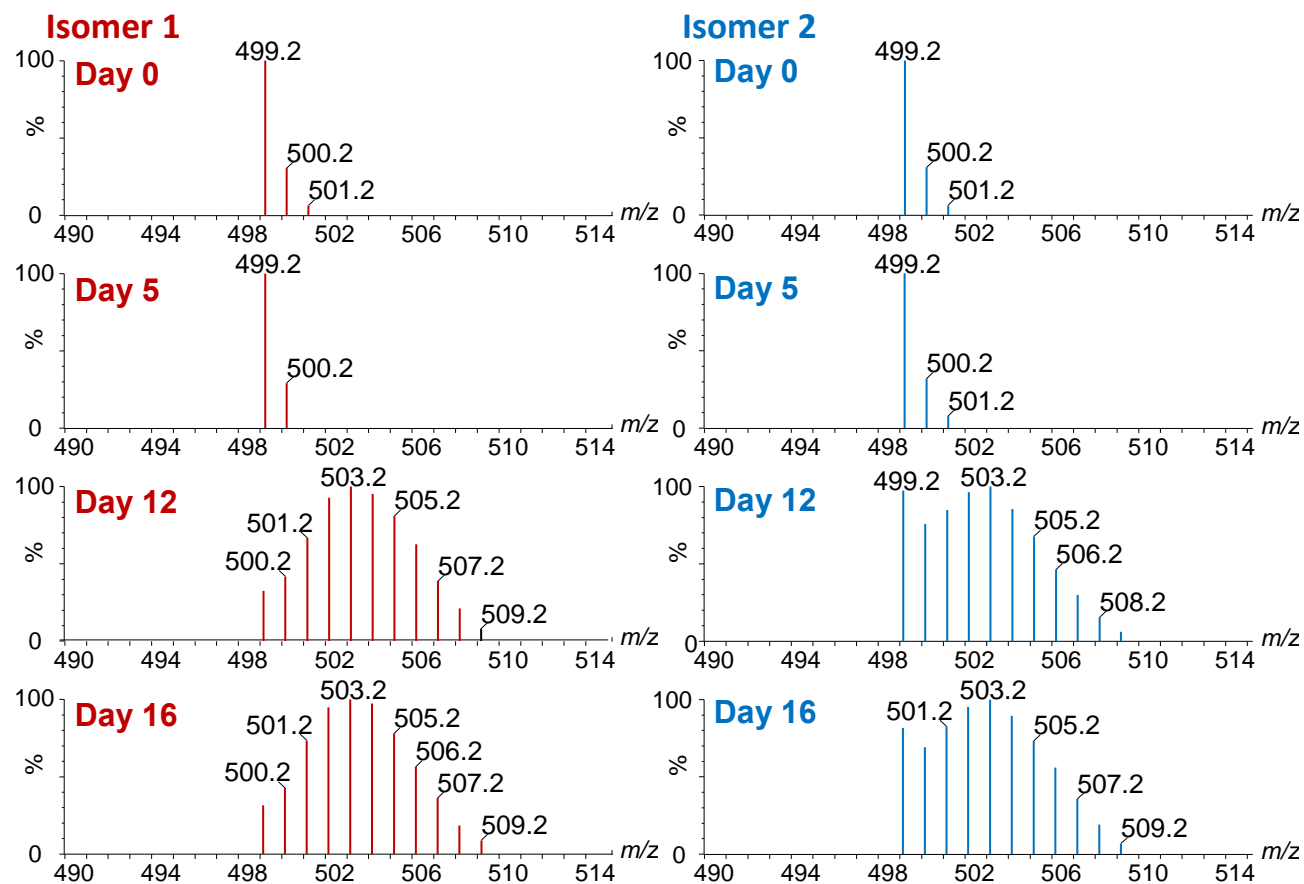
Figure 2.9. Positive-ion mode ESI MS spectra of strictosamide isomer 1 (red) and isomer 2 (blue) from labeled *Camptotheca acuminata* at Days 0, 5, 12 and 16 from 10-week old seedlings. Spectra of molecular ion of strictosamide (*m/z* 499.2) isomer 1 (1st eluting) demonstrate a single mode isotopologue distribution while isomer 2 (2nd eluting) demonstrates a bimodal isotopologue distribution.

Iridoids such as secologanic acid showed about 16 mol % enrichment in both groups of seedlings. When observing the collision-induced dissociation spectrum of secologanic acid (Figure 2.4), it is evident that all fragment ions, including terpenoid and glucose moieties, contain significant amounts of ^{13}C-labeling (Figure 2.4. and 2.8). Similarly, the trend is also observed when analyzing the fragment ions of downstream pathway intermediates. For example, the isotopologue profile generated without mass-selective precursor ion isolation of strictosamide isomers shows that the fragment ions that are produced via Multiplexed CID at 40 V have label incorporated evenly throughout the molecule (Figure 2.8).

2.3.7. Unusual isotopologue distributions in isomers of downstream metabolites

Strictosidinic acid is proposed to be the first camptothecin precursor produced from the condensation of tryptamine with secologanic acid, and it undergoes subsequent dehydration to form strictosamide. Both of these intermediates exist as multiple isomers in *C. acuminata* (Figure 2.5). A closer look at the isotopologues of these downstream metabolites in 10-week old seedlings reveal a bimodal isotopic distribution in one isomer. For example, in the mass spectra of the first-eluting strictosamide isomer, named isomer 1, from Day 0 to 16 represents a distribution in the isotopologues that is different from the second-eluting isomer 2 which shows a bimodal distribution by Day 12 of the labeling experiment (Figure 2.9). For example, the abundance of the monoisotopic [M+H]$^+$ at *m/z* 499.2 is greater than the abundance of the next peak at *m/z* 500.2. A similar observation comes from comparison of the two isomers of strictosidinic acid, where isomer 2 reflects a bimodal isotopologue distribution that is not observed in isomer 1. The observation of bimodal isotopic distribution could be explained by the unlabeled pools of intermediates that were present before the label was introduced into the plants. The 10-week old plants were already making significant amounts of intermediates by the time they were introduced into the experiment. Hence, there is accumulation of non-labeled

metabolites along with newly synthesized labeled metabolites. Interestingly, the trend of bimodal distribution is not observed in strictosidinic acid and strictosamide isomers from 5-week old seedlings. This anomaly supports the idea that the age of the plants that are introduced into the labeling experiment plays a key role in whether a metabolite is synthesized by the plant. For example, the 5-week old seedlings not only produce low amounts of intermediates across the board, but are also slow to synthesize new intermediates (Figure 2.6).

2.4 Conclusions

Significant incorporation of ^{13}C across all functional groups of the biosynthetic pathway (terpenoid, carbohydrate, and indole) intermediates was observed in less than two weeks of labeling. The labeled compounds generated using this approach can be used to isolate, purify and characterize the low abundant metabolites using semi-preparative HPLC and 2-dimensional NMR analysis. NMR spectroscopy would be a valuable tool for analyzing these labeled metabolites without having to isolate milligram quantities. This is because NMR signals that are based on ^{13}C will be enhanced when analyzing metabolites which have levels of ^{13}C that are more than the natural abundance of carbon at 1.1%. The isolated compounds can also be used as tracers in feeding studies to clarify other intermediates involved in the pathway.

The simple and inexpensive design and functionality of the method built in this study for $^{13}CO_2$ labeling can be introduced easily into any laboratory. The maximum labeling of plants under a ^{13}C atmosphere generally do not reach beyond 20 % because the bacteria in the soil will use up the remaining $^{13}CO_2$. These findings show that short labeling times of two weeks are sufficient to obtain maximum labeling even in complex specialized metabolite structures depending on the developmental age of the plant. Whole-plant labeling allows for non-targeted approaches in tracing the transfer of carbon atoms within the succession of

intermediates in the biosynthetic pathway, as well as flux measurements that allow system-wide metabolic regulation of specialized metabolites.

The results lead to the conclusion that there is a metabolic bottleneck in specialized metabolite production in younger plants, in this case at 5 weeks post-germination. Examination of label incorporation from specific metabolite substructures via fragments observed on the mass spectra shows that the ^{13}C is incorporated into the early building blocks including sucrose, tryptamine and secologanic acid, but is not efficiently incorporated into downstream pathway metabolites. Results indicate that the rates of ^{13}C incorporation of iridoids are not affected by the age of the plant. However, the developmental stage of the plant has a great effect on rates of biosynthesis of downstream pathway intermediates. It is proposed that younger plants may produce volatile iridoids as chemical defenses against predators before plant tissues have developed alternative physical (*e.g.* lignin formation) or chemical (e.g. camptothecin) defenses. Absence of formation of equivalent levels of intermediates in the 5-week old plants indicate that the developmental stage of the plant is important when researching into specific metabolites. For example, when studying the up or down regulation of downstream metabolites, one would not choose young seedlings that are younger than 10 weeks old.

Extraction and isolation of specialized metabolites is often necessary to further study the interaction of the intermediates with candidate enzymes, and for structure characterization purposes. When ^{13}C labeled precursors are incorporated into downstream metabolites, the NMR signals based on ^{13}C will be markedly increased, in turn, accommodating analysis of low abundant intermediates in the µg scale [30]. This study concludes that labeling of plants that are younger than 10 weeks would have limited utility if the researcher is aiming for labeled downstream metabolites since the plant is not turning labeled precursors into the more complex downstream intermediates.

Finally, the results from this study will aid in the discovery of new metabolites and close the gaps in the current knowledge of pathway intermediates and their transformations through the cycle of stable isotopic labeling, isolation and purification, and structure elucidation of the metabolite.

REFERENCES

REFERENCES

1.   Lorence, A. and C.L. Nessler, *Camptothecin, over four decades of surprising findings.* Phytochemistry, 2004. **65**(20): p. 2735-2749.

2.   Lorence, A., F. Medina-Bolivar, and C.L. Nessler, *Camptothecin and 10-hydroxycamptothecin from Camptotheca acuminata hairy roots.* Plant Cell Reports, 2004. **22**(6): p. 437-441.

3.   Wall, M.E., et al., *Plant Antitumor Agents. I. The Isolation and Structure of Camptothecin, a Novel Alkaloidal Leukemia and Tumor Inhibitor from Camptotheca acuminata1,2.* Journal of the American Chemical Society, 1966. **88**(16): p. 3888-3890.

4.   Verpoorte, R., et al., *Metabolic engineering of plant secondary metabolite pathways for the production of fine chemicals.* Biotechnology Letters, 1999. **21**(6): p. 467-479.

5.   van Hengel, A.J., R.M. Buitelaar, and H.J. Wichers, *Camptotheca acuminata Decne: In Vitro Culture and the Production of Camptothecin*, in *Medicinal and Aromatic Plants VII*, Y.P.S. Bajaj, Editor. 1994, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 98-112.

6.   Tabata, M., et al., *Pigment formation in callus cultures of Lithospermum erythrorhizon.* Phytochemistry, 1974. **13**(6): p. 927-932.

7.   Yukimune, Y., et al., *Methyl jasmonate-induced overproduction of paclitaxel and baccatin III in Taxus cell suspension cultures.* Nat Biotechnol, 1996. **14**(9): p. 1129-32.

8.   Sakato, K., et al., *Isolation and Identification of Camptothecin from Cells of Camptotheca acuminata Suspension Cultures.* Agricultural and Biological Chemistry, 1974. **38**(1): p. 217-218.

9.   Lorence, A., F. Medina-Bolivar, and C.L. Nessler, *Camptothecin and 10-hydroxycamptothecin from Camptotheca acuminata hairy roots.* Plant Cell Rep, 2004. **22**(6): p. 437-41.

10.   Lopez-Meyer, M., C.L. Nessler, and T.D. McKnight, *Sites of accumulation of the antitumor alkaloid camptothecin in Camptotheca acuminata.* Planta Med, 1994. **60**(6): p. 558-60.

11.   Pardo, E., et al., *De novo production of six key grape aroma monoterpenes by a geraniol synthase-engineered S. cerevisiae wine strain.* Microbial Cell Factories, 2015. **14**: p. 136.

12.   Ajikumar, P.K., et al., *Isoprenoid Pathway Optimization for Taxol Precursor Overproduction in Escherichia coli.* Science, 2010. **330**(6000): p. 70-74.

13. Ding, M.-z., et al., *Biosynthesis of Taxadiene in <italic>Saccharomyces cerevisiae </italic>: Selection of Geranylgeranyl Diphosphate Synthase Directed by a Computer-Aided Docking Strategy.* PLoS ONE, 2014. **9**(10): p. e109348.

14. Qu, X., et al., *Molecular Cloning, Heterologous Expression, and Functional Characterization of an NADPH-Cytochrome P450 Reductase Gene from Camptotheca acuminata, a Camptothecin-Producing Plant.* PLoS One, 2015. **10**(8): p. e0135397.

15. Arabidopsis Genome, I., *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.* Nature, 2000. **408**(6814): p. 796-815.

16. Goffeau, A., et al., *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546, 563-7.

17. Leonard, E., et al., *Opportunities in metabolic engineering to facilitate scalable alkaloid production.* Nat Chem Biol, 2009. **5**(5): p. 292-300.

18. Siddiqui, M.S., et al., *Advancing secondary metabolite biosynthesis in yeast with synthetic biology tools.* FEMS Yeast Research, 2012. **12**(2): p. 144-170.

19. Canel, C., et al., *Effects of over-expression of strictosidine synthase and tryptophan decarboxylase on alkaloid production by cell cultures of Catharanthus roseus.* Planta, 1998. **205**(3): p. 414-9.

20. Dunn, W.B. and D.I. Ellis, *Metabolomics: Current analytical platforms and methodologies.* TrAC Trends in Analytical Chemistry, 2005. **24**(4): p. 285-294.

21. Glinski, M. and W. Weckwerth, *The role of mass spectrometry in plant systems biology.* Mass Spectrometry Reviews, 2006. **25**(2): p. 173-214.

22. Fan, T.W., et al., *Stable isotope-resolved metabolomics and applications for drug development.* Pharmacol Ther, 2012. **133**(3): p. 366-91.

23. Huege, J., et al., *GC-EI-TOF-MS analysis of in vivo carbon-partitioning into soluble metabolite pools of higher plants by monitoring isotope dilution after $^{13}CO_2$ labelling.* Phytochemistry, 2007. **68**(16-18): p. 2258-72.

24. Krishnan, P., N.J. Kruger, and R.G. Ratcliffe, *Metabolite fingerprinting and profiling in plants using NMR.* J Exp Bot, 2005. **56**(410): p. 255-65.

25. Breitling, R., A.R. Pitt, and M.P. Barrett, *Precision mapping of the metabolome.* Trends in Biotechnology, 2006. **24**(12): p. 543-548.

26. Gu, L., A.D. Jones, and R.L. Last, *Broad connections in the Arabidopsis seed metabolic network revealed by metabolite profiling of an amino acid catabolism mutant.* Plant J, 2010. **61**(4): p. 579-90.

27. Wang, Z., A.D. Jones, P*rofiling of Stable Isotope Enrichment in Specialized Metabolites Using Liquid Chromatography and Multiplexed Nonselective Collision-Induced Dissociation.* Analytical Chemistry, 2014. **86**(21): p. 10600-10607.

28. Raines, C.A., M.J. Paul, *Products of leaf primary carbon metabolism modulate the developmental programme determining plant morphology.* Journal of Experimental Botany, 2006. **57**(9): p. 1857-1862.

29. Stöckigt, J. and M.H. Zenk, *Isovincoside (strictosidine), the key intermediate in the enzymatic formation of indole alkaloids.* FEBS Letters, 1977. **79**(2): p. 233-237.

30. Exarchou V., et al., *LC-NMR coupling technology: recent advancements and applications in natural products analysis.* Magn Reson Chem: MRC. 2005;**43**(9):681–687.

Chapter 3: Strategies for profiling of specialized metabolites in non-transformed and transgenic lines of *Camptotheca acuminata*

This chapter presents part of the research described in the following journal article: Radin Sadre, Maria Magallanes-Lundback, Sujana Pradhan, Vonny Salim, A. Daniel Jones, Dean DellaPenna, Metabolite Diversity in alkaloid biosynthesis: A multi-lane highway for camptothecin production in *Camptotheca acuminata*. *The Plant Cell*. Advance Publication 18th July 2016, tpc. 00193.2016

3.1. Introduction

Monoterpene indole alkaloids (MIA) account for thousands of specialized metabolites produced by plant species from the orders Cornales and Gentianales. Much of this chemical diversity originates from the common precursor 3-alpha(*S*)-strictosidine, which is formed by stereospecific condensation of the indole metabolite tryptamine and the monoterpene secologanin [1]. MIA metabolic diversity likely evolved as components of plant adaptation to changing environments and as defensive agents against various biotic stresses. Many MIAs, including camptothecin, target specific cellular processes in mammals and have important pharmacological activities and medicinal uses [2].

Camptotheca acuminata (Cornales, Nyssaceae) produces the monoterpene indole alkaloid camptothecin, a potent anti-cancer drug approved by FDA for colon, ovarian, cervical and small-cell lung cancers. Camptothecin was first identified during the 1960s as a novel antitumor alkaloid extracted from a tree native to Southern China [3]. Its mode of action is the specific inactivation of topoisomerase I resulting in cell death by apoptosis [4]. Semisynthetic analogs derived from camptothecin exhibit improved pharmacological properties and clinical efficacy relative to camptothecin, and are widely used to treat lung, colorectal, cervical and ovarian cancers [5]. As for many alkaloids, multiple-step and complex synthesis of camptothecin is impractical and the production of semi-synthetic derivatives with improved pharmacological properties relies entirely on camptothecin isolated from the bark and seeds of *C. acuminata* and of *Nothapodytes nimmoniana*. Unfortunately, even in these plants, camptothecin is produced only in tiny amounts [6]. Due to the need for isolation solely from plants, the cost of camptothecin treatments and development of improved analogs has increased with the growing worldwide demand for

camptothecin. Hence, there is a need for the exploration of more economical and sustainable alternatives for the production of camptothecin.

Despite the medicinal importance of camptothecin, its biosynthesis by plants remains mysterious. Since its original discovery in *C. acuminata*, camptothecin was found to accumulate in several Asian and African tropical and subtropical plant species from unrelated orders and families, including species of *Ophiorrhiza* [7, 8]. Early radiolabeling experiments in *C. acuminata* demonstrated that tryptophan and its decarboxylation product tryptamine were metabolic precursors, as was a mixture of the monoterpene alcohols geraniol and its isomer nerol, leading to classification of camptothecin as a MIA [9] despite it having a quinoline, and not an indole, ring system. Subsequent stable isotope labeling and NMR experiments demonstrated that strictosamide served as a precursor for camptothecin in *C. acuminata* [10]. The co-occurrence of camptothecin in *Ophiorrhiza pumila* with the structurally related alkaloids pumiloside and deoxypumiloside suggested they might be intermediates in the camptothecin biosynthetic pathway [11]. Notably, pumiloside was also detected in C. *acuminata* [12, 13]. Recent studies of camptothecin biosynthesis in *O. pumila* hairy roots demonstrated that strictosidine, derived from the condensation of tryptamine and the seco-iridoid secologanin, is required for camptothecin synthesis in this species [14, 15].

The seco-iridoid pathway, which is a form of metabolism of the monoterpene alcohol geraniol, that leads to strictosidine was only recently elucidated in the MIA-producing plant *Catharanthus roseus*. Geraniol 8-oxidase (CYP76B6) catalyzes hydroxylation of the 2-C-methylerythritol-4-phosphate (MEP) pathway-derived isoprenoid geraniol to 8-hydroxygeraniol, which undergoes further oxidation to the di-aldehyde 8-oxogeranial by 8-hydroxygeraniol oxidoreductase [16]. The enzyme iridoid synthase catalyzes the subsequent conversion of 8-oxogeranial to iridodial [17],

which is oxidized further to 7-deoxyloganetic acid by 7-deoxyloganetic acid synthase, which is a cytochrome P450 enzyme designated as CYP76A26 [18]. 7-Deoxyloganetic acid is glycosylated to yield 7-deoxyloganic acid [19], which undergoes subsequent hydroxylation by another cytochrome P450, CYP72A224, to yield loganic acid, whose carboxyl group is then O-methylated to form loganin [20, 21]. An additional cytochrome P450, ecologanin synthase (CYP72A1) then catalyzes the oxidative ring-opening of loganin to secologanin, which contains an aldehyde group suitable for condensation with the amino group of tryptamine [22]. Strictosidine synthase then catalyzes the stereospecific condensation of tryptamine and secologanin to 3-alpha(*S*)-strictosidine [23]. It was proposed that strictosidine is subsequently channeled into the biosynthesis of the downstream MIAs strictosamide, pumiloside and deoxypumiloside to yield camptothecin in *O. pumila* [14]. However, post-strictosidine reaction steps, enzymes and their sequence(s) remain unknown in camptothecin-producing plants.

The Medicinal Plant Consortium recently made the assembled transcriptome, gene expression and metabolite profiles for *C. acuminata* publicly available (http://medicinalplantgenomics.msu.edu/, http://metnetdb.org/mpmr_public/). Here, to identify the genes involved in production of key intermediates that are part of the biosynthetic pathway of camptothecin, metabolic phenotypes in non-transformed and silenced *Camptotheca acuminata* are assessed by profiling and quantifying metabolites in root, stem, shoot apex, young leaf and old leaf tissues.

3.2. Materials and methods

3.2.1. Reagents

HPLC grade acetonitrile and formic acid (88% aqueous solution) were purchased from VWR Scientific. Tryptamine and loganic acid were purchased from ChromaDex, camptothecin from MP Biomedicals, LLC and telmisartan from Toronto Research Chemicals.

3.2.2. Plants

Root, stem, shoot apex, young leaf and old leaf tissues of non-transformed and silenced *C. acuminata* were obtained from the laboratory of Professor Dean DellaPenna at Michigan State University.

3.2.3. Extraction of metabolites

For each tissue sample, approximately 30 mg of frozen powdered plant tissue and 500 of μl acetonitrile/water (7/3, v/v) containing 1.25 μM telmisartan (internal standard) was added, and mixture was vortexed for 5 sec and incubated in the dark at 4 °C for 16 hours. Samples were then centrifuged at 4 °C and 10,000 × g for 30 min, supernatants were transferred to fresh tubes and centrifuged at 4 °C and 10,000 × g for 15 min and 40-μl aliquots diluted by addition of 150 μL deionized water. Immediately prior to HPLC/MS analysis, 10 μL of 10% formic acid was added by the autosampler to each extract and mixed by drawing the liquid into the autosampler syringe and ejecting back into the sample vial. This procedure minimized acid-catalyzed degradation while vials remained in the autosampler tray. Individual standard solutions of tryptamine, loganic acid, camptothecin and telmisartan were prepared over a range of concentrations from 0 - 70 μM, and were analyzed together with each set of plant tissue extracts.

3.2.4. LC/MS experiments

LC/MS analyses were performed using a Shimadzu LC-20AD ternary pump coupled to a SIL-5000 autosampler, column oven, and Waters LCT Premier mass spectrometer equipped with an electrospray ionization source. A 10-µL volume of each extract was analyzed using either a 52-min or a 15-min gradient elution method on an Ascentis Express C18 column (2.1×100 mm, 2.7 µm) with mobile phases consisting of 10 mM ammonium formate in water, pH 2.85 (solvent A) and methanol (solvent B). The 52-min method gradient was as follows: 2% B at 0.00-2.00 min, linear gradient to 20% B at 20.00 min, linear gradient to 55% B at 43.00 min, then a step to 99% B at 43.01 min, then return to 2% B over 47.01-52.00 min. The 15-min method employed 8% B at 0.00-1.00 min, linear gradient to 40% B at 3.00 min, linear gradient to 70% B at 11.00 min, then step to 99% B at 11.01 and held until 13.00 min, followed by a return to 8% B and held from 14.00-15.00 min. For both gradients, the flow rate was 0.3 ml/min and the column temperature was 45 °C. The mass spectrometer was operated using V optics in positive-ion mode with a typical resolution of ~ 4000 at full width at half maximum. Source parameters were as follows: capillary voltage 3200 V, sample cone voltage 10 V, desolvation temperature 350 °C, source temperature 100 °C, cone gas flow 40 L/h and desolvation gas flow 350 L/h. Mass spectrum acquisition was performed over m/z 50 to 1,500 with scan time of 0.1 s, using dynamic range extension. Mass spectra containing fragment ions were generated by rapid switching of aperture 1 voltage over four parallel data acquisition functions (20, 40, 60 and 80 V) (Gu et al., 2010). Accurate masses and fragments were confirmed in UHPLC/MS/MS analyses (Xevo G2-S QTOF mass spectrometers, Waters). The ESI positive mode conditions for Xevo G2-S QTOF was follows: capillary voltage, 3 kV; sample cone voltage, 35 V; source temperature, 100 °C; desolvation gas temperature, 350 ºC, cone gas flow, 50 L/h, desolvation gas flow, 600 L/h. Mass spectra were acquired over m/z 50-

600 using centroid peak acquisition and extended dynamic range in resolution mode (M/ΔM ~ 22 000, fwhm). Not data-dependent, scanned at different retention times, collision energy start at 15eV and reaches 80 eV at the end of each of the 7 functions.

For quantitative analyses, the 15-min gradient HPLC/MS method was used. The extracted ion chromatograms for each target analyte were integrated, and analytes were quantified using QuanLynx software (Waters) with a mass window allowance of 0.2. Sodium adduct peaks [M+Na]$^+$ were quantified for iridoid carboxylic acids and [M+H]$^+$ signals were measured for downstream nitrogen-containing metabolites, respectively. All calculated peak areas were normalized to the internal standard telmisartan and tissue fresh weight.

3.3. Results and discussion

The findings in this study are in contrast to other well studied MIA producing plants, including *O. pumila* (Yamazaki et al., 2003b; Asano et al., 2013). Perhaps most notable is that *C. acuminata* does not synthesize strictosidine and instead uses an alternative seco-iridoid pathway to produce strictosidinic acid. Unlike other camptothecin-producing plants, a mixture of isomers of strictosidinic acid was formed as judged from accurate *m/z* values of [M+H]$^+$ ions (Table 3.1). None of these alternative isomers appear to accumulate, and it is postulated that all strictosidinic acid isomers are channeled into the camptothecin pathway, with multiple isomers of intermediates downstream of strictosidinic acid are retained throughout the pathway. This can be considered analogous to a multi-lane highway in which parallel lanes lead to a single destination, in this case the single isomer of camptothecin. Based on current knowledge, it is believed to be an unusual, if not unique, example of either enzyme promiscuity or the action of multiple enzymes, each acting on a single isomeric intermediate.

The analysis of tissue extracts relies on information-rich mass spectrometry strategies for metabolite annotation and measurement. Figure 3.1 shows the total ion LC/TOF MS chromatogram of wild type *C. acuminata* root tissue showing metabolites annotated with retention time and *m/z* values using accurate mass measurements of $[M+H]^+$ and fragment ions generated by multiplexed non-selective CID.
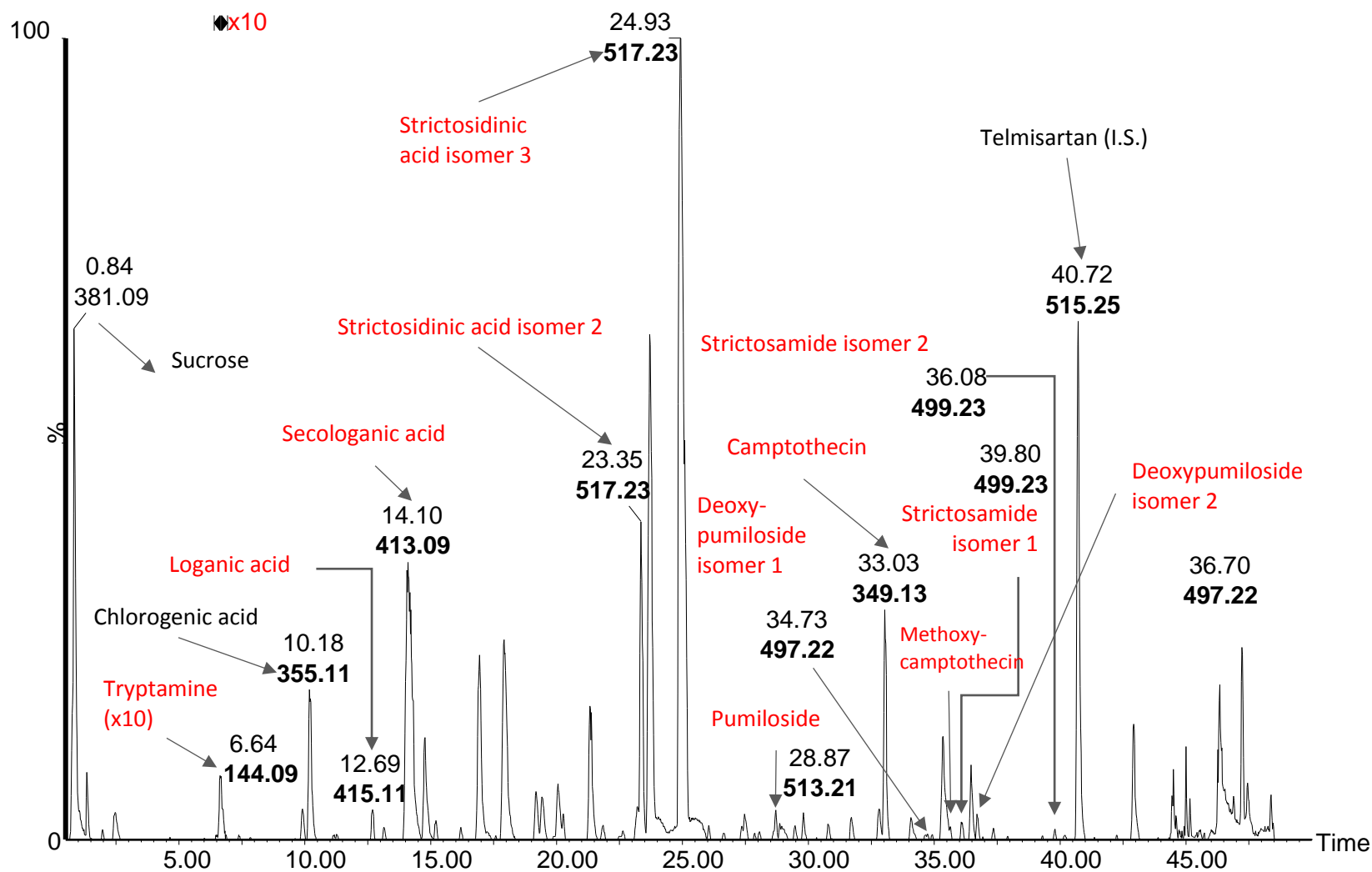
Figure 3.1. Base peak ion LC/TOF MS chromatogram of *Camptotheca acuminata* root tissue in positive ion mode showing metabolites annotated with retention time in minutes (top number) and base peak *m/z* values (bottom number in bold) using accurate measurements of pseudomolecular and fragment ions generated by multiplexed non-selective CID.

Table 3.1. Relevant compounds detected in wild-type *C. acuminata*. Metabolites in root, stem, shoot apex and leaf extracts were separated by a 52-minute HPLC/MS method and are listed with their precursor and fragment ions. Fragment ions obtained after loss of a neutral glucose unit (162 Da) are highlighted in bold. Only stem tissue contained detectable levels of all metabolite isomers listed here.

Annotated metabolite	Formula	Retention time (min)	Calculated $[M+H]^+$ (*m/z*)	Experimental $[M+H]^+$ (*m/z*)	Fragment ion(s) from MS/MS product ion spectra (*m/z*); products of $[M+H]^+$
tryptamine	$C_{10}H_{12}N_2$	8.0	161.1079	161.1065	144
loganic acid	$C_{16}H_{24}O_{10}$	13.6	377.1448	377.1450	359, **215**, 197, 179, 161, 151, 137, 133, 123, 109, 81
secologanic acid	$C_{16}H_{22}O_{10}$	15.1	375.1291	375.1290	**213**, 195, 177, 151, 125, 109, 107, 95, 79, 77
strictosidinic acid isomer 1	$C_{26}H_{32}N_2O_9$	23.9	517.2186	517.2198	**355**, 338, 320, 269/268, 251, 194, 180/181, 168/170, 151, 156, 144, 130, 125
strictosidinic acid isomer 2	$C_{26}H_{32}N_2O_9$	24.8	517.2186	517.2210	**355**, 338, 320, 269/268, 251, 194, 180/181, 168/170, 151, 156, 144, 130, 125
strictosidinic acid isomer 3	$C_{26}H_{32}N_2O_9$	26.5	517.2186	517.2200	**355**, 338, 320, 269/268, 251, 194, 180/181, 168/170, 151, 156, 144, 130, 125
strictosamide isomer 1	$C_{26}H_{30}N_2O_8$	37.6	499.2080	499.2081	**337**, 319, 267, 171, 144
strictosamide isomer 2	$C_{26}H_{30}N_2O_8$	41.5	499.2080	499.2086	**337**, 319, 267, 171, 144
strictosamide epoxide isomer 1	$C_{26}H_{30}N_2O_9$	22.8	515.2030	515.2021	**353**, 335, 309, 291, 283, 265, 263, 237, 209, 183, 184, 155, 144
strictosamide epoxide isomer 2	$C_{26}H_{30}N_2O_9$	24.8	515.2030	515.2031	**353**, 335, 309, 291, 283, 265, 263, 237, 209, 183, 184, 155, 144
strictosamide epoxide isomer 3	$C_{26}H_{30}N_2O_9$	26.5	515.2030	515.2026	**353**, 335, 309, 291, 283, 265, 263, 237, 209, 183, 184, 155, 144
strictosamide diol	$C_{26}H_{32}N_2O_{10}$	20.1	533.2135	533.2170	**371**, 353, 283, 265, 185, 160, 142, 132
strictosamide ketolactam isomer 1	$C_{26}H_{30}N_2O_{10}$	22.5	531.1979	531.1953	**369**, 351, 341, 299, 281, 271, 253, 194, 176, 158, 148, 130, 124, 106
strictosamide ketolactam isomer 2	$C_{26}H_{30}N_2O_{10}$	22.7	531.1979	531.1979	**369**, 351, 341, 299, 281, 271, 253, 194, 176, 158, 148, 130, 124, 106
pumiloside isomer 1	$C_{26}H_{28}N_2O_9$	30.2	513.1873	513.1890	**351**, 333, 315, 305, 281, 235, 140
pumiloside isomer 2	$C_{26}H_{28}N_2O_9$	32.9	513.1873	513.1899	**351**, 333, 315, 305, 281, 235, 140
deoxypumiloside isomer 1	$C_{26}H_{28}N_2O_8$	36.1	497.1924	497.1930	**335**, 265, 247, 219, 183, 169, 142, 97
deoxypumiloside isomer 2	$C_{26}H_{28}N_2O_8$	38.2	497.1924	497.1931	**335**, 265, 247, 219, 183, 169, 142, 97
camptothecin	$C_{20}H_{16}N_2O_4$	34.3	349.1188	349.1198	305, 277, 249, 219/220, 168
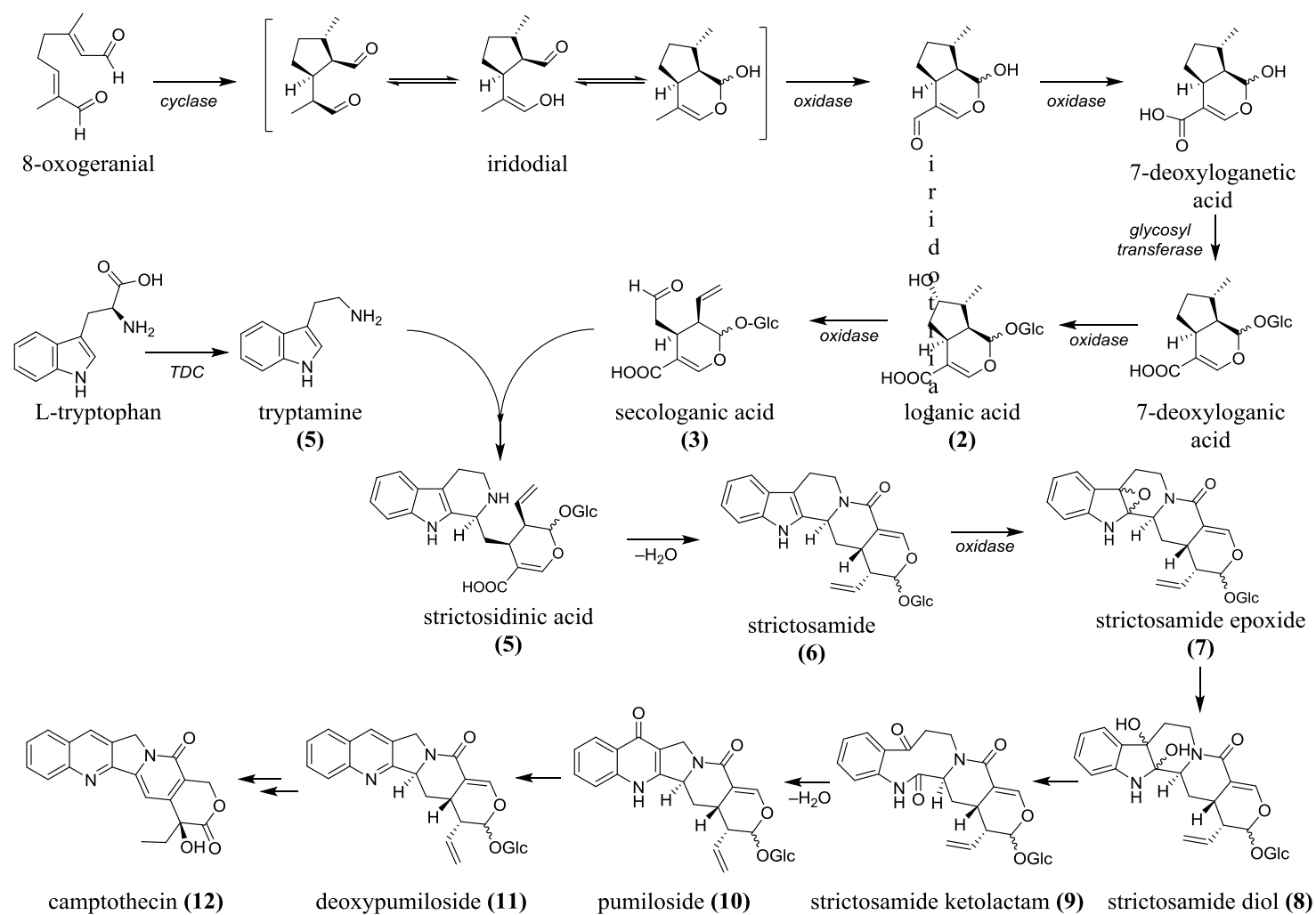
Figure 3.2. Proposed pathway for camptothecin biosynthesis in *C. acuminata*. **Camptothecin (12) is synthesized from the central precursor strictosidinic acid (5) derived from condensation of tryptamine (4) and the iridoid secologanic acid (3), a monoterpenoid glycoside.**

3.3.1. Identification and characterization of new camptothecin pathway intermediates in *Camptotheca acuminata*

To assess the abundance of putative intermediates of the camptothecin pathway in *C. acuminata,* five different tissues (root, green stems, shoot apex and young and mature leaves) were collected from wild-type plants grown under greenhouse conditions, and 70% acetonitrile extracts were subjected to non-targeted HPLC/MS analysis using a 52-min reverse phase chromatographic separation. Nineteen metabolites were annotated based on accurate mass measurements of positive ions and fragments observed in HPLC/MS and UHPLC/MS/MS mass spectra (Table 3.1). Wild-type *C. acuminata* accumulated tryptamine (**4**), the iridoids loganic acid (**2**) and secologanic acid (**3**) as well as the MIA strictosidinic acid (**5**) but lacked detectable levels of their methyl-esterified derivatives loganin, secologanin and strictosidine. This suggests that in contrast to the best-studied MIA producing plants *Rauvolfia serpentina* and *C. roseus* that use methyl-esterified intermediates [24], *C. acuminata* uses an alternative seco-iridoid pathway with carboxylic acid intermediates, rather than methyl esters, up to strictosidinic acid. Several post-strictosidinic acid metabolites that we propose as intermediates in the camptothecin pathway were also detected (Table 3.2). In addition to the previously reported metabolites strictosamide (**6**) and pumiloside (**10**) [10, 12, 13], we identified metabolites with exact molecular masses and fragment ions consistent with strictosamide epoxide (**7**), strictosamide diol (**8**), strictosamide ketolactam (**9**), and deoxypumiloside (**11**) (Table 3.1). Surprisingly, most of these MIAs were present in multiple isomeric forms with identical exact molecular and fragment ion masses that were resolved and numbered according to their relative elution order. Strictosidinic acid (**5**) and strictosamide epoxide (**7**) each exhibited three isomers, while two isomers each were detected for strictosamide (**6**), strictosamide ketolactam (**9**), pumiloside (**10**) and deoxypumiloside (**11**).
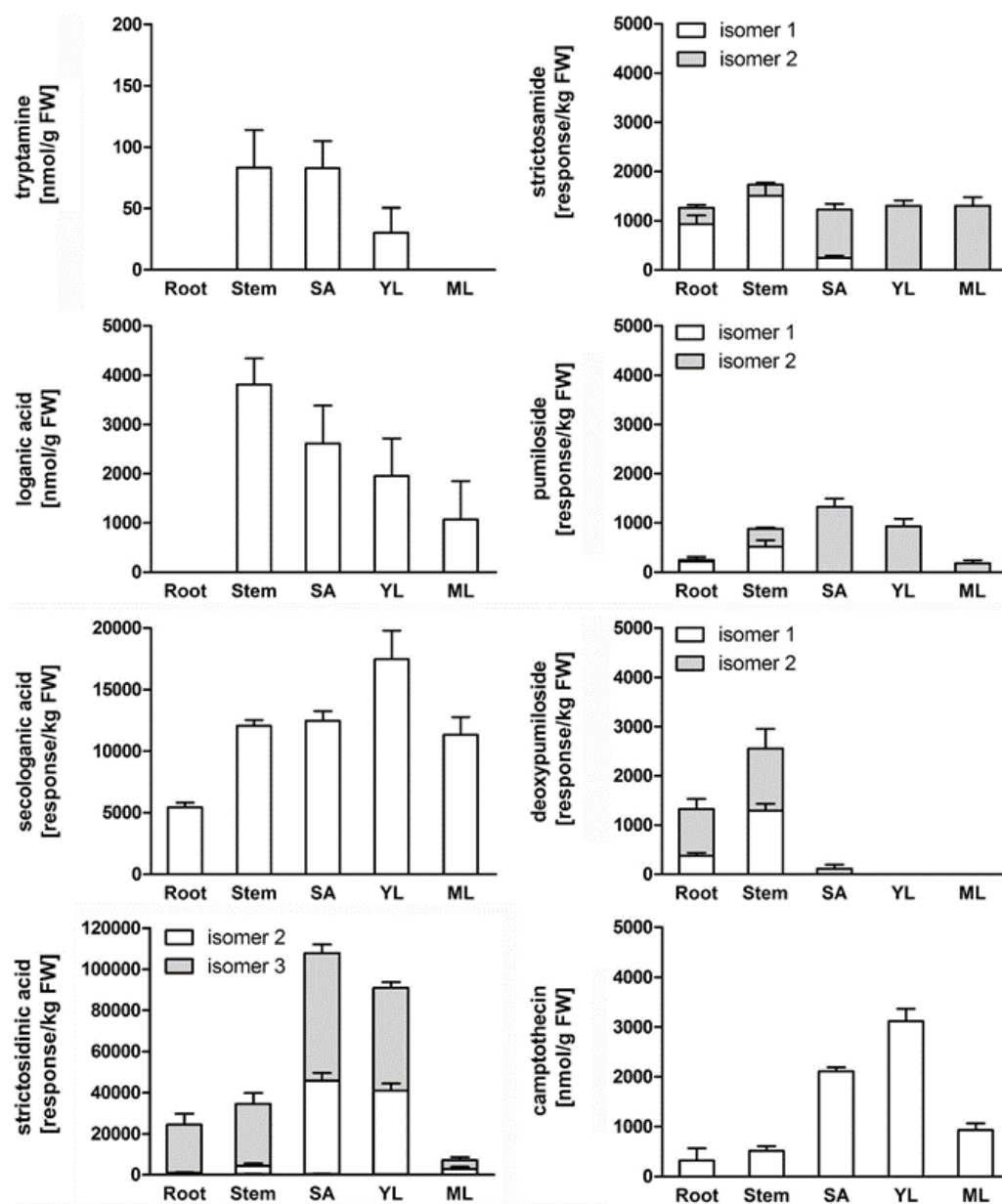
Figure 3.3. Tissue distribution profiles of proposed camptothecin pathway metabolites in wild-type *C. acuminata*. Tissues were collected from wild-type plants that had been under greenhouse cultivation in soil for eight months, and 70% acetonitrile extracts were analyzed using a 15-min gradient elution method for UHPLC/MS. Multiple isomers were detected for strictosidinic acid and post-strictosidinic acid metabolites. Average values are shown with SD (n= 3) for the most abundant and quantifiable isomers (SA, shoot apex; YL: young leaf; ML: mature leaf).

73

Next, the relative levels of these metabolites and their isomers were detected and quantified from extracts of five different tissues for three wild-type plants (Figure 3.3). As expected, shoot apex and young leaf shared similar metabolite profiles and accumulated most metabolites at distinctly higher levels, normalized to tissue mass, compared to root and/or mature leaf. Tryptamine (**4**) was only detected in stem, shoot apex and young leaf. Loganic acid (**2**) content was highest in stem, lower in other photosynthetic tissues and absent from root. Secologanic acid (**3**) was detected in all five tissues with levels in photosynthetic tissues two- to three-fold higher than in roots. The major accumulation sites for strictosidinic acid (**5**) were shoot apex and young leaf with progressively lower levels in stem, root and mature leaf. Surprisingly, the levels of strictosamide (**6**) were nearly identical in all tissues. Strictosamide epoxide (**7**), strictosamide diol (**8**), and strictosamide ketolactam (**9**) were detected (signal/noise > 3) but could not be reliably quantified due to their low abundances in all *C. acuminata* tissues. The tissue distribution profile of pumiloside (**10**) was similar to that of strictosidinic acid (**5**), with highest levels being in shoot apex and young leaf and lowest levels in mature leaf.

Deoxypumiloside (**11**) was only detected in stem, root and at extremely low levels in shoot apex.

All tissues contained camptothecin (**12**) with young leaf and shoot apex accumulating the highest levels. It is noteworthy that the relative amounts of isomers of strictosidinic acid (**5**) and post-strictosidinic acid intermediates varied among tissues (Figure 3.5). In most tissues, one of the two major isomers was more abundant, for example strictosamide (**6**) isomer 1 predominates in root and stem while isomer 2 predominates in shoot apex and leaves. In general, root and stem have similar isomer profiles as do young leaf and mature leaf.

On the basis of these observations, a model pathway for camptothecin synthesis in *C. acuminata* is proposed (Figure 3.2) where tryptamine (**4**) and the aldehyde-containing iridoid

secologanic acid (**3**) (rather than secologanin) are coupled via a Pictet-Spengler reaction, leading to strictosidinic acid (**5**). Intramolecular dehydrative cyclization of strictosidinic acid yields strictosamide (**6**). Conversion of the indole ring system to the quinoline ring is postulated to occur via multi-step oxidation to strictosamide ketolactam (**9**) followed by condensation and elimination of water to yield the quinolinone ring of pumiloside (**10**). The subsequent reduction to deoxypumiloside (**11**), deglycosylation, and other as yet undefined metabolic conversions lead to the end product camptothecin (**12**). This model pathway takes into account the formation and retention of multiple glycosidic isomers for pathway intermediates from 7-deoxyloganetic acid to deoxypumiloside (**11**) as discussed in Chapter 4.

3.3.2. Cyclase and tryptophan decarboxylase gene silenced plants

The genes CYC1 (annotated as progesterone 5-beta reductase based on sequence similarity, but termed cyclase locus 23532) and TDC1 (tryptophan decarboxylase 1) were selected as targets for reverse genetics analyses by gene silencing. Plants transformed with RNA interference (RNAi) for these genes were prepared by the DellaPenna laboratory using *Agrobacterium tumefaciens*, and transformed cells were induced to regenerate transgenic shoots, and later, shoots by adjusting the growth medium. After shoot elongation, selected plants were transferred to soil. With the exception of decreased resistance to red spider mite infestation, RNAi plants did not show any obvious phenotypic differences compared to wild-type plants (Figure 3.4).

Figure 3.4. *C. acuminata* **wild-type (WT) and two RNAi lines.** The plants had been under greenhouse cultivation in soil for approximately sixteen months. TDC1-RNAi and CYC1-RNAi do not show any obvious phenotypic differences compared to wild-type plants. Note that the tops of all plants were trimmed to reduce their height and encourage side branching.
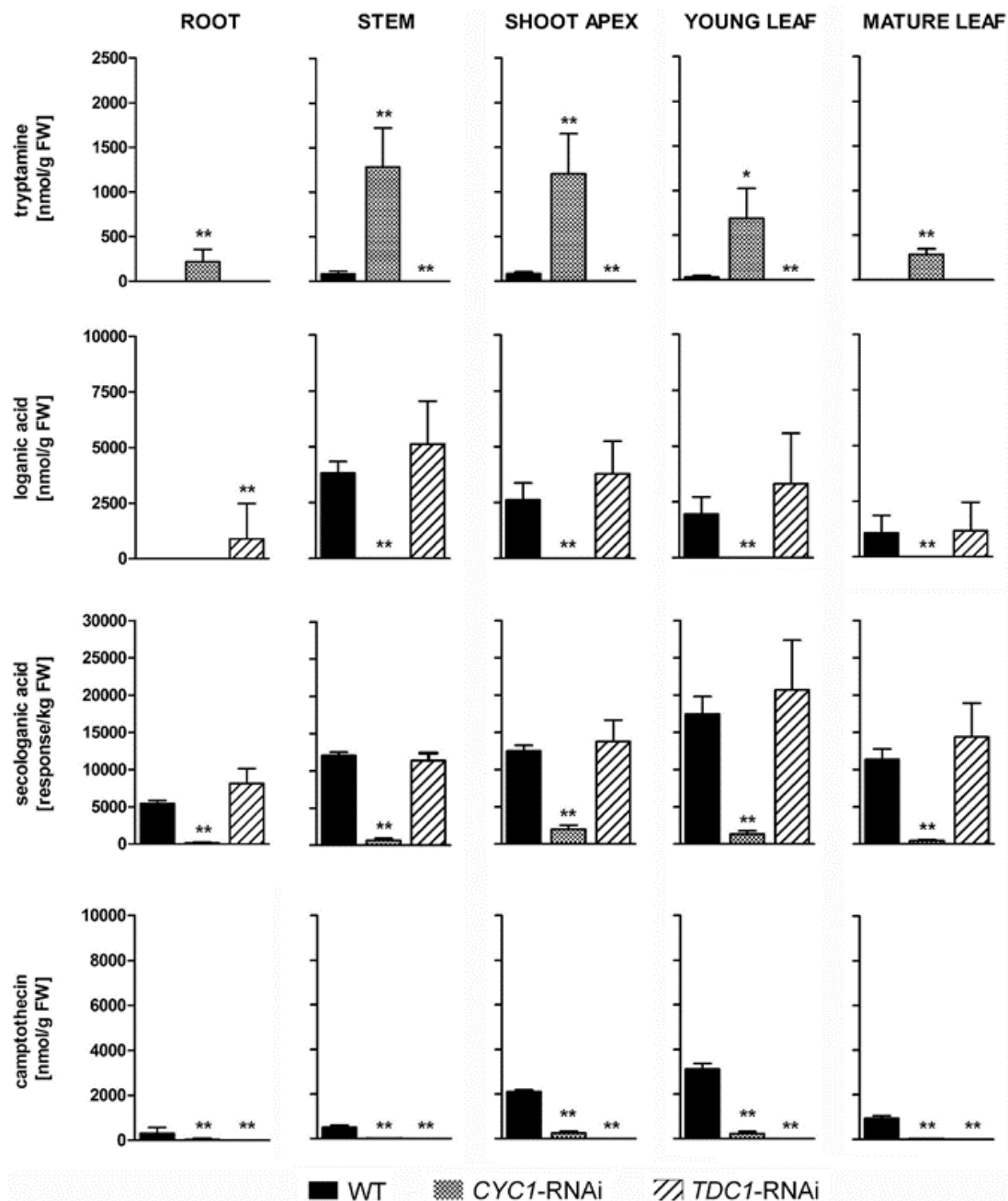
Figure 3.5. Levels of tryptamine (4), loganic acid (2), secologanic acid (3) and camptothecin (12) in different tissues of wild-type *C. acuminata* and CYC1-RNAi and TDC1-RNAi plants as measured using LC/MS. Tissues were collected from plants that had been under greenhouse cultivation for eight months. Average levels are shown with SD for wild-type (WT, n = 3), CYC1-RNAi (n = 5) and TDC1-RNAi lines (n = 5). Asterisks indicate significantly different metabolite levels in CYC1-RNAi lines and TDC1-RNAi (unpaired t-test; *, P<0.05; **, P<0.001) relative to wild-type.

3.3.3. Analysis of metabolites in TDC1-RNAi plants

A previous study demonstrated that *C. acuminata* possesses two differentially expressed tryptophan decarboxylase genes, TDC1 and TDC2, annotated as catalyzing synthesis of tryptamine from tryptophan [25]. Transcriptome data from the Medicinal Plant Genomics Resource (MPGR) database [26] indicated that TDC1 expression is moderate to high in leaves, roots and young bark of *C. acuminata* while TDC2 transcript is barely detectable in most plant tissues, with the exception of immature and mature fruit (not shown). Since TDC1 is expressed in tissues that accumulate camptothecin, it is considered the most promising candidate for synthesizing the indole precursor of camptothecin. The predicted TDC1 protein shares 68% and 74% sequence identities with the tryptophan decarboxylases of *C. roseus* [27] and *O. pumila* [28] respectively.

To evaluate the impact of TDC1 gene silencing on camptothecin biosynthesis, extracts from root, stem, shoot apex and young and mature leaves were subjected to HPLC/MS analyses. In contrast to the wild-type controls, TDC1-RNAi lines did not accumulate detectable levels of tryptamine (**4**), strictosidinic acid (**5**) or any of the post-strictosidinic acid metabolites (Figure 3.1), including the newly proposed pathway intermediates strictosamide epoxide (**7**), strictosamide diol (**8**) and strictosamide ketolactam (**9**). The levels of the iridoids loganic acid (**2**) and secologanic acid (**3**) in stem, shoot apex and leaves of TDC1-RNAi plants were not significantly different from wild-type levels (Figure 3.5). However, all TDC1-RNAi lines accumulated loganic acid (**2**) in roots, a metabolite undetectable in wild-type roots (Figure 3.5). This finding suggests that knockout/knockdown of TDC1 removes a sink for loganic acid in root tissues.

3.3.4. Analysis of metabolites in CYC1-RNAi plants

The impact of CYC1-RNAi was assessed by analyzing putative pathway metabolite levels in roots, stem, shoot apex and young and mature leaves of five independent CYC1-RNAi lines, and

comparing levels to those in wild-type (non-transformed) plants at the same stage of development. In addition to a sharp reduction in camptothecin levels in all tissues to as low as 2% of wild-type levels, all tissues of CYC1-RNAi lines accumulated substantially higher levels of tryptamine (**4**) compared to wild-type (Figure 3.5). In contrast to wild-type and TDC1-RNAi plants, which lacked detectable levels of tryptamine in roots and mature leaves, CYC1-RNAi roots and mature leaves contained about 250 nmol tryptamine per gram fresh weight (Figure 3.5). In stem, shoot apex, and young leaf of the CYC1-RNAi lines, tryptamine levels approached or exceeded 1000 nmol per gram fresh weight (15- to 23-fold increases relative to wild-type). None of the CYC1-RNAi lines accumulated detectable levels of loganic acid (**2**) in any tissue and in most tissues, the levels of secologanic acid (**3**), strictosidinic acid (**5**) and the post-strictosidinic acid intermediates were also reduced at least six-fold relative to wild-type, (Figure 3.5 and Table 3.2) with strictosamide epoxide (**7**), strictosamide diol (**8**) and strictosamide ketolactam (**9**) levels reduced to below the limit of quantification which is defined as signal/noise = 10. With regard to the isomer composition of strictosidinic acid (**5**) and post-strictosidinic acid metabolites, the levels of all identified isomers were decreased in CYC1-RNAi lines relative to those in wild-type, in some tissues to below the limits of quantification (Table 3.3). Isomer compositions were also radically altered in certain tissues of the CYC1-RNAi lines compared to wild-type. For example, strictosidinic acid (**5**) was composed of 39% isomer 2 and 61% isomer 3 in mature leaf of wild-type, whereas in CYC1-RNAi lines, isomer 3 was dominant (94%). Strictosamide (**6**) in root and stem of wild-type consisted mainly of isomer 1 (74% and 87%, respectively) while isomer 2 was exclusively accumulated in CYC1-RNAi lines. Pumiloside (**10**) in stem of wildtype was composed of 41% isomer 2 whereas in CYC1-RNAi lines, pumiloside isomer 1 was the only detectable isomer. Surprisingly, deoxypumiloside (**11**) isomer compositions did not significantly differ between wild-type and

CYC1-RNAi plants. These combined results for the effect of CYC1-RNAi are consistent with the

metabolite profiles of TDC1-RNAi lines and the proposed pathway for camptothecin biosynthesis

(Figure 3.1 and Figure 3.2).

Table 3.2. Isomer compositions of strictosidinic acid and post-strictosidinic acid metabolites in different tissues of CYC1-RNAi lines compared to wild-type. Tissues were collected from plants grown under greenhouse cultivation for eight months and metabolite levels determined by HPLC/MS. Average levels expressed as response per kg fresh weight are shown with SD for tissues from wild-type (n = 3) and CYC1-RNAi (n = 5) plants. Asterisks indicate significantly different metabolite levels in CYC1-RNAi lines (unpaired t-test; *, P<0.05; **, P<0.001) relative to wild-type. n.d. below limit of quantification, defined as signal/noise = 10.

Metabolite	Isomer	Wild-type					CYC1-RNAi				
		Root	Stem	Shoot Apex	Young Leaf	Mature Leaf	Root	Stem	Shoot Apex	Young Leaf	Mature Leaf
Strictosidinic acid	1	n.d.	447±147	475±87	342±54	n.d.	n.d.	n.d.	130±116**	42±58**	n.d.
	2	997±298	3,970±1098	45,252±3958	40,524±3540	2,808±1220	155±112**	1,025±822*	3,436±1811**	2,742±884**	2,070±733
	3	23,477±5219	30,066±5284	62,147±4213	50,004±2840	4,400±1379	3,020±1826**	1,209±712**	8,022±4086**	2,381±1108**	49±67**
Strictosamide	1	936±175	1,511±264	249±41	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
	2	323±65	222±24	975±121	1,305±107	1,306±170	165±74*	21±13**	125±35**	132±35**	149±127**
Pumiloside	1	219±90	520±130	n.d.	n.d.	n.d.	n.d.	152±90*	n.d.	n.d.	n.d.
	2	28±11	360±25	1,329±167	928±153	179±65	n.d.	n.d.	127±45**	n.d.	n.d.
Deoxypumiloside	1	373±66	1,291±141	n.d.	n.d.	n.d.	63±61**	372±169**	n.d.	n.d.	n.d.
	2	955±203	1,266±395	111±88	n.d.	n.d.	165±85**	436±108*	n.d.	n.d.	n.d.

3.3.5. Isomeric, acidic metabolic intermediates in the seco-iridoid pathway for camptothecin biosynthesis in *C. acuminata*

The results indicate that *C. acuminata* (order Cornales) uses a seco-iridoid pathway that fundamentally differs from that in *C. roseus* and *R. serpentina*, two plant species in the order Gentianales in which seco-iridoid and MIA synthesis have been most extensively studied [24]. *C. roseus* and *R. serpentina* form the methyl esters loganin and secologanin to produce the key MIA precursor strictosidine. These compounds are absent from *C. acuminata* and instead, the analogous carboxylic acids loganic acid (**2**) and secologanic acid (**3**) accumulate and are utilized as precursors of strictosidinic acid (**5**) (Table 3.1). In this regard, *C. acuminata* also differs from *O. pumila*, the only other camptothecin-producing species for which some biochemical and molecular data exist. *O. pumila* is also in the order Gentianales and like *C. roseus* and *R. serpentina*, utilizes methyl ester intermediates to produce strictosidine [14]. Metabolomics data generated in the current investigation indicate that at least two routes for camptothecin have evolved in plants and allow proposal of a model pathway for *C. acuminata* (Figure 3.1) that is identical to the "Gentianales-type" seco-iridoid pathway only up to loganic acid formation and thereafter differs [14, 18, 29].

In *C. acuminata*, tryptamine (**4**) is condensed with secologanic acid (**3**) to form strictosidinic acid (**5**), instead of with secologanin to form strictosidine as in the order Gentianales. Moreover, plant species of the order Gentianales form a single isomer of strictosidine as the initial MIA intermediate [14, 29] while *C. acuminata* produces a mixture of strictosidinic acid (**5**) isomers as its initial MIA intermediate (Figure 3.2 and Table 3.2).

Silencing of two essential genes for synthesis of the monoterpene and indole components of camptothecin, CYC1 and TDC1, respectively, provides strong evidence for the involvement of specific metabolites in the proposed camptothecin pathway in *C. acuminata*. CYC1 RNAi lines

with low CYC1 transcript levels had levels of the iridoid loganic acid below detection limits, and CYC1 RNA interference significantly lowered levels of secologanic acid compared to wild-type (Figure 3.3). Additionally, the levels and isomers of all strictosidinic acid (**5**) and post-strictosidinic acid intermediates (including camptothecin) were significantly decreased in CYC1-RNAi lines relative to wild-type (Figure 3.3 and Table 3.2). Silencing of TDC1 greatly impaired the plants' production of tryptamine (**4**) and all indole-containing intermediates/isomers of the proposed camptothecin pathway (Figure 3.3). Most significantly, TDC1-RNAi and CYC1-RNAi lines impacted, to different degrees, the same strictosidinic acid (**5**) and post-strictosidinic acid intermediates/isomers.

Combining non-targeted metabolite profiling in wild-type and RNAi lines allowed us to identify new pathway intermediates of extremely low abundance in wild-type (*i.e.,* strictosamide epoxide, diol and ketolactam) that are consistent with the biosynthetic scheme shown in Figure 3.2. These three novel metabolites are likely formed from isomeric strictosamide (**6**) yielding first strictosamide epoxide (**7**) that is further converted to strictosamide diol (**8**) and then to its ketolactam (**9**), which undergoes intramolecular cyclization of its enolate to carbonyl followed by dehydration, in a manner similar to the well-known aldol condensation. This process channels strictosamide (**6**) towards pumiloside (**10**) synthesis by completing the quinolone ring topology that provides the framework for camptothecin. In the proposed pathway, pumiloside isomers are then reduced to deoxypumiloside (**11**) isomers. Subsequent metabolic conversions involve deglycosylation and oxidations that convert the C-atom site of glycosylation (illustrated in Figure 3.1 as having variable stereochemical configuration) to a carbonyl (as described in chapter 4), with concomitant loss of chirality so that only one product of a single stereochemistry, camptothecin, is formed.

The abundance of glucoside isomer mixtures for the majority of the camptothecin biosynthetic pathway, from loganic acid (**2**) to deoxypumiloside (**11**) is both remarkable and unprecedented in MIA biosynthesis. This has profound implications from a mechanistic perspective because it suggests that some, if not all, of the enzymes from 7-deoxyloganetic acid to deoxypumiloside (**11**) are able to accommodate multiple isomers or that multiple enzymes exist for each step that are specific for a single stereoisomer. Given that the metabolic transformations from strictosidinic acid (**5**) to deoxypumiloside (**11**) are remote from the site of glycosylation, the consistent presence of multiple intermediate isomers in this portion of the pathway suggests that a single enzyme could accommodate both stereoisomer substrates.

However, the variation in the isomeric composition of strictosidinic acid (**5**) and post-strictosidinic acid compounds between different tissues in wild-type as well as the marked changes in the isomer composition of certain metabolites in some CYC1-RNAi tissues (Table 3.2) indicate that additional stereoselective sinks or processes are present in *C. acuminata*.

3.4. Conclusions

Recent advances in genomics, computational biology, and analytical chemistry now allow efficient investigations into natural product biosynthesis in previously intractable and non-model medicinal plant species. In this study, quantitative metabolite profiling demonstrates functions of candidate biosynthetic genes in a non-model medicinal plant. Metabolic phenotypes in non-transformed and silenced *Camptotheca acuminata* were assessed by profiling and quantifying the metabolites in root, stem, shoot apex, young leaf and old leaf tissues. Metabolites were annotated using accurate measurements of $[M+H]^+$ and fragment ions generated by non-selective

Multiplexed CID which provides a fast and efficient way of performing non-targeted metabolite profiling to identify metabolites based on their functional groups.

Cyclase-silenced transgenic lines exhibited decreased levels of key terpenoid intermediates involved in the biosynthetic pathway of camptothecin. Cyclase knockout plants showed near-complete elimination of loganic acid, an upstream iridoid intermediate generated following reductive cyclization of 8-oxogeranial. Similarly, secologanic acid, strictosidinic acid and strictosamide were absent in these transgenic plants. Furthermore, levels of tryptamine were significantly higher in the cyclase silenced lines. This is attributed to the accumulation of unused tryptamine since the cyclase knockout plants do not produce iridoid intermediates that condense with tryptamine. In addition, the silencing of tryptophan decarboxylase yielded dramatic reductions in all downstream intermediates. The data revealed that the levels of camptothecin in the cyclase-silenced lines were decreased to less than 10% of the levels in the non-transformed plant tissues, while the tryptophan decarboxylase-silenced lines showed less than 1% of the same controls. These data have allowed for the discovery of the cryptic chemical transformations required of the intermediates involved in the biosynthetic pathway of camptothecin. Two essential genes for camptothecin biosynthesis, tryptophan decarboxylase and iridoid synthase, were identified and down regulated by RNAi, and the consequences for metabolite and gene expression profiles determined. The presence of multiple isomers throughout the pathway is consistent with their use in synthesis before finally being resolved to a single camptothecin isomer, much like a multi-lane highway allows parallel tracks to converge at a common destination. A model "racemic" pathway for camptothecin biosynthesis in *C. acuminata* is proposed that fundamentally differs from previously studied MIA pathways and may be the first "racemic" metabolic pathway identified in plants.

REFERENCES

REFERENCES

1.   De Luca, V., et al., *Making iridoids/secoiridoids and monoterpenoid indole alkaloids: progress on pathway elucidation.* Current Opinion in Plant Biology, 2014. **19**: p. 35-42.

2.   O'Connor, S.E. and J.J. Maresh, *Chemistry and biology of monoterpene indole alkaloid biosynthesis.* Nat Prod Rep, 2006. **23**(4): p. 532-47.

3.   Wall, M.E., et al., *Plant Antitumor Agents. I. The Isolation and Structure of Camptothecin, a Novel Alkaloidal Leukemia and Tumor Inhibitor from Camptotheca acuminata1,2.* Journal of the American Chemical Society, 1966. **88**(16): p. 3888-3890.

4.   Wright, C.M., et al., *DNA topoisomerase I domain interactions impact enzyme activity and sensitivity to camptothecin.* J Biol Chem, 2015. **290**(19): p. 12068-78.

5.   Liu, Y.Q., et al., *Perspectives on biologically active camptothecin derivatives.* Med Res Rev, 2015. **35**(4): p. 753-89.

6.   Lorence, A., F. Medina-Bolivar, and C.L. Nessler, *Camptothecin and 10-hydroxycamptothecin from Camptotheca acuminata hairy roots.* Plant Cell Reports, 2004. **22**(6): p. 437-441.

7.   Wink, M., *Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective.* Phytochemistry, 2003. **64**(1): p. 3-19.

8.   Gopalakrishnan, R. and B. Shankar, *Multiple shoot cultures of Ophiorrhiza rugosa var. decumbens Deb and Mondal--a viable renewable source for the continuous production of bioactive Camptotheca alkaloids apart from stems of the parent plant of Nothapodytes foetida (Wight) Sleumer.* Phytomedicine, 2014. **21**(3): p. 383-9.

9.   Sheriha, G.M. and H. Rapoport, *Biosynthesis of Camptotheca acuminata alkaloids.* Phytochemistry, 1976. **15**(4): p. 505-508.

10. Hutchinson, C.R., et al., *Biosynthesis of camptothecin. 3. Definition of strictosamide as the penultimate biosynthetic precursor assisted by carbon-13 and deuterium NMR spectroscopy.* Journal of the American Chemical Society, 1979. **101**(12): p. 3358-3369.

11. Aimi, N., et al., *Pumiloside and deoxypumiloside; plausible intermediates of camptothecin biosynthesis.* Tetrahedron Letters, 1989. **30**(37): p. 4991-4994.

12. Carte, B.K., et al., *Isolation and characterization of a presumed biosynthetic precursor of camptothecin from extracts of Camptotheca acuminata.* Tetrahedron, 1990. **46**(8): p. 2747-2760.

13. Montoro, P., et al., *Metabolite fingerprinting of Camptotheca acuminata and the HPLC–ESI-MS/MS analysis of camptothecin and related alkaloids.* Journal of Pharmaceutical and Biomedical Analysis, 2010. **51**(2): p. 405-415.

14. Asano, T., et al., *Suppression of camptothecin biosynthetic genes results in metabolic modification of secondary products in hairy roots of Ophiorrhiza pumila.* Phytochemistry, 2013. **91**: p. 128-39.

15. Yamazaki, M., T. Asano, and K. Saito, *Camptothecin Production and Biosynthesis in Plant Cell Cultures*, in *50 Years of Phytochemistry Research: Volume 43*, R.D. Gang, Editor. 2013, Springer International Publishing: Cham. p. 43-54.

16. Hofer, R., et al., *Geraniol hydroxylase and hydroxygeraniol oxidase activities of the CYP76 family of cytochrome P450 enzymes and potential for engineering the early steps of the (seco)iridoid pathway.* Metab Eng, 2013. **20**: p. 221-32.

17. Geu-Flores, F., et al., *An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis.* Nature, 2012. **492**(7427): p. 138-42.

18. Salim, V., et al., *7-deoxyloganetic acid synthase catalyzes a key 3 step oxidation to form 7-deoxyloganetic acid in Catharanthus roseus iridoid biosynthesis.* Phytochemistry, 2014. **101**: p. 23-31.

19. Asada, K., et al., *A 7-deoxyloganetic acid glucosyltransferase contributes a key step in secologanin biosynthesis in Madagascar periwinkle.* Plant Cell, 2013. **25**(10): p. 4123-34.

20. Murata, J., et al., *The leaf epidermome of Catharanthus roseus reveals its biochemical specialization.* Plant Cell, 2008. **20**(3): p. 524-42.

21. Salim, V., et al., *Virus-induced gene silencing identifies Catharanthus roseus 7-deoxyloganic acid-7-hydroxylase, a step in iridoid and monoterpene indole alkaloid biosynthesis.* Plant J, 2013. **76**(5): p. 754-65.

22. Irmler, S., et al., *Indole alkaloid biosynthesis in Catharanthus roseus: new enzyme activities and identification of cytochrome P450 CYP72A1 as secologanin synthase.* Plant J, 2000. **24**(6): p. 797-804.

23. Bracher, D. and T.M. Kutchan, *Strictosidine synthase from Rauvolfia serpentina: analysis of a gene involved in indole alkaloid biosynthesis.* Arch Biochem Biophys, 1992. **294**(2): p. 717-23.

24. De Luca, V., et al., *Making iridoids/secoiridoids and monoterpenoid indole alkaloids: progress on pathway elucidation.* Curr Opin Plant Biol, 2014. **19**: p. 35-42.

25. Lopez-Meyer, M. and C.L. Nessler, *Tryptophan decarboxylase is encoded by two autonomously regulated genes in Camptotheca acuminata which are differentially expressed during development and stress.* Plant J, 1997. **11**(6): p. 1167-75.

26.   Gongora-Castillo, E., et al., *Development of transcriptomic resources for interrogating the biosynthesis of monoterpene indole alkaloids in medicinal plant species.* PLoS One, 2012. **7**(12): p. e52506.

27.   De Luca, V., C. Marineau, and N. Brisson, *Molecular cloning and analysis of cDNA encoding a plant tryptophan decarboxylase: comparison with animal dopa decarboxylases.* Proc Natl Acad Sci U S A, 1989. **86**(8): p. 2582-6.

28.   Yamazaki, Y., et al., *Camptothecin biosynthetic genes in hairy roots of Ophiorrhiza pumila: cloning, characterization and differential expression in tissues and by stress compounds.* Plant Cell Physiol, 2003. **44**(4): p. 395-403.

29.   Miettinen, K., et al., *The seco-iridoid pathway from Catharanthus roseus.* Nat Commun, 2014. **5**: p. 3606.

Chapter 4: Purification and structure elucidation of abundant intermediates of biosynthesis of the monoterpene indole alkaloid camptothecin

4.1. Introduction

Camptotheca acuminata has gained the attention of many medical researchers and medicinal chemists around the world since the discovery of camptothecin in the stem tissues of the plant in 1966 [1]. A variety of specialized metabolites have been isolated from different tissues of the plant, including alkaloids, ellagic acids and flavonoids [2]. To date, approximately 3 dozen alkaloids have been reported from *C. acuminata*, and most of these metabolites are monoterpenoid indoles or a variation thereof [2, 3]. However, few publications have described the chemical diversity of alkaloid metabolites in this medicinal plant.

Metabolite profiling of wild-type and RNAi lines identified plausible intermediates for missing pathway steps and demonstrated nearly all camptothecin pathway intermediates are present as multiple isomers, as presented in Chapter 3. Unlike previously characterized MIA-producing plants, *C. acuminata* does not synthesize 3-alpha(*S*)-strictosidine, which contains a methyl ester group, as its central MIA intermediate. Instead, multiple isomers of strictosidinic acid were observed. These results have been interpreted as evidence that the plant uses an alternative seco-iridoid pathway to produce these isomers, but information regarding the structural differences between isomers would help guide conclusions regarding the roles of these isomeric metabolites.

Though isomerism in the *C. acuminata* metabolome is not limited to strictosidinic acid, isomers of this compound were chosen for purification in this study because they present the most abundant pair of isomers in *C. acuminata* of potential precursors of camptothecin. Young leaf and shoot apex have been observed to house the highest quantities of strictosidinic acid (Figure 3.3), but due to the low availability of young leaves, mature leaves of *Camptotheca acuminata* were also harvested in order to perform large-scale isolation and elucidation of metabolites. The exact structure of the strictosidinic acid isomers will help answer questions such as the nature of

strictosidinic acid isomerism and may suggest a biochemical basis for the formation of multiple

isomers.
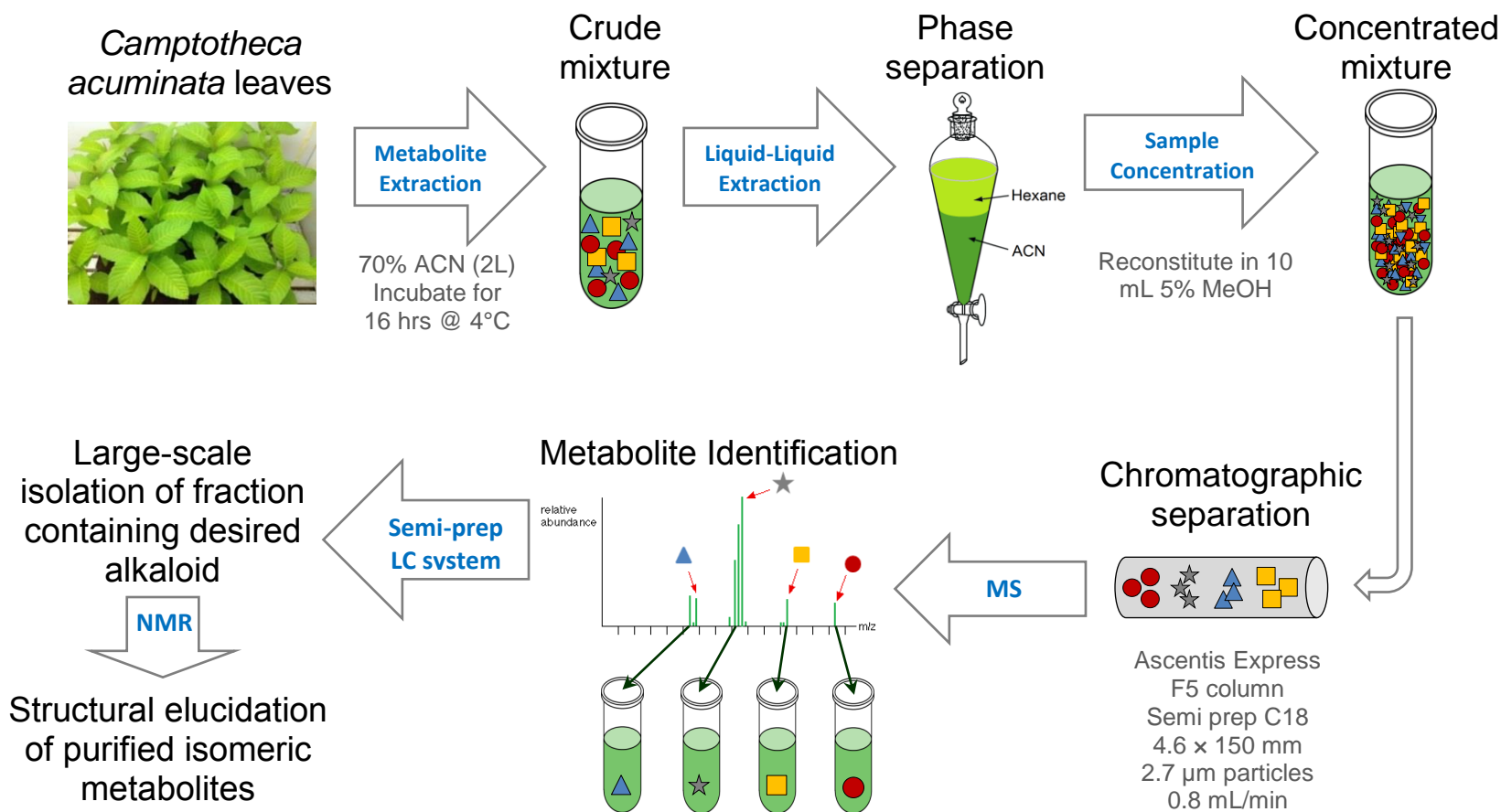
Figure 4.1. General workflow depicting the process used to separate and purify the strictosidinic acid isomers. Freshly harvested *C. acuminata* leaves (350 g) were extracted using 2 L of 70% acetonitrile/30% water to obtain a crude mixture that was subjected to liquid-liquid phase partitioning against hexane. The lower acetonitrile phase was collected and dried under reduced pressure and dissolved in 10 mL of 5% methanol/95% water to produce a concentrated mixture. Aliquots (150 µL) were injected into the LC/MS system for metabolite separation and purification. The purified materials were analyzed using 1D and 2D Nuclear Magnetic Resonance (NMR) spectroscopy. Colored shapes represent metabolites that undergo chromatographic separation in the F5 (pentafluorophenylpropyl) column.

4.2 Materials and methods

4.2.1. Reagents

HPLC grade acetonitrile, methanol and formic acid (88% aqueous solution) were purchased from VWR Scientific. n-Hexane 95%, Omnisolv grade, was from EMD Millipore Chemicals. Deuterated methanol (CD$_3$OD 99.8%) was purchased from Sigma-Aldrich, St. Louis, MO, USA.

4.2.2. Purification of strictosidinic isomers

As depicted in Figure 4.1, leaf tissue (350 g fresh weight) was freshly harvested from greenhouse-cultivated *C. acuminata* plants, transferred to a 4 L amber glass bottle, and extracted using 2 L of 70% HPLC grade acetonitrile/30% water for 16 hours at 4°C. The extract was ultrasonicated for 2 minutes followed by liquid-liquid partitioning against two volumes of 500 mL hexane. The lower acetonitrile layer was collected, and solvent was removed under reduced pressure in a rotary evaporator and reconstituted in 10 mL of methanol/water (5/95 v/v) immediately before UHPLC/MS metabolite profiling. UHPLC/MS metabolite profiling was performed using a Shimadzu LC-20AD ternary pump coupled to a SIL-5000 autosampler, column oven. Mass spectra were acquired using a Waters LCT Premier mass spectrometer equipped with an electrospray ionization source and operated in positive-ion mode. Preparative HPLC fractionation was performed using a Waters Model 2795 HPLC system coupled to a LKB BROMMA 221 fraction collector. An Ascentis Express F5 (pentafluorophenylpropyl) column (4.6 x 150 mm, 2.7 μm particles; Supelco Sigma-Aldrich) was used with a solvent flow rate of 0.8 mL/min for profiling and preparative fractionation. The flow was split post-column, and approximately 0.3 mL/min was diverted to the mass spectrometer. The mobile phase consisted of water (Solvent A) and methanol (Solvent B) using linear gradients: 10% B during 0-1 min, to 30%

B at 2 min, to 50% B at 14 min, to 65% B at 15 min, to 90% B at 25 min, then 90-99% B over 25-32 min followed by a hold until 36 min, then return to 10% B over 36.01-40 min. Fractions were collected at 15-second intervals, and 50 injections of 150 µL were made in order to accumulate sufficient material for Nuclear Magnetic Resonance (NMR) analysis.

Two major strictosidinic acid isomers, denoted isomers 2 and 3 based on the order of their elution times, were collected in fractions 42-46 and 60-66, respectively. Fractions for each isomer were combined and dried under reduced pressure.

4.2.3. NMR experiments

For all NMR experiments, compounds were dried for 48 hours under reduced pressure to minimize the NMR signals from the solvents used during the extraction procedure. Dried compounds were redissolved in 99.8% CD_3OD and placed into CD_3OD susceptibility matched Shigemi NMR tubes. NMR spectra (^{1}H, J-resolved ^{1}H, ^{1}H-^{1}H COSY, HSQC, HMBC) were acquired on a 21.1 T Bruker Avance-900 NMR spectrometer equipped with TCI inverse triple-resonance cryoprobe at 900 MHz (^{1}H) and 225 MHz (^{13}C) at the Michigan State University Max T. Rogers NMR Facility. Residual non-deuterated solvent signal at δ^1H (CD_3OD) = 3.34 ppm and $\delta^{13}C$ (CD_3OD) = 49.08 ppm were used to calibrate NMR spectra.

4.3. Results and discussion

4.3.1. Analysis of metabolites using HPLC/MS

A

B
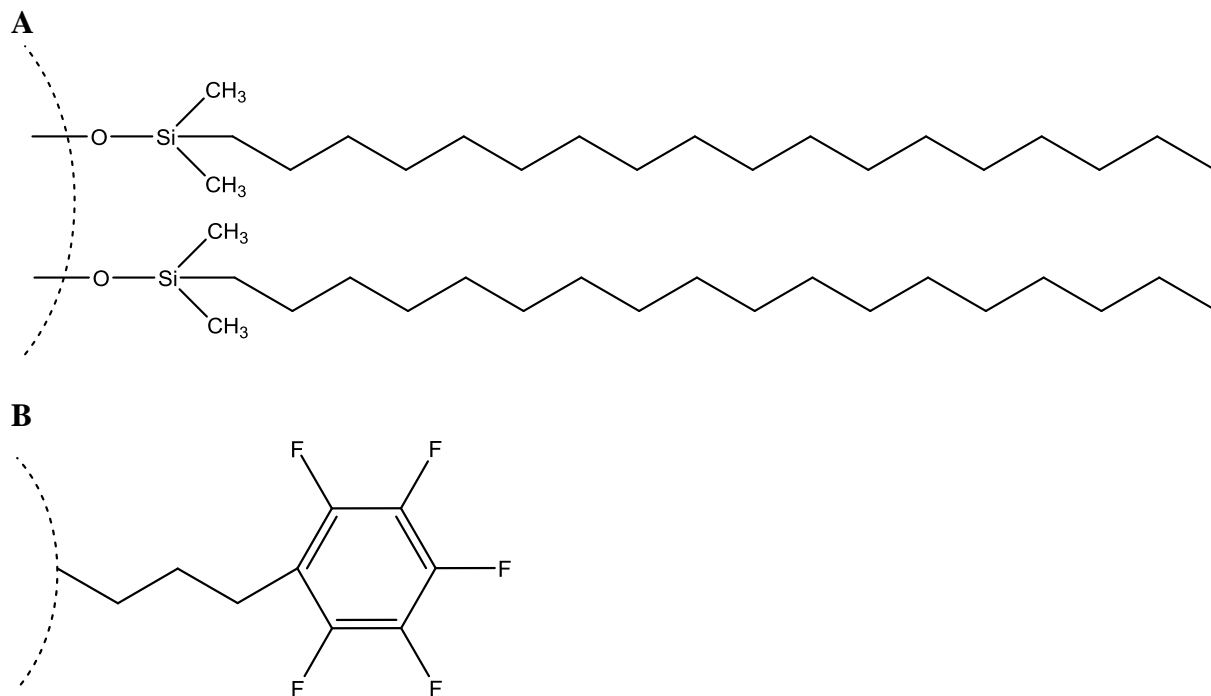
Figure 4.2. (A) General structure of the stationary phase of (A) C18 column and (B) pentafluorophenylpropyl (PFPP) column. Dotted line represents the surface of the particle.

Based on the LC separation shown in Chapter 3 (Figure 3.1), the two most abundant strictosidinic acid isomers elute within 2 mins of each other during the 52-min gradient mentioned in Chapter 3. The 1.7-minute separation in retention times was sufficient for profiling of individual isomers on the analytical column (2.1 × 100 mm, 2.7 μm), however, when transferring the gradient unto a larger column (4.6 x 150 mm, 2.7 μm particles) for large-scale purification, the retention times of the two peaks were within 0.5 min of each other. The decrease in separation can be attributed to the reduced theoretical plates characteristic of larger column diameters [4] and the large injection volume relative to total column volume. To improve chromatographic resolution of

96

the strictosidinic acid isomers, a F5 column was useful for separation of the two isomers, yielding a difference of 4.5 mins in retention times of the two isomers of strictosidinic acid.   The column chemistry of the F5 column significantly differs from that of the C18 column. C18 columns have a hydrophobic 18-carbon chain that is suitable for separating non-polar metabolites. Though the C18 column is generally useful for separation of strictosidinic acid isomers from other metabolites (Figure 3.1), more selectivity in retention was desired for separating the isomers from each other. The pentafluorophenylpropyl group in the F5 column has shown stronger retention than C18 phases for basic nitrogen-containing compounds that has been attributed to a hydrophobically-assisted cation-exchange mechanism that has yet to be fully explained [5]. The fluorines also impart a partial positive charge on the internal region of the ring while imparting a partial negative charge on the outer edge. The 3 conjugated π bond system in the phenyl group presents a large localized dipole, and is able to better retain the electron-rich alkaloid metabolites.
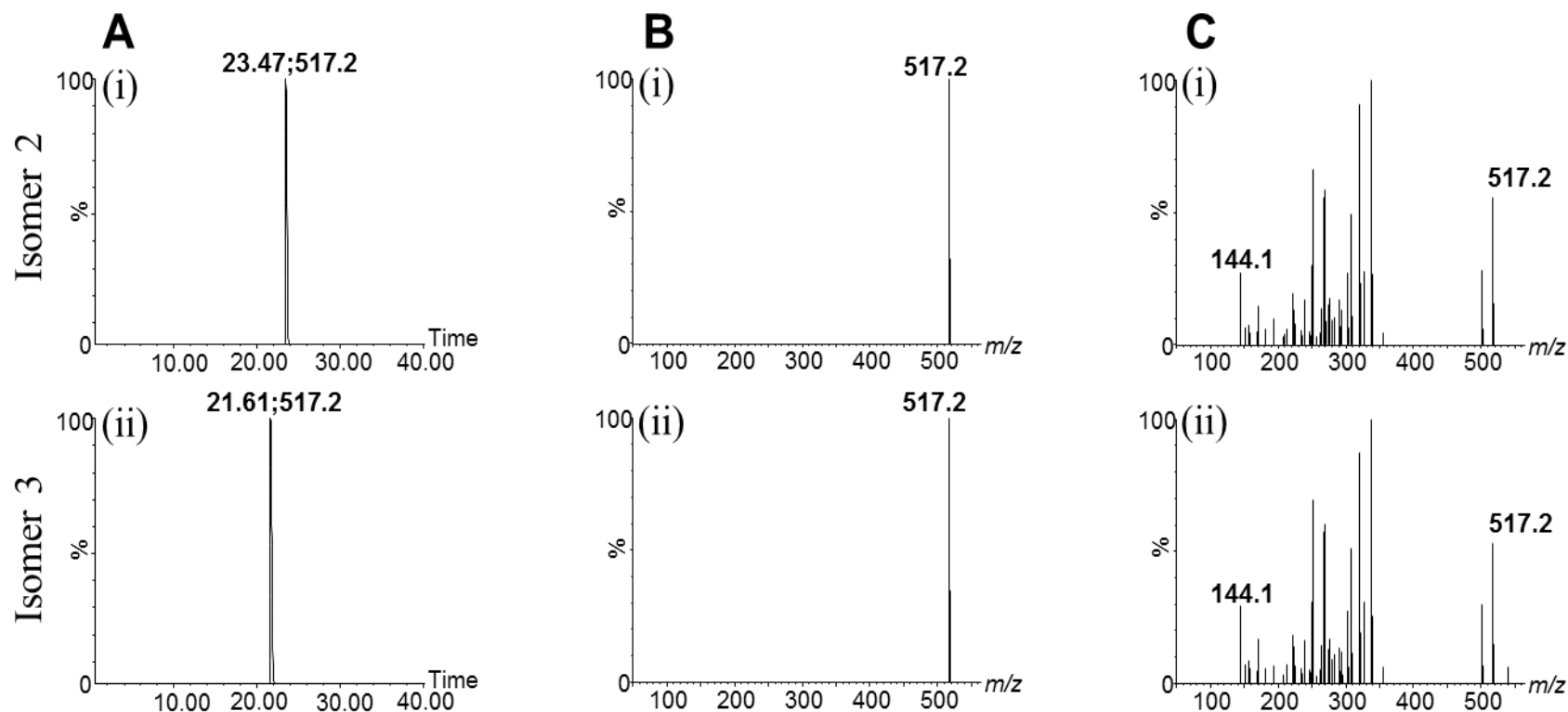
Figure 4.3. (A) UHPLC/MS base peak ion chromatogram generated from analysis of purified strictosidinic acid isomers (i) 2 and (ii) 3. (B) Positive ion mode multiplexed CID mass spectrum at collision potential (at aperture 1) of 15 V for isomers (i) 2 and (ii) 3 strictosidinic acid detected as [M+H]⁺ at *m/z* 517.2. (C) Positive ion mode multiplexed CID mass spectrum at collision potential (at aperture 1) of 60 V for isomers (i) 2 and (ii) 3 strictosidinic acid at *m/z* 517.2. Both isomers displayed identical fragments of the same *m/z* values, including the smallest fragment at *m/z* 144.1. Analysis was performed using the 52-min gradient mentioned in chapter 3.

4.3.2. Structure elucidation of strictosidinic acid isomers

The most HPLC/MS suitable conditions, including column and solvent gradient, were determined for isolating the two most abundant isomers of strictosidinic acid from crude extracts of *C. acuminata* leaf tissue. After purification and drying under reduced pressure, both compounds were isolated as white amorphous powders. A small portion of each compound was tested on the HPLC/MS to ensure that there was no change in mass or evidence of degradation from the purification and drying process. Based on MS analysis of the two chromatographically separated isomers, the molecular formula was assigned as $C_{26}H_{32}N_2O_9$ with the molecular ion present at *m/z* 517 $[M+H]^+$ for both isomer 2 and 3. The major fragment representing neutral loss of a glucose moiety was observed at *m/z* 355 of $[M+H-162]^+$ for isomer 2 (experimental *m/z* 355.1808; theoretical *m/z* 355.1652; Δm= 43.9) and isomer 3 (experimental *m/z* 355.1738; theoretical *m/z* 355.1652; Δm=24.2). At higher collision energies in the Multiplexed CID process, more fragments were generated. Among the smallest fragments to be observed was at *m/z* 144, which contains the indole moiety in an ion attributed to $C_{10}H_{10}N^+$ (Figure 4.3).

The two purified isomers of strictosidinic acid were dissolved in CD_3OD and subjected to 1D and 2D NMR analysis. For isomer 2, the ^{1}H NMR spectrum exhibited an ortho-substituted phenyl correlating to ring A, with aromatic protons at δ_H 7.55 (1H, *d*, *J* = 8.2 Hz, H-9), δ_H 7.45 (1H, *d*, *J* = 8.2 Hz, H-12), δ_H 7.23 (1H, *t*, *J* = 7.6 Hz, H-11) and δ_H 7.13 (1H, *t*, *J* = 7.3 Hz, H-10) (Table 4.1). Likewise, for isomer 3, the ^{1}H NMR spectrum exhibited an ortho-substituted phenyl correlating to ring A, with aromatic protons at δ_H 7.56 (1H, *d*, *J* = 7.9 Hz, H-9), δ_H 7.42 (1H, *d*, *J* = 8.2 Hz, H-12), δ_H 7.23 (1H, *t*, *J* = 7.6 Hz, H-11) and δ_H 7.15 (1H, *t*, *J* = 7.4 Hz, H-10) (Table 4.1). By comparing the aforementioned chemical shifts corresponding to ring A of both isomers, it can be concluded that there is no discernable difference in the indole portion of the structure.

Three olefinic protons, corresponding to a terminal double bond, were detected at δ_H 6.19 (1H, *dd*, *J* = 10.2, 17.2 Hz, H-19), δ_H 5.47 (1H, *d*, *J* = 17.3 Hz, H-18) and δ_H 5.39 (1H, *d*, *J* = 11.0 Hz, H-18) for isomer 2. The same protons were observed at δ_H 5.96 (1H, *dd*, *J* = 10.3, 17.5 Hz, H-19), δ_H 5.43 (1H, *d*, *J* = 17.3 Hz, H-18) and δ_H 5.32 (1H, *d*, *J* = 10.7 Hz, H-18) for isomer 3. Yet again, the differences in chemical shifts between the two isomers are negligible (Table 4.1). For isomer 2, the proton spectrum showed an anomeric proton at δ_H 4.85 (1H, *d*, *J* = 7.9 Hz, H-1') and an anomeric carbon at δ_C 98.8 corresponding to the single glucose unit. Similarly, for isomer 3, the anomeric proton at δ_H 4.92 (1H, *d*, *J* = 8.0 Hz, H-1') and anomeric carbon at δ_C 98.9 corresponds to a glucose unit (Table 4.1). The Heteronuclear Multiple Bond Correlation (HMBC) spectrum exhibited key correlations of the same anomeric proton at δ_H 4.85 (1H, *d*, *J* = 7.9 Hz, H-1') to C-21 (δ_C = 95.5) for isomer 2. Likewise, the correlation of the anomeric proton at δ_H 4.92 (1H, *d*, *J* = 8.0 Hz, H-1') to C-21 (δ_C 95.1) for isomer 3 was observed. These HMBC correlations suggest that the glucosyl moiety is attached to C-21 on both isomers (Figure 4.8 and 4.13).

The ^{1}H-^{1}H correlation spectroscopy (COSY) spectra showed marked differences between the two isomers. The signal corresponding to the coupling of the single protons on C-20 (δ_H 2.74, *m*) and C-21 (δ_H 2.70, *m*) in the COSY spectrum for isomer 2 and isomer 3 showed a difference of 0.6 ppm along the axis pertaining to the chemical shift of the proton on C-21. This indicates that the major differences in the two isomers lies in the stereochemistry of position C-21 (Figure 4.9 and 4.14). The chemical shift of position C-19 in isomer 2 (δ_H 6.19) is farther downfield compared to C-19 of isomer 3 (δ_H 5.96) according to their Heteronuclear Single Quantum Coherence (HSQC) spectra. The HSQC spectra for both isomers also show differences in chemical shifts for position C-21 (Figure A.2 and A.8). Based on NMR and MS experiments, the chemical shifts and coupling constants for both isomers are shown in Table 4.1, and the structures are shown in Figure 4.4.

Subsequent coupled HSQC NMR spectra were generated to assess whether relative stereochemical configuration of hydrogens at chiral positions could be assessed by measuring the ^{1}H-^{13}C coupling constants, which are sensitive to orientation relative to other bonds, and have been used to establish relative bond orientations in a variety of metabolites (Marquez et al. 2001). These coupling constants ($^1J_{C-H}$) were 178 Hz for isomer 2 and 170 Hz for isomer 3, consistent with two epimers differing in the stereochemistry of C-glycosylation at position 21 (Table 4.1). The data suggest that the stereochemistry of position C-21 is the only major difference between the two isomers. The presence of similar dominant isomer pairs in further downstream metabolites (e.g. strictosamide, deoxypumiloside, and pumiloside) suggest that all MIAs in the pathway are also present as isomeric mixtures and are converted into isomeric intermediates downstream. Prior to glycosylation, equilibration between the open- and closed-ring conformations of 7-deoxyloganetic acid yields stereoisomers at the C2 carbon atom, which is subsequently glycosylated (Figure 4.5). This proposed mechanism is consistent with the detection of two isomers for loganic acid and secologanic acid (Sadre et al, 2016) as well as the presence of isomer mixtures for strictosidinic acid and most post-strictosidinic acid intermediates (Table 3.1).

Based on similarities to NMR spectra reported for 21(*S*)-strictosidine, strictosidinic acid isomer 3 was assigned to have the 21(*R*) configuration and strictosidinic acid isomer 2 to have 21(*S*) configuration, the latter being in common with the published structure and coupling constants of 21(*S*)-strictosidine (Patthy-Lukáts et al. 1997). NMR data for stereocenters at positions 3, 15, 20, and the carbohydrate resonances for isomers 2 and 3 did not differ substantially, consistent with the difference being the configuration at position 21. These two diastereomer configurations are most likely formed earlier in iridoid biosynthesis by the spontaneous ring opening and closing of 7-deoxyloganetic acid that forms C2-hydroxy epimers which are then

glycosylated (Figure 4.5). The presence of multiple isomers for most of the later biosynthetic intermediates suggests that *C. acuminata* 7-deoxyloganetic acid glucosyltransferase is either a promiscuous enzyme that can accommodate both R- and S-isomers of 7-deoxyloganetic acid or that multiple stereospecific *C. acuminata* 7-deoxyloganetic acid glycosyltransferases exist. Under the 52-min chromatographic conditions mentioned in chapter 3, the first iridoids, loganic acid (2) and secologanic acid (3), appeared as single chromatographic peaks, however, two isomers were resolved using a different chromatography column and gradient (Sadre, 2016). Thus, the isomers observed throughout the pathway are likely glucosidic isomers that have their origin early in iridoid synthesis.

Table 4.1. NMR chemical shifts and coupling constants for strictosidinic acid isomers isolated from *Camptotheca acuminata* leaf tissue as measured from J-resolved ^{1}H spectra and ^{1}H-^{13}C HSQC spectra. Notable differences in ^{1}H chemical shifts and $^1J_{C-H}$ coupling constants were observed for position 21 (highlighted in bold in the table).

Carbon number and group	Isomer 2		Isomer 3	
	^{1}H shift (ppm) and $^3J_{H-H}$ coupling constants (Hz)	^{13}C shift (ppm) and $^1J_{C-H}$ coupling constants (Hz)	^{1}H shift (ppm) and $^3J_{H-H}$ coupling constants (Hz)	^{13}C shift (ppm) and $^1J_{C-H}$ coupling constants (Hz)
3 (CH)	4.45 (dd), ^3J = 3.4, 7.7 Hz	50.8, ^1J = 144 Hz	4.56 (dd), ^3J = 3.6, 7.9 Hz	53.1, ^1J = 143 Hz
5 (CH$_2$)	3.29 (m) 3.75 (m)	41.3, ^1J = 143, 145 Hz;	3.45 (m) 3.50 (m)	39.6, ^1J = 139, 141 Hz
6 (CH$_2$)	3.01 (m) 3.29 (m)	18.1, ^1J = 131, 133 Hz	2.99 (m) 3.04 (m)	18.2, ^1J = 129, 131 Hz
9 (CH)	7.56 (d), ^3J = 7.9 Hz	117.7, ^1J = 159 Hz	7.55 (d), ^3J = 7.9 Hz	117.6, ^1J = 158 Hz
10 (CH)	7.15, ^3J = 7.4 Hz	119.1, ^1J = 159 Hz	7.13, ^3J = 7.3 Hz	118.9, ^1J = 158 Hz
11 (CH)	7.23, ^3J = 7.6 Hz	121.9, ^1J = 159 Hz	7.23, ^3J = 7.6 Hz	121.8, ^1J = 158 Hz
12 (CH)	7.42 (d), ^3J = 8.2 Hz	110.7, ^1J = 159 Hz	7.45 (d), ^3J = 8.2 Hz	110.8, ^1J = 158 Hz
14 (CH$_2$)	2.15 (dd), ^3J = 4.6, 11.1 Hz; 2.40 (dd), ^3J = 5.9, 11.6 Hz	33.7, ^1J = 128, 129 Hz	1.98 (m) 2.46 (m)	34.1, ^1J = 130, 131 Hz
15 (CH)	3.00 (m)	32.6, ^1J = 137 Hz	2.96 (m)	32.9, ^1J = 135 Hz
17 (CH)	7.67 (s)	151.7, ^1J = 191 Hz	7.35 (s)	148.4, ^1J = 192 Hz
18 (CH$_2$)	5.32 (d), ^3J = 10.7 Hz; 5.43 (d), ^3J = 17.3 Hz	117.5, ^1J = 155, 158 Hz	5.39 (d), ^3J = 11.0 Hz; 5.47 (d), ^3J = 17.3 Hz	118.4, ^1J = 154, 161 Hz
19 (CH)	5.96 (dd), ^3J = 10.3, 17.5 Hz	134.8, ^1J = 156 Hz	6.19 (dd), ^3J = 10.2, 17.2 Hz	134.8, ^1J = 152 Hz
20 (CH)	2.70 (m)	44.2, ^1J = 134 Hz	2.74 (m)	44.9, ^1J = 132 Hz
21 (CH)	**5.93 (d), ^3J = 9.5 Hz**	**95.1, ^1J = 170 Hz**	**5.64 (d), ^3J = 8.2 Hz**	**95.5, ^1J = 178 Hz**
1′ (CH)	4.92 (d), ^3J = 8.0 Hz	98.9, ^1J = 163 Hz	4.85 (d), ^3J = 7.9 Hz	98.8, ^1J = 162 Hz
2′ (CH)	3.44 (m)	76.5, ^1J = 145 Hz	3.41 (m)	76.5, ^1J = 145 Hz
3′ (CH)	3.41 (m)	77.3, ^1J = 145 Hz	3.35 (m)	77.3, ^1J = 145 Hz
4′ (CH)	3.26 (m)	70.2, ^1J = 145 Hz	3.28 (m)	70.3, ^1J = 145 Hz
5′ (CH)	3.24 (m)	73.3, ^1J = 145 Hz	3.24 (m)	73.3, ^1J = 145 Hz
6′ (CH$_2$)	3.68 (dd), ^3J = 4.95, 12.2 Hz; 4.12 (dd), ^3J = 2.17, 12.1 Hz	61.6, ^1J = 141, 142 Hz	3.76 (dd), ^3J = 6.61, 12.3 Hz; 4.05 (dd), ^3J = 2.40, 12.6 Hz	61.5, ^1J = 131, 132 Hz
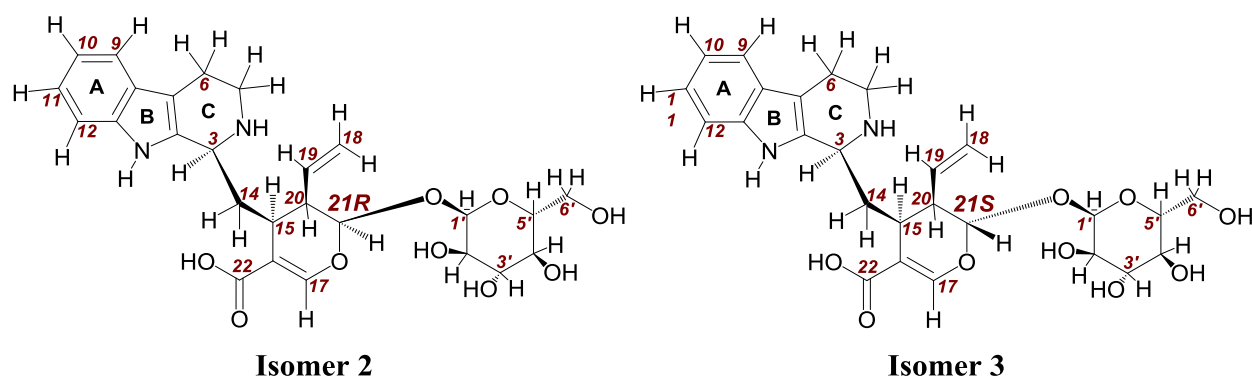
Figure 4.4. Structures of isomers 2 and 3 of strictosidinic acid determined using 1D and 2D NMR and MS analyses.
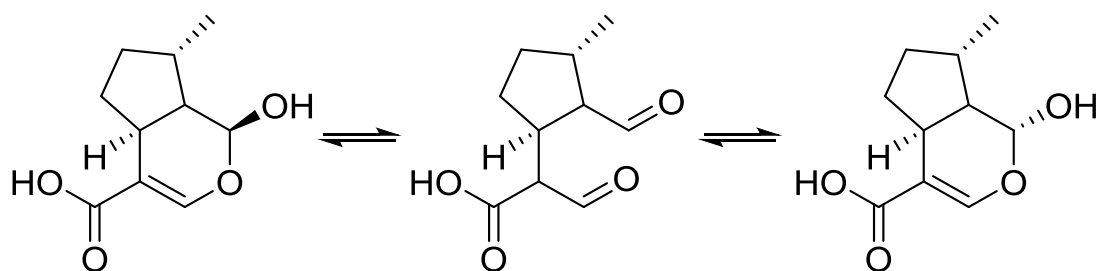
Figure 4.5. Formation of iridoid stereoisomers. Equilibration between open- and closed-ring conformation yielding stereoisomers at the C2 hydroxyl group is depicted for 7-deoxyloganetic acid.

4.4 Conclusions

The research performed in this chapter concludes that the presence of multiple isomeric MIA metabolites in UHPLC/MS metabolite profiles of *C. acuminata* tissues with indistinguishable MS/MS spectra (Table 3.1) suggested they differ in stereochemical configurations, but their chromatographic resolution indicates they are diastereomers. The most substantial differences were observed in the ^{1}H chemical shifts at position 21 (Table 4.1), which is the position of glycosylation.

The abundance of glucoside isomer mixtures for the majority of the camptothecin biosynthetic pathway, from loganic acid to deoxypumiloside, is both remarkable and unprecedented in MIA biosynthesis. This has profound implications from a mechanistic perspective because it suggests that some, if not all, of the enzymes from 7-deoxyloganetic acid to deoxypumiloside are able to accommodate multiple isomers or that multiple enzymes exist for each step that are specific for a single stereoisomer. Given that the metabolic transformations from strictosidinic acid to deoxypumiloside are remote from the site of glycosylation, the consistent presence of multiple intermediate isomers in this portion of the pathway suggests that a single enzyme could accommodate both stereoisomer substrates. However, the variation in the isomeric composition of strictosidinic acid (5) and post-strictosidinic acid compounds between different tissues in wild-type as well as the marked changes in the isomer composition of certain metabolites in some CYC1-RNAi tissues (Table 3.2) indicate that additional stereo-selective sinks or processes are active in *C. acuminata*.

To assess the nature of these isomeric compounds, the two most abundant isomers of strictosidinic acid were purified and subjected to an assortment of 1-D and 2-D NMR spectroscopic analyses, including MS analyses. The tactic of using ^{1}H-^{1}H correlation signals with one-bond ^{1}H-

[13]C correlation signals to decipher fragments of protonated carbons, and subsequent analyses through long-rang [1]H-^{13}C correlation is very useful for building the skeletal structures of complicated natural products. The structure elucidation of the two isomers of strictosidinic acid presented in this chapter is a great example of how 1-D and 2-D NMR analyses have been combine to make progress in the field of natural product research.

APPENDIX
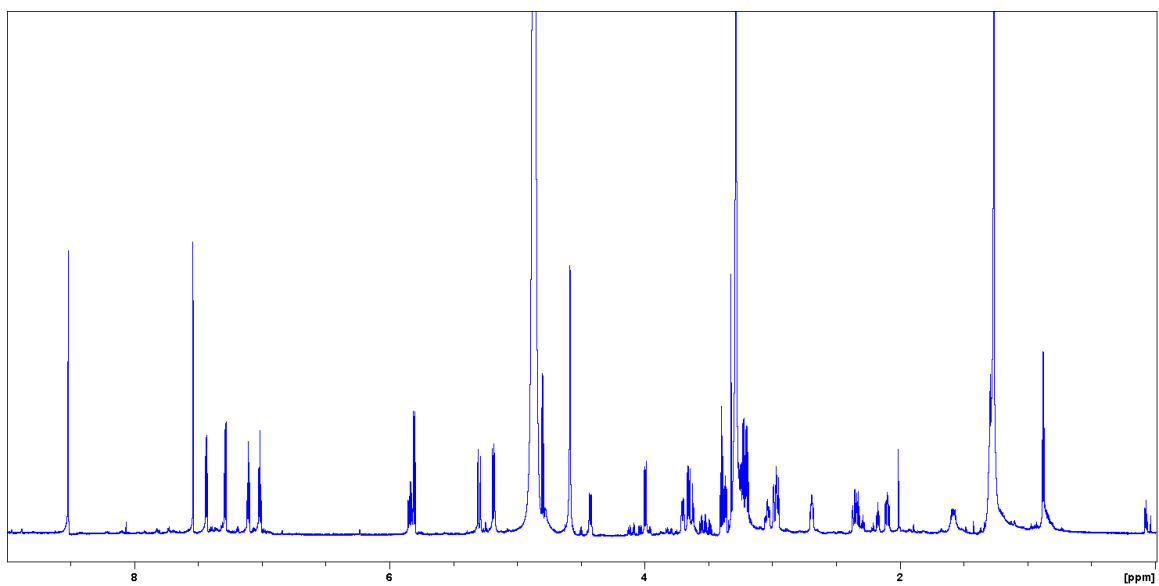
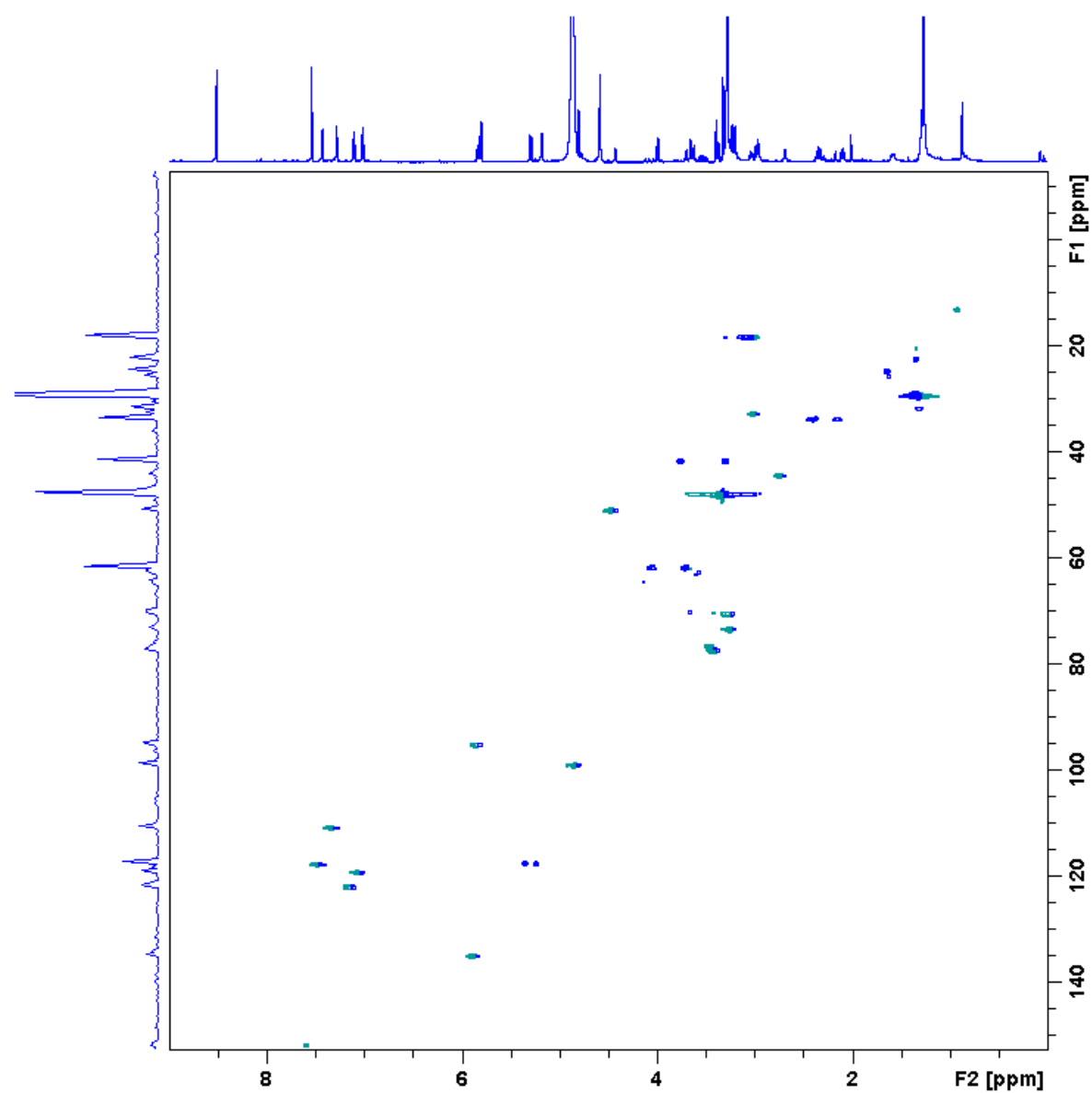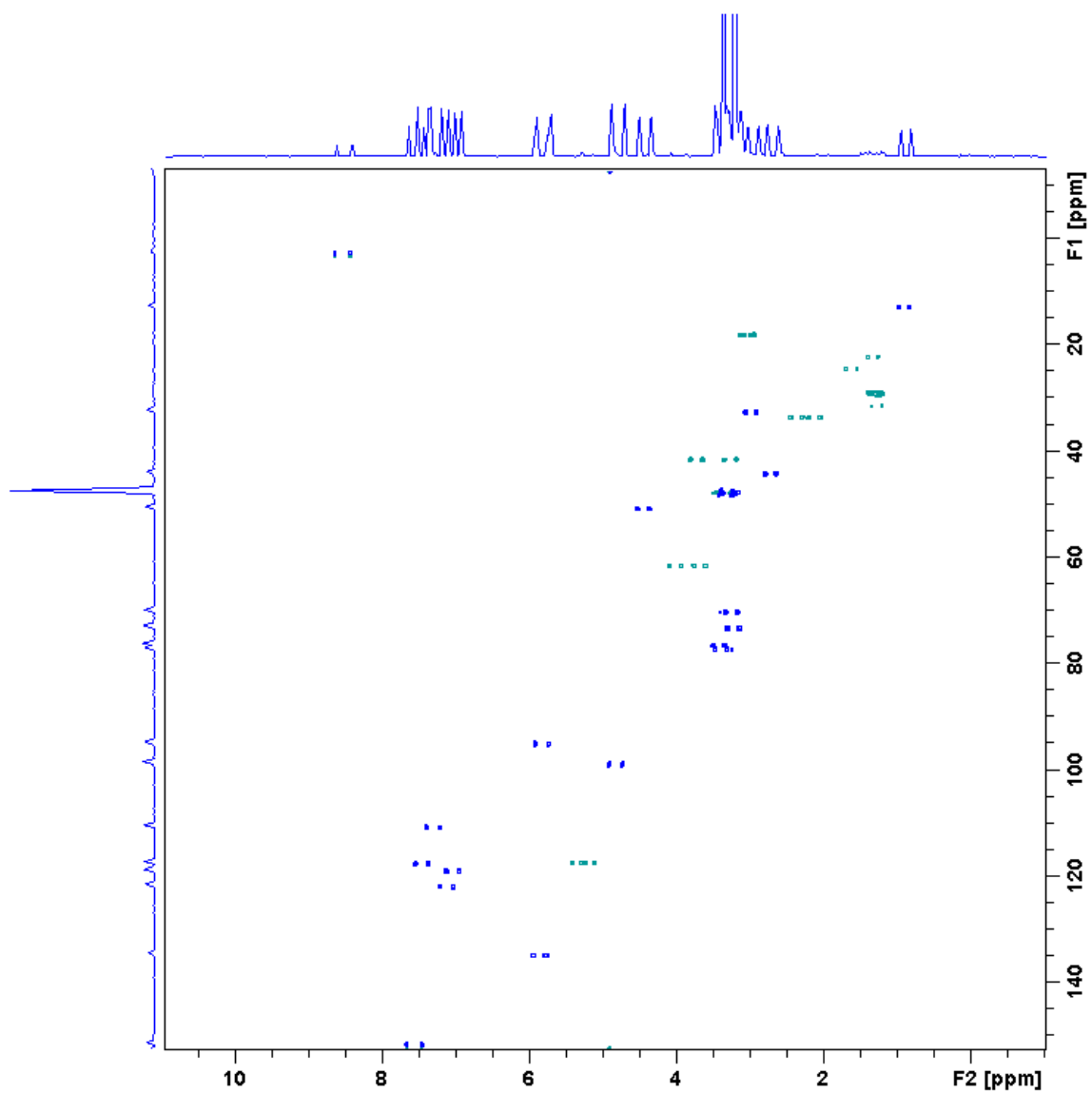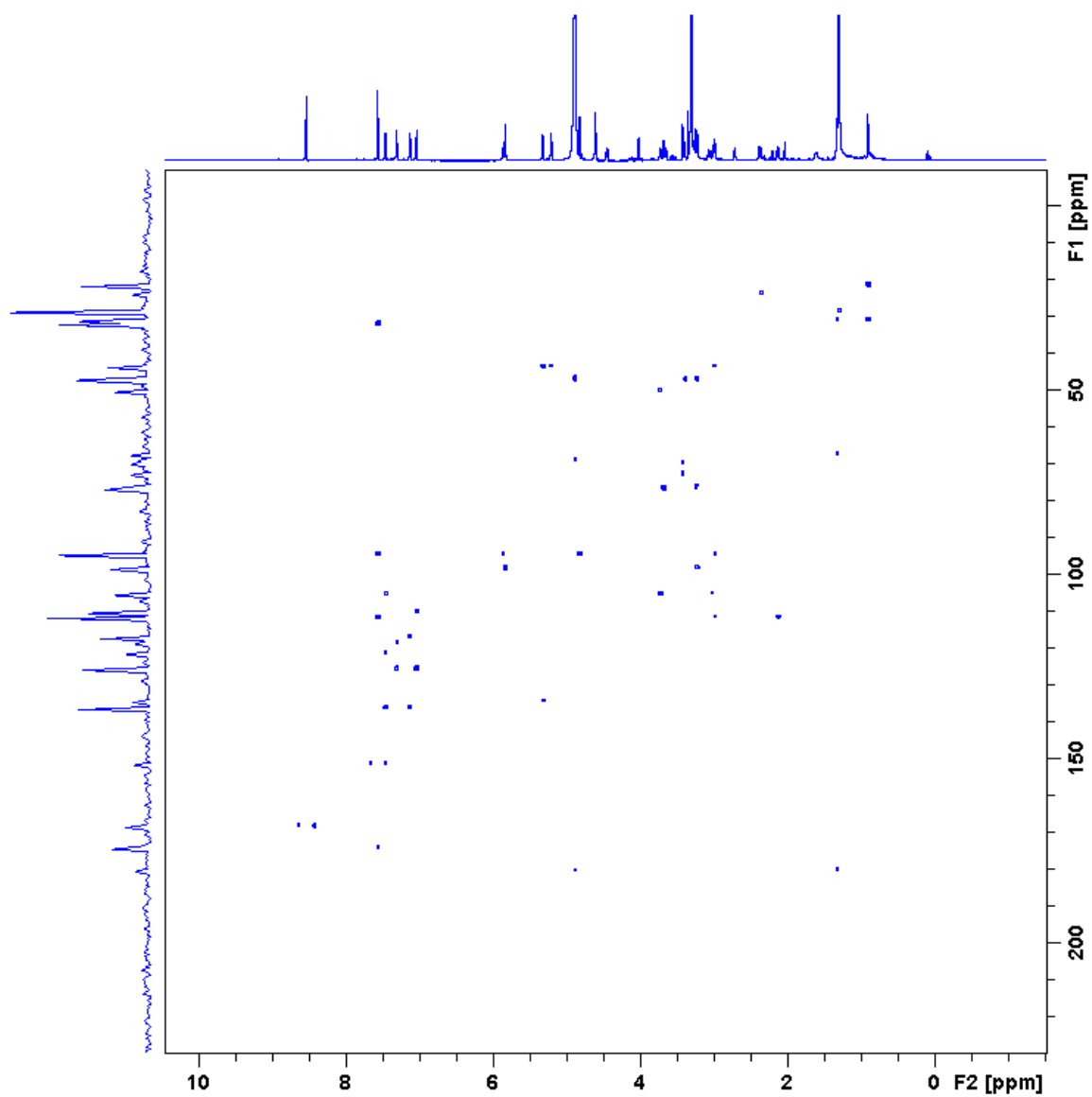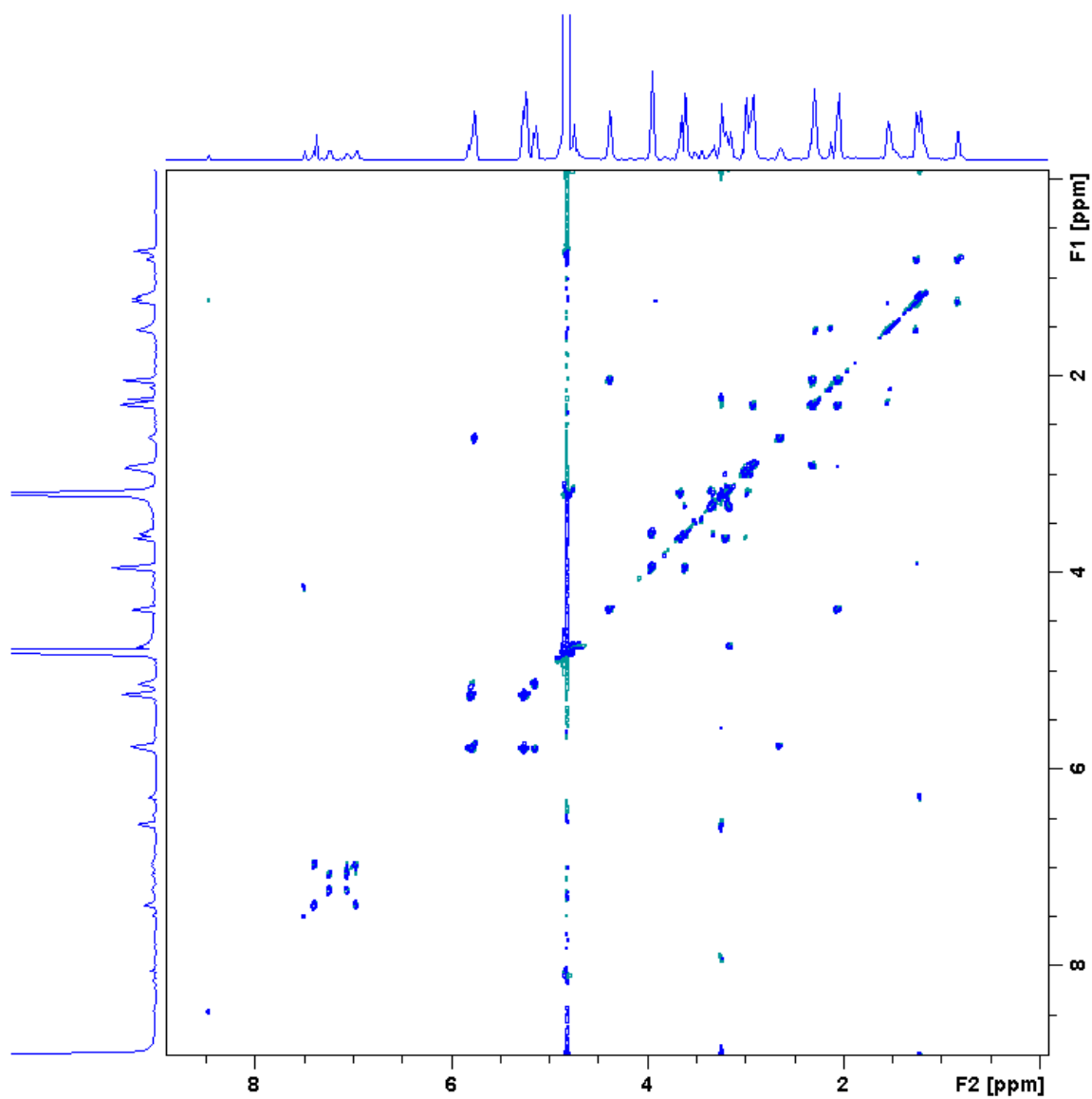**Figure A.1. 1H NMR spectrum of the isolated strictosidinic acid isomer 2**

Figure A.2. HSQC spectrum of the isolated strictosidinic acid isomer 2

Figure A.3. cHSQC spectrum of the isolated strictosidinic acid isomer 2

Figure A.4. HMBC spectrum of the isolated strictosidinic acid isomer 2

Figure A.5. COSY spectrum of the isolated strictosidinic acid isomer 2
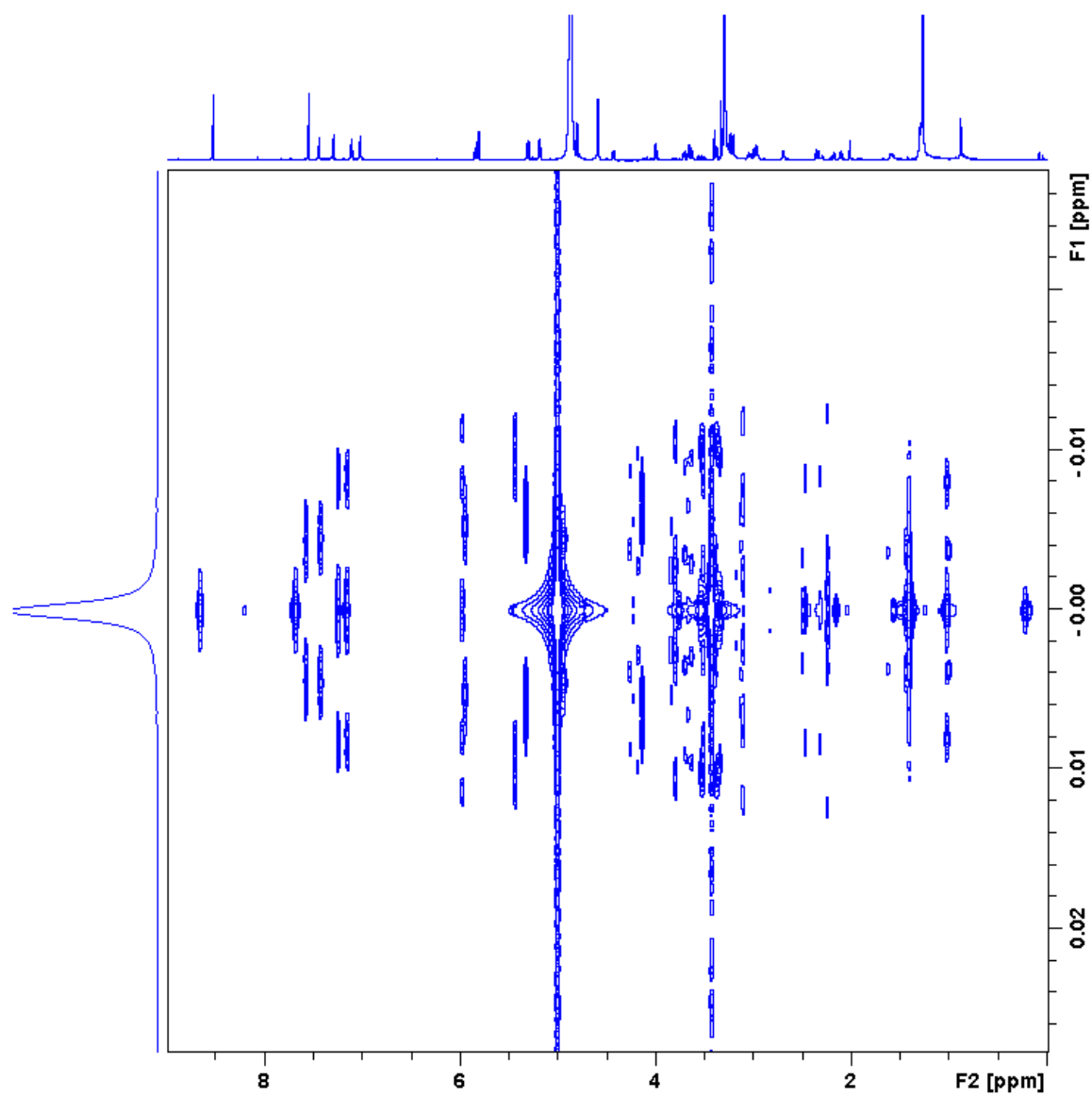
Figure A.6. *J*-resolved spectrum of the isolated strictosidinic acid isomer 2

Figure A.7. ^{1}H NMR spectrum of the isolated strictosidinic acid isomer 3

Figure A.8. HSQC spectrum of the isolated strictosidinic acid isomer 3

Figure A.9. cHSQC spectrum of the isolated strictosidinic acid isomer 3

Figure A.10. HMBC spectrum of the isolated strictosidinic acid isomer 3

Figure A.11. COSY spectrum of the isolated strictosidinic acid isomer 3
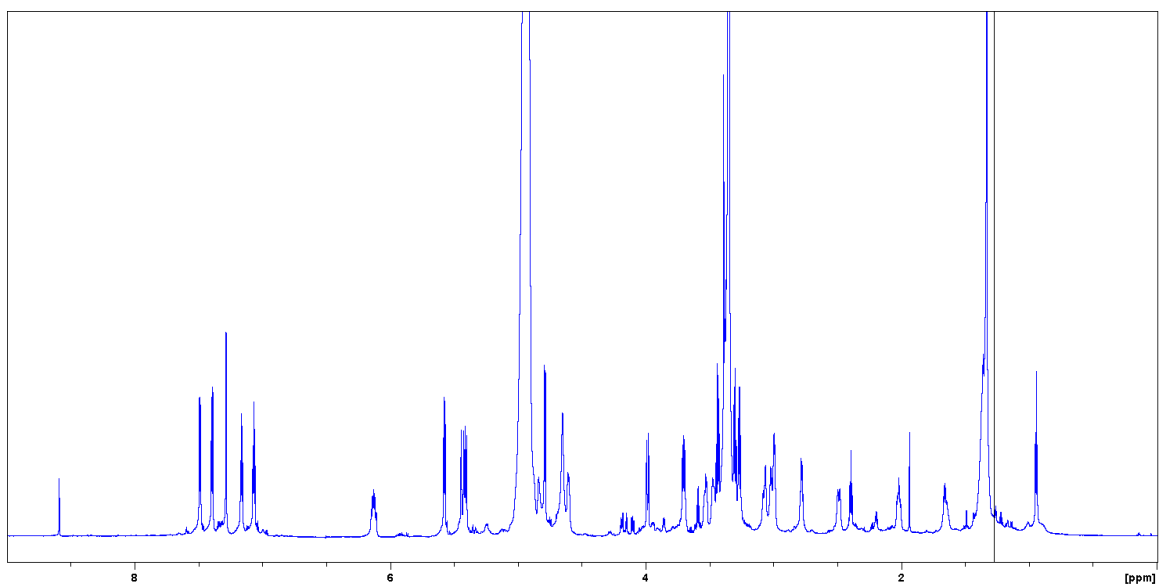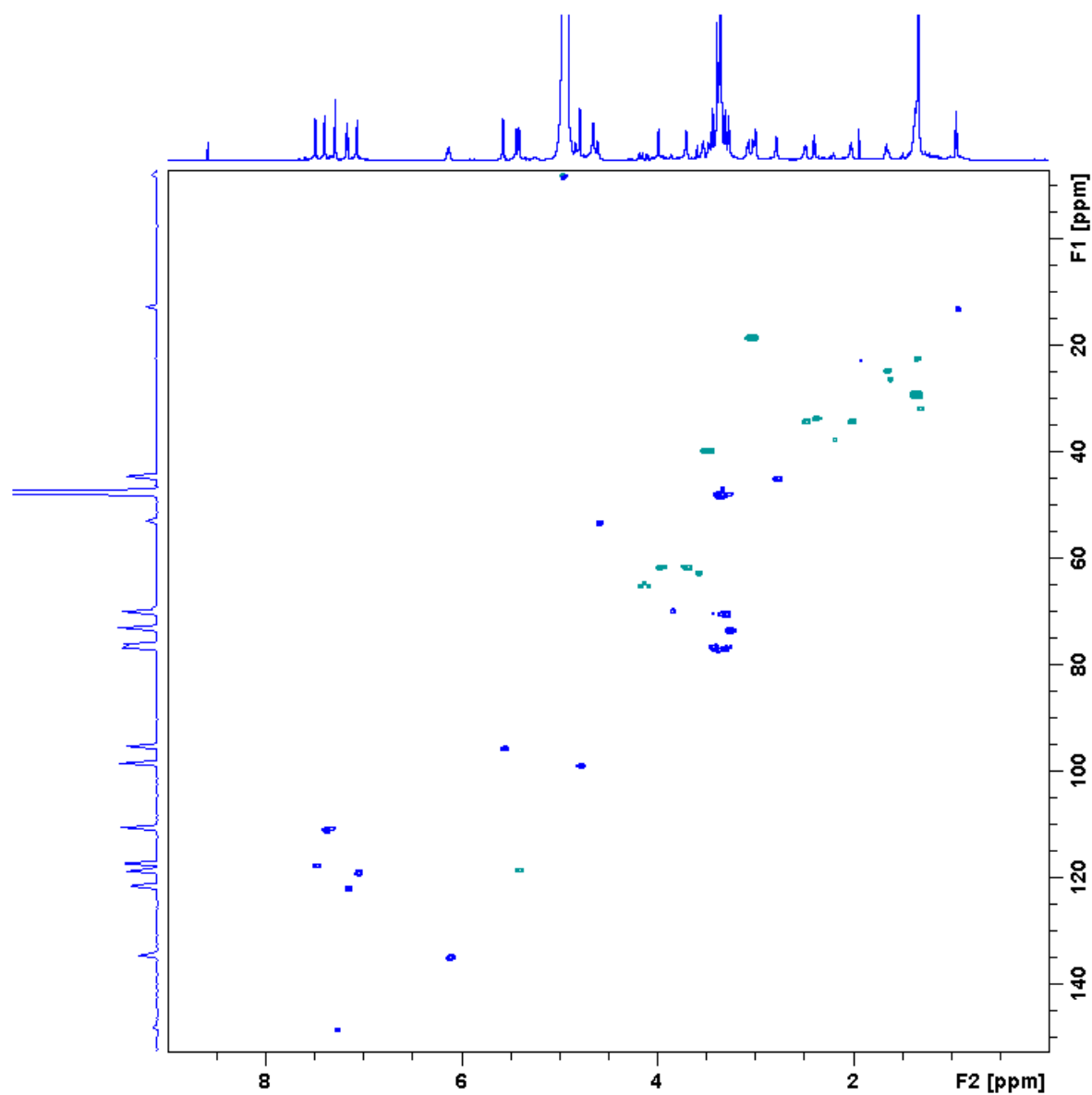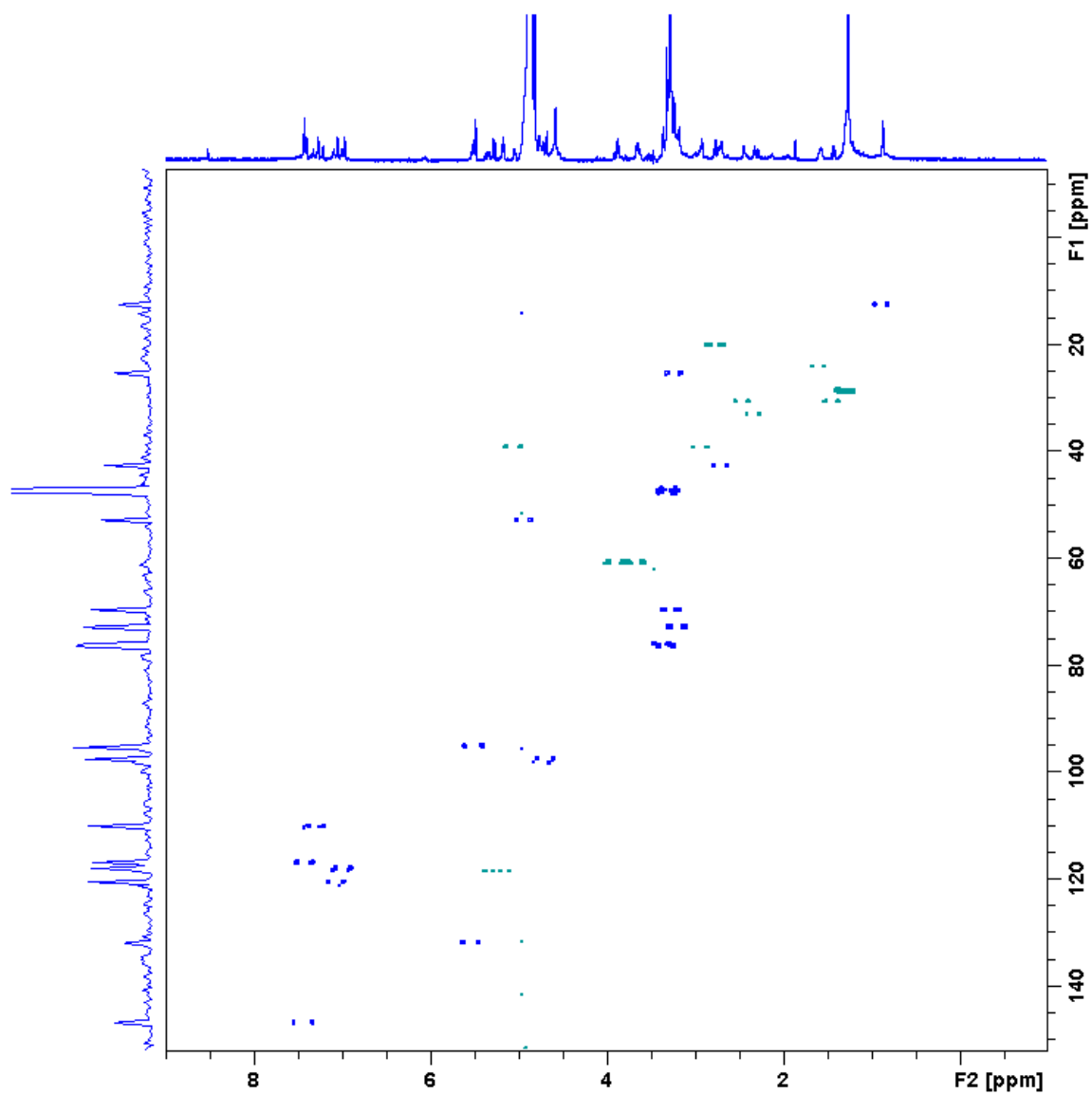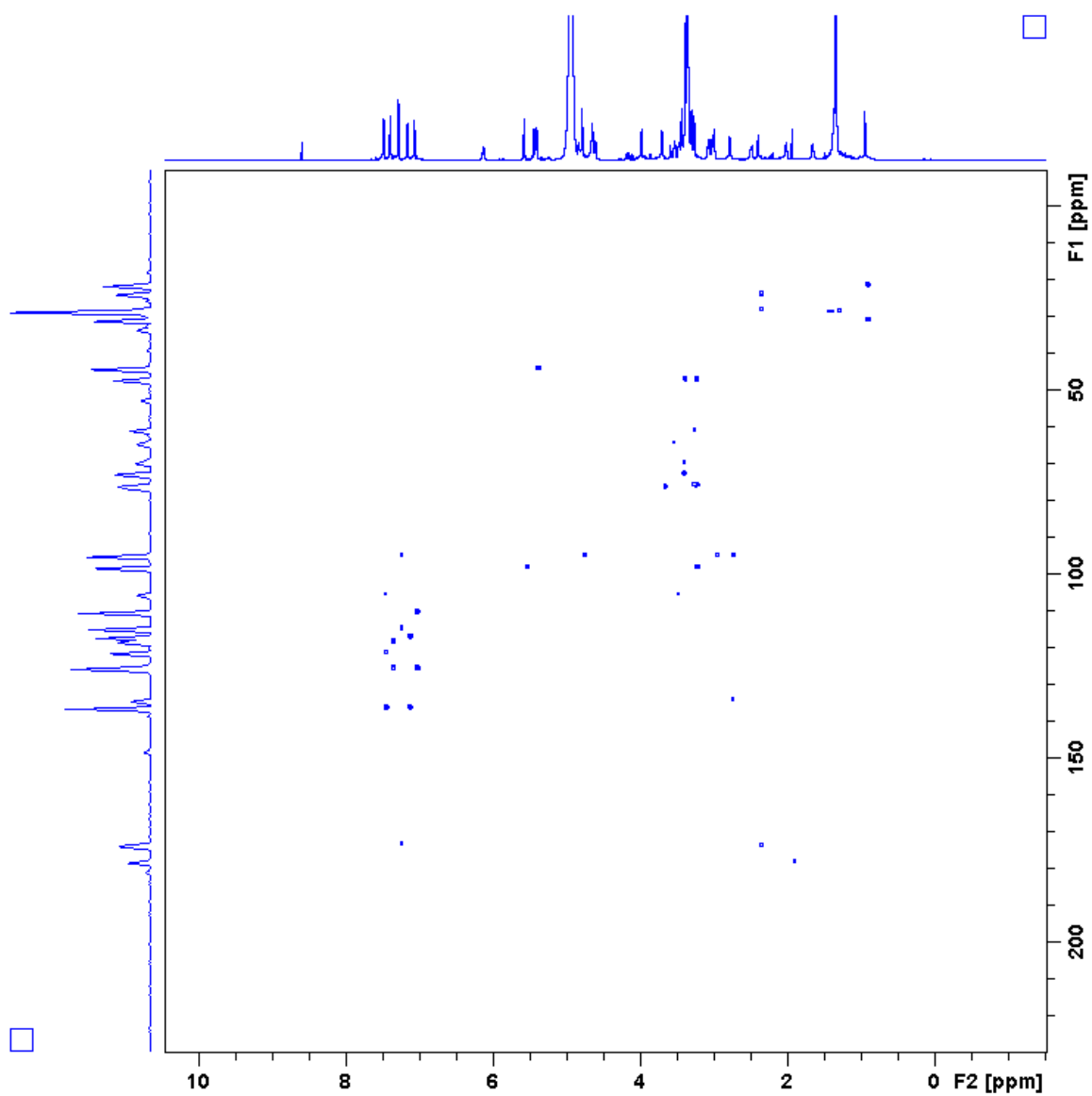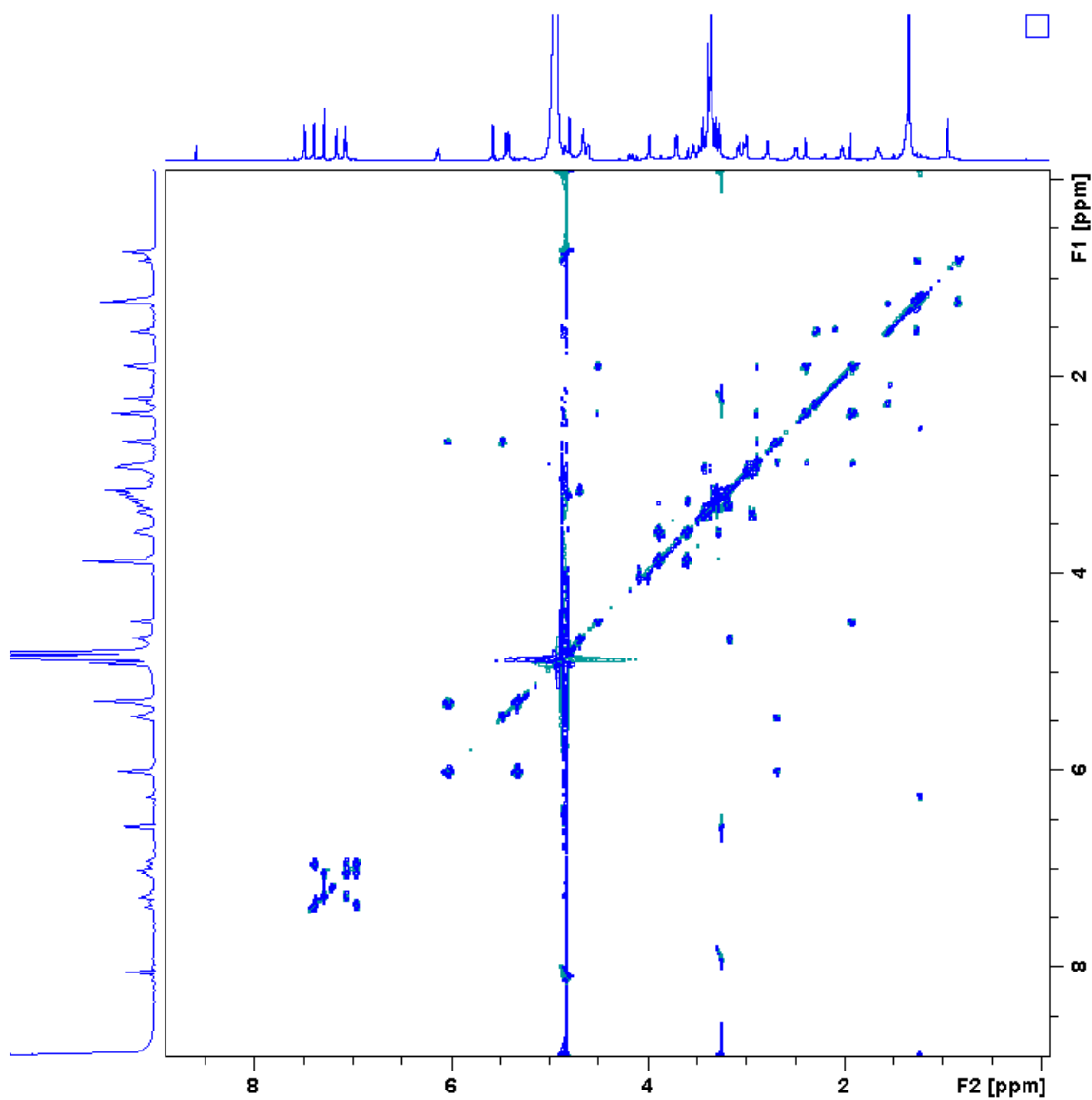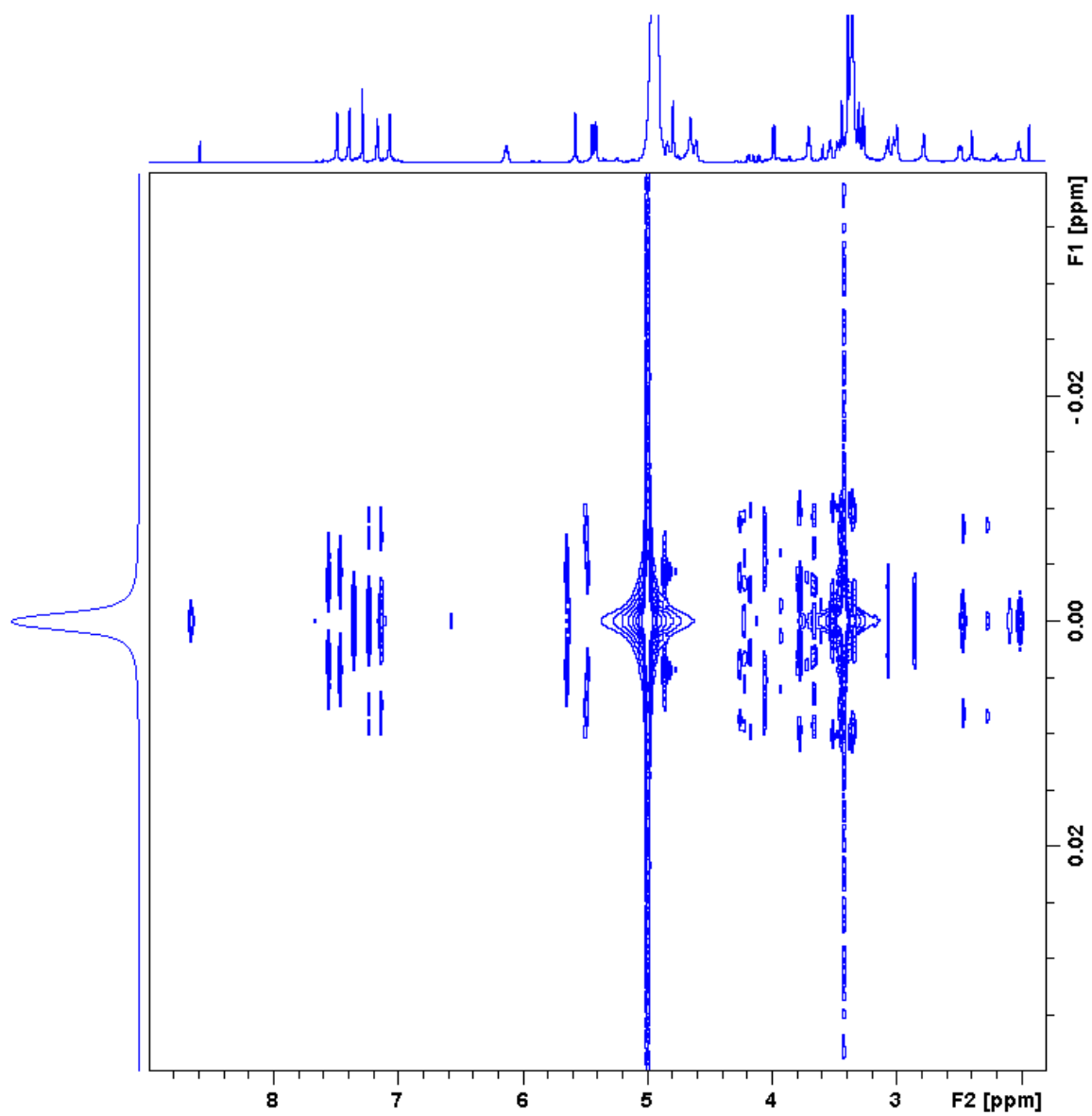
Figure A.12. *J*-resolved spectrum of the isolated strictosidinic acid isomer 3

Table A. 1 NMR Metadata

NMR Metadata	
Instrument Type	Bruker Avance-900 MHz NMR
NMR Experiment Type	1D, 2D
Field Frequency Lock	Deuterium
NMR Solvent	Methanol-d_4
Standard Concentration	None
Spectrometer Frequency	900 MHz
NMR Tube Size	Shigemi 5 mm tube
NMR Probe	Bruker TCI triple-resonance inverse detection cyroprobe
Shimming Method	topshim
Pulse Sequence	^{1}H, ^{13}C, HSQC, HMBC, COSY
Water Suppression	None
Pulse Width	^{1}H: 27 µs ^{13}C: 9.225 µs HSQC: 46.4 µs HMBC: 46.4 µs COSY: 61.8 µs
Receiver Gain	^{1}H: 14.3 ^{13}C: 41285 HSQC: 46341 HMBC: 26008 COSY: 80.6
Offset Frequency	^{1}H: 5557 Hz ^{13}C: 22629 Hz HSQC: 4496Hz, 15829 Hz HMBC: 4047 Hz, 24888 Hz COSY: 4047 Hz, 4047 Hz
Chemical Shift Reference Compound	TMS
Temperature	298K
Number of Scans	^{1}H: 128 ^{13}C: 3000 HSQC: 16 HMBC: 40 COSY: 16
Dummy Scans	^{1}H: 2 ^{13}C: 8 HSQC: 16 HMBC: 32 COSY: 24

Table A.1. (cont'd)

	Isomer A (1st eluting)
Aqcuisition Time	^{1}H: 4 s ^{13}C: 0.6 s HSQC: 0.2 s HMBC: 0.2 s COSY: 0.13 s
Spectral Width	^{1}H: 18518 Hz ^{13}C: 54200 Hz
Number of Data Points (TD) Acquired	^{1}H: 148K ^{13}C: 65K HSQC: 4308, 400 HMBC: 4096, 480 COSY: 2048, 300
Real Data Points (SI)	^{1}H: 262K ^{13}C: 131K HSQC: 8192, 1024 HMBC: 4096, 2048 COSY: 4096, 4096
Line Broadening	^{1}H: 0 Hz ^{13}C: 2 Hz
Baseline Correction Method	Polynomial
Chemical Shift Reference Standard	None

REFERENCES

REFERENCES

1.   Wall, M.E., et al., *Plant Antitumor Agents. I. The Isolation and Structure of Camptothecin, a Novel Alkaloidal Leukemia and Tumor Inhibitor from Camptotheca acuminata1,2.* Journal of the American Chemical Society, 1966. **88**(16): p. 3888-3890.

2.   Zhang, Z., et al., *New camptothecin and ellagic acid analogues from the root bark of Camptotheca acuminata.* Planta Med, 2004. **70**(12): p. 1216-21.

3.   Wang, P., et al., *New indole glucosides as biosynthetic intermediates of camptothecin from the fruits of Camptotheca acuminata.* Fitoterapia, 2015. **103**: p. 1-8.

4.   Shen, Y. and M.L. Lee, *General Equation for Peak Capacity in Column Chromatography.* Analytical Chemistry, 1998. **70**(18): p. 3853-3856.

5.   Bell, D.B. and Jones, A.D., *Solute attributes and molecular interactions contributing to "U-shape" retention on a fluorinated high-performance liquid chromatography stationary phase*, Journal of Chromatography A, 2005. **1073**(1): p. 99-109.

Chapter 5: Concluding remarks

The tremendous diversity of MIAs found in nature poses several fundamentals questions about the evolution of the pathway, especially in light of the data presented in this dissertation for *C. acuminata*. Synthesis of loganic acid, an early pathway precursor, has been documented for numerous other plant species from unrelated plant orders and families [1-11] and it appears to represent a trait that arose relatively early in plant evolution. The highly conserved proteins for early seco-iridoid biosynthesis in the unrelated species *C. acuminata* and *C. roseus* are consistent with this interpretation. In contrast, the differences in seco-iridoid biosynthesis that yield exclusively strictosidinic acid in *C. acuminata* (Cornales) and strictosidine in the Gentianales (*C. roseus*, *R. serpentina* and *O. pumila*) indicate that post-loganic acid steps likely evolved later. In this context, it must be mentioned that although *C. roseus*, *R. serpentina* and *O. pumila* all produce exclusively strictosidine, several other members of the Gentianales, most notably other *Ophiorrhiza* species, are able to produce strictosidinic acid [12-18]. As additional biochemical and genetic information is not available for these strictosidinic acid producing Gentianales, a comparison of them with *C. roseus* and *C. acuminata* is not possible but would help clarify if strictosidinic acid arose independently (i.e., by convergent evolution) or whether the lack of secologanic acid methylation in *C. acuminata* is an ancestral trait.

The first study in this dissertation was able to bring attention to differences in metabolite levels in the plant depending on its developmental stage. The question that remains to be answered is why younger plants are able to still produce camptothecin without accumulating significant levels of intermediates. Future researchers may utilize these results to compare the transcriptome data which may lead to enzymes that are upregulated at certain stages in the life of the plant, allowing for manipulation of those steps to reduce levels of intermediates and increase levels of final product.

In the second and third chapters of this dissertation, metabolite profiling has shown multiple isomeric forms of intermediates throughout the pathway and the differences in the ratios of isomer abundance found in plant tissues raise questions on whether there may be multiple enzymes working together to perform the transformation of a certain intermediate or single enzyme that is able to accept multiple isomers of the same intermediate. Furthermore, it is interesting to note that the end of the pathway leads to camptothecin which does not show isomerism. It is currently unclear why the plant possesses a multi-lane highway approach to making a single end compound, or if the downstream products from one of the highway lanes do not feed into the final natural product.

The third chapter in this dissertation showed that the main difference in the isomers of strictosidinic acid lie in the glucosidic linkage. However, it is unclear whether this is the case for the isomeric forms of all other intermediates leading up to camptothecin. Strictosidinic acid was chosen for the analysis as described in Chapter 3 because it is the most abundant in leaf tissues, and due to the abundance of the amount leaf tissue available for extraction experiments. Therefore, in order to study the isomers of the other intermediates which are approximately 25-50 fold more dilute in any part of the plant, an impractical amount of raw tissue would be necessary. However, the research presented in Chapter 2 can be used to produce C-13 labeled isomeric intermediates. These labeled intermediates would mitigate the need for milligram quantities of purified compounds to be used for structure elucidation via NMR spectroscopic methods.

REFERENCES

REFERENCES

1. Skaltsounis, A.L., et al., *Iridoids from Scaevola racemigera1*. Planta Med, 1989. **55**(2): p. 191-2.

2. Graikou, K., et al., *Cantleyoside-dimethyl-acetal and other iridoid glucosides from Pterocephalus perennis--antimicrobial activities.* Z Naturforsch C, 2002. **57**(1-2): p. 95-9.

3. Muller, A.A. and M. Weigend, *Iridoids from Loasa acerifolia.* Phytochemistry, 1998. **49**(1): p. 131-5.

4. Rastrelli, L., et al., *Iridoids from Lippia graveolens.* Phytochemistry, 1998. **49**(6): p. 1829-1832.

5. Fan, G., et al., *Metabolic discrimination of Swertia mussotii and Swertia chirayita known as "Zangyinchen" in traditional Tibetan medicine by (1)H NMR-based metabolomics.* J Pharm Biomed Anal, 2014. **98**: p. 364-70.

6. Han, Q.B., et al., *A simple method to identify the unprocessed Strychnos seeds used in herbal medicinal products.* Planta Med, 2008. **74**(4): p. 458-63.

7. Serrilli, A.M., et al., *Iridoidic pattern in endemic Sardinian plants: the case of Galium species.* Nat Prod Res, 2008. **22**(7): p. 618-22.

8. Aberham, A., et al., *Analysis of iridoids, secoiridoids and xanthones in Centaurium erythraea, Frasera caroliniensis and Gentiana lutea using LC-MS and RP-HPLC.* J Pharm Biomed Anal, 2011. **54**(3): p. 517-25.

9. Zhang, T., et al., *Two novel secoiridoid glucosides from Tripterospermum chinense.* J Asian Nat Prod Res, 2012. **14**(12): p. 1097-102.

10. Asano, T., et al., *Suppression of camptothecin biosynthetic genes results in metabolic modification of secondary products in hairy roots of Ophiorrhiza pumila.* Phytochemistry, 2013. **91**: p. 128-39.

11. Zhou, Z., et al., *A new iridoid glycoside and potential MRB inhibitory activity of isolated compounds from the rhizomes of Cyperus rotundus L.* Nat Prod Res, 2013. **27**(19): p. 1732-6.

12. Arbain, D., et al., *(R)-Magnocurarine from Evodia cf. trichotoma.* Planta Med, 1993. **59**(3): p. 290.

13. Hamzah, A.S., N.H. Lajis, and M.V. Sargent, *Kaempferitrin from the leaves of Hedyotis verticillata and its biological activity.* Planta Med, 1994. **60**(4): p. 388-9.

14.   Reanmongkol, W., et al., *Analgesic and Antipyretic Activities of n-Butanol Alkaloids Extracted from the Stem Bark Hunteria Zeylanica and its Major Constituent, Strictosidinic Acid, in Mice.* Pharm Biol, 2000. **38**(1): p. 68-73.

15.   Cardoso, C.L., et al., *Indole glucoalkaloids from Chimarrhis turbinata and their evaluation as antioxidant agents and acetylcholinesterase inhibitors.* J Nat Prod, 2004. **67**(11): p. 1882-5.

16.   Atsuko, T., et al., *Non-invasive assessment of arterial stiffness in appropriate-for-gestational age newborn infants.* Early Hum Dev, 2008. **84**(6): p. 403-8.

17.   Olusegun, A.F., O.C. Ehis, and O. Richard, *Proportion of Urinary Schistosomiasis among HIV-Infected Subjects in Benin City, Nigeria.* Oman Med J, 2011. **26**(3): p. 175-7.

18.   Farias, F.M., et al., *Strictosidinic acid, isolated from Psychotria myriantha Mull. Arg. (Rubiaceae), decreases serotonin levels in rat hippocampus.* Fitoterapia, 2012. **83**(6): p. 1138-43.