BIOCHEMICAL INVESTIGATION OF EARLY ENZYME PATHWAYS IN

HAPALINDOLE-TYPE ALKALOID BIOSYNTHESIS

by

Kuljira Ittiamornkul

B.Eng. Chemical Engineering, Thammasat University, 2011

Submitted to the Graduate Faculty of

The Dietrich School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

THE DIETRICH SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Kuljira Ittiamornkul

It was defended on

November 21, 2014

and approved by

Professor Steve Weber, Department of Chemistry

Professor Kabirul Islam, Department of Chemistry

Committee chair: Professor Xinyu Liu, Department of Chemistry

BIOCHEMICAL INVESTIGATION OF EARLY ENZYME PATHWAYS IN

HAPALINDOLE-TYPE ALKALOID BIOSYNTHESIS

Kuljira Ittiamornkul, M.S.

University of Pittsburgh, 2014

Stigonematalean cyanobacteria are producers of countless secondary metabolites with diverse structures that show a variety of bioactivities. Hapalindole-type indole alkaloids produced by *Fischerella* sp., and *Hapalosiphon* sp. exhibit pharmaceutical potentials. These hapalindole-type indole alkaloids share common molecular features including an indole core, isonitrile group, which is attached with a monoterpene unit to form a tri- or tetra cylic structure. The relative stereochemical diversity across C10-C12 is conserved in the same stigonematalean species, but differs between species. The origin of sterochemical diversity was intriguing but the biosynthesis knowledge at the molecular level was absent. The initial biosynthetic hypothesis was proposed by Moore, which suggested that the biosynthesis involved a chloronium ion or proton-catalyzed polycyclization of 3-(2'-isocyanoethanyl) indole **1** (*E* or *Z*) with β-ocimene **2** (*E* or *Z)* to provide tricyclic hapalindole core intermediate, which can be oxidized to other hapalindole-type molecules.

Our group started an effort to identify the biosynthetic gene clusters of ambiguines from *F. ambigua* UTEX 1903 and welwitindolinones from *H. welwitschii* UTEX B1830. This effort revealed common biosynthetic precursors in both pathways, including GPP and *(Z)*-**1** that contradicted the early hypothesis by Moore. It also resulted in the identification of *(Z)*-**1** biosynthetic genes that encode three isonitrile synthases I1, I2 and I3. This discovery contrasted the early study by Brady who has shown that IsnA/B are two enzymes responsible for

biosynthesis of *(E)*-**1**, where I1/I2 are homologues of IsnA and I3 is homologue of IsnB. In addition to this discovery, it was found that I1-I3 was required for *(Z)*-**1** production *in vivo*, whereas I1/I3 were suffice *in vitro*. This finding regarding the possible redundancy of I2 led to studies, including I2 truncation, complementation and *in vitro* reconstitution, to investigate the role of I2. These studies collectively demonstrated the possible role of I2 is an isomerase to convert ribose-5-phosphate to ribulose-5-phosphate.

Furthermore, promiscuity of I1-I3 were investigated by using various halogenated tryptophans as substrates by *in vitro* and *in vivo*. It can be deduced that the promiscuity of I1-I3 was limited by the size of substituted halogens but not by the position of the halogens on the indole ring.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS

BME   β-mercaptoethanol

DMAPP   dimethylallyl pyrophosphate

DNA   deoxyribonucleic acid

DTT   dithiothreitol

EDTA   ethylenediaminetetraacetic acid

HEPES   2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid

HPLC   high-pressure liquid chromatography

HRMS   high resolution mass spectrometry

HSQC   heteronuclear single quantum coherence

GPP   geranyl pyrophosphate

IPP   isopentenyl pyrophosphate

IPTG   isopropyl β-D-1-thiogalactopyranoside

LB   Luria Broth

NMR   nuclear magnetic resonance

PCR   polymerase chain reaction

PEG   polyethylene glycol

SC   *Saccharomyces cerevisiae*

SDC   sodium dodecanoyl sarcosine

SDS   sodium dodecyl sulfate

TE   tris-EDTA

UV   ultraviolet

YPD   yeast-extract-peptone-dextrose

1.0   INTRODUCTION

Cyanobacteria (blue-green algae) are photosynthetic prokaryotes that exist in diverse habitats worldwide including aquatic, marine, terrestrial environments and in more extreme environments ranging from hot springs to the Arctic [1]. The diversity of cyanobacteria arises from structural and functional characteristics of cell morphology and differences in metabolic strategies, motility, cell division, amongst others [2]. Traditionally, morphological variations have been used to divide cyanobacteria into five subsections [3]. The morphologically most complex one is stigonematalean or subsection V cyanobacteria, which are filamentous, differentiate heterocysts (specialized N_2-fixing cells) and exhibit true-branching. The morphological variety places them amongst the most sophisticated of prokaryotes for which reason they were believed to be ancestors of all eukaryotes [3-5].

Branched filamentous stigonematalean cyanobacteria are known to be producers of countless secondary metabolites with diverse structures that show a variety of bioactivities [6] such as insecticidal [7], antibacterial [8], antifungal [9], and anticancer [10]. Of all species of branched filamentous stigonematalean cyanobacteria, *Fischerella* sp., and *Hapalosiphon* sp., are two core prolific producers of bioactive small molecular natural products, including isonitrile-containing hapalindole-type terpenoid indole alkaloids [11], polyketide/nonribosomal peptide hybrids hapalosins [12], fischerellins [13] as well as halogenated polyphenol ambigols [14].

Of all these bioactive natural products produced by stigonematalean cyanobacteria, hapalindole-type indole alkaloids or isonitrile containing indole alkaloids such as hapalindoles, ambiguine isonitriles, fischerindoles and welwitindolinones, have been explored for their pharmaceutical potentials [15-16]. The pioneer work by Moore and co-workers in 1984 resulted in the isolation and characterization of hapalindoles A and B from *H. fontinalis* sp. The compounds showed antifungal and antimycotic activities [17]. In 1994, Smith *et al.* demonstrated that *N*-methylwelwitindolinone C isothiocyanate, *N*-methylwelwitindolinone C nitrile and welwitindolinone C isothiocyanate isolated from *H. welwitschii* to have differed levels of anticancer drug resistance attenuation [10]. In 2008, Mo *et al.* isolated five isonitrile-containing ambiguines K-O from *F. ambigua* and found that they exhibited various levels of antibacterial activities against *Mycobacterium tuberculosis* and *Bacilus anthracis* [8]. Since the first report by Moore in 1984, more than 70 related hapalindole-type natural products have been identified from the stigonematalean cyanobacteria [11-12].

1.1 STRUCTURAL DIVERSITIES AND EXISTING BIOSYNTHETIC PROPOSAL OF HAPALINDOLE-TYPE INDOLE ALKALOIDS

The family of hapalindole-type indole alkaloids shares a few common structural characteristics such as an indole core and isonitrile functional group (Figure 1). The indole and the isonitrile groups are connected by C10-C11, which is then attached with a monoterpene unit, forming a tricyclic structure with cyclohexane unit connected to C-3 position of the indole [12]. Exomethylene carbon C-16 fuses with the indole core at C-4, providing an additional ring to form tetracyclic hapalindoles. On the other hand, the fusion of C-16 with the indole core at C-2

gives fischerindoles which are proposed precursors to welwitindolinones. Pentacyclic ambiguines arise from the decoration of tetracyclic hapalindoles with a *tert*-prenyl group at C-2 of the indole, followed by additional oxidative diversifications [11]. Despite the structural diversities among the family of hapalindole-type molecules, the relative stereochemical diversity across C-10, C-11 and C-12 is well conserved in the same stigonematalean species. The relative stereochemical diversity across C10-C11 and C11-C12 of all hapalindole-type indole alkaloids produced by *F. ambigua* UTEX 1903 are *syn-syn*, whereas those produced from *H. welwitschii* UTEX B1830 are *syn-anti*. Lastly, those produced from *F. musicola* UTEX LB 1829 are *anti-syn* and *anti-anti* (Table 1.).
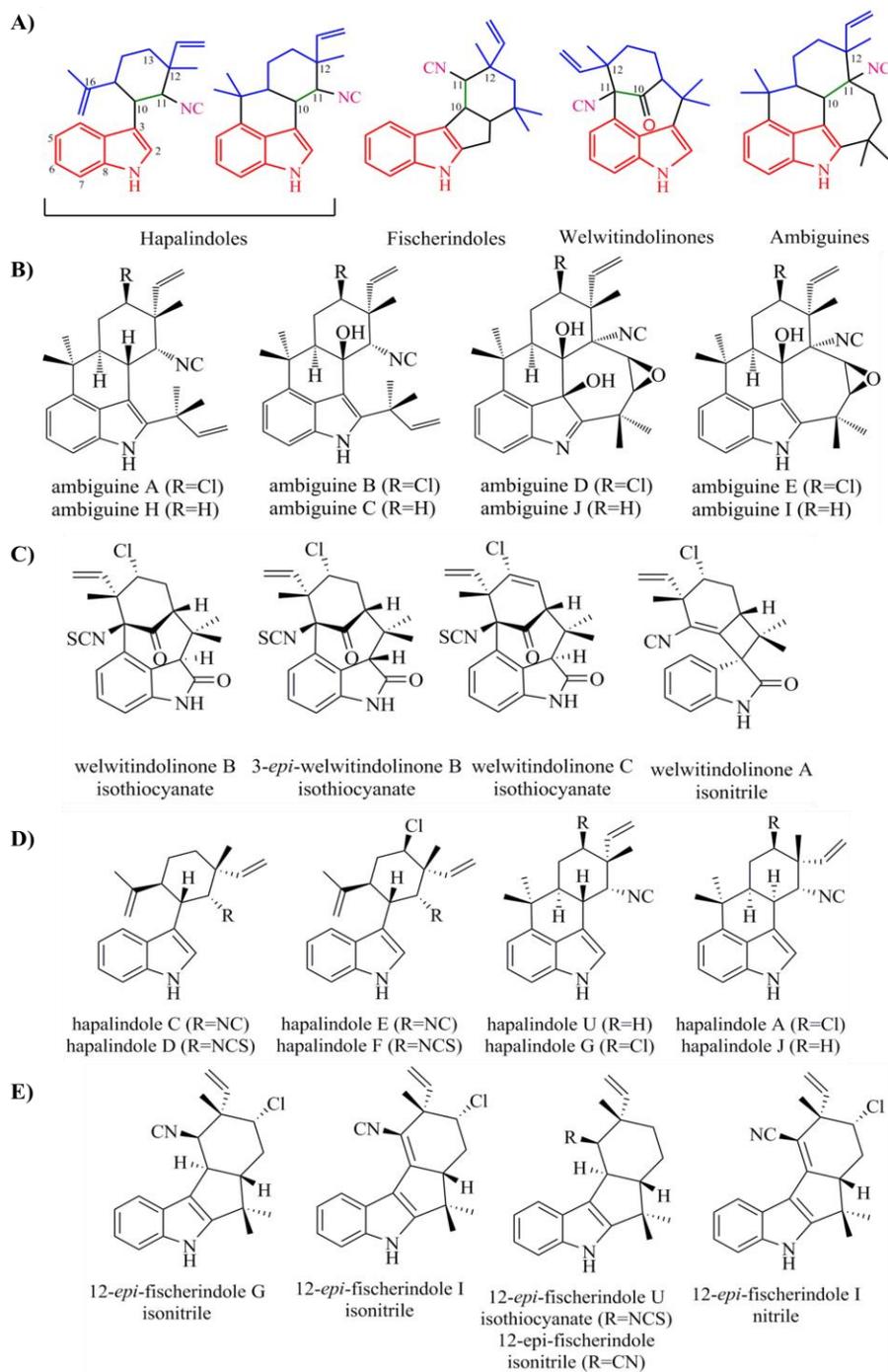
Figure 1. The structure diversity of hapalindole-type terpenoid indole alkaloids A) Typical monoterpenoid structural elements in hapalindole-type natural products B) structure of ambiguines produced by *F. ambigua* UTEX1903 C) structures of welwitindolinones produced by *H. welwitschii* UTEX B1830 D) structure of hapalindoles C, D, E, F produced by *H. welwitschii* UTEX B1830, and hapalindoles G, U, A, J produced by *F. ambigua* UTEX1903, E) structures of fischerindoles G, U, I produced by *H. welwitschii* UTEX B1830

Hapalindole G from
F. ambigua UTEX 1903

12-*epi*-Fischerindole G
from *H.*
welwitschii UTEX B1830

Hapalindole A from
F. musicola UTEX LB 1829

Fischerindole L
from
F. musicola UTEX LB1829

● C10
● C11
● C12
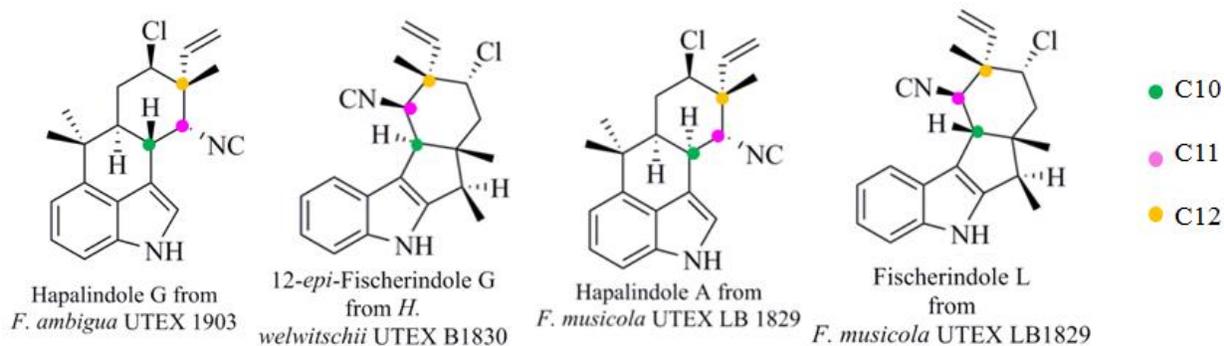
	C10 (indole) C11 (isonitrile)	C11 (isonitrile) C12 (vinyl)
UTEX 1903	*syn*	*syn*
UTEX B1830	*syn*	*anti*
UTEX LB1829	*anti*	*syn*
UTEX LB1829 (fischerindole L)	*anti*	*anti*

Table 1. Relative stereochemistry across C10, C11 and C12 showing the conserved stereochemical diversity across C10-C11 and C11-C12 in the same species of stigonematalean cyanobacteria

Due to the complexities of structural architectures of these hapalindole-type alkaloids and their promising biological activities, it has led to a number of synthetic studies by many research groups, resulting in over a dozen of total chemical synthesis of hapalindole-type molecules [12]. In spite of numerous total synthetic efforts, the knowledge regarding the biosynthesis of these molecules at molecular levels was absent because no genes or enzymes involved in the biogenesis have been identified [11].

The initial biosynthetic hypothesis regarding the generation of hapalindole-type alkaloids was proposed by Moore and more recently by others [18-20]. Based on the structural similarities among hapalindole-type alkaloids, they proposed a common biosynthetic origin that involve a chloronium ion or a proton-catalyzed condensation of a tryptophan-derived 3-(2'-

isocyanoethanyl) indole **1** (*E* or *Z*) with a monoterpene β-ocimene **2** (*E* or *Z*) to provide tricyclic

hapalindole core intermediate. The tricyclic hapalindole core can then undergo further oxidation

and cyclization to other classes of hapalindole-type molecules. Based upon this biosynthetic

hypothesis, the strain-specific stereochemical diversity should arise from the biosynthetic

precursors in different stereoconfigurations (*E* or *Z*) (Figure 2). For example, the *syn-syn*

configuration across C10-C11 and C11-C12 of hapalindole G produced by *F. ambigua* UTEX

1903 would arise from *(Z)*-**1** and *(Z)*-**2**. On the other hand, the *syn-anti* configuration of

hapalindoles/fischerindoles in *H. welwitschii* UTEX B1830 would arise from *(Z)*-**2** and *(E)*-**1**.
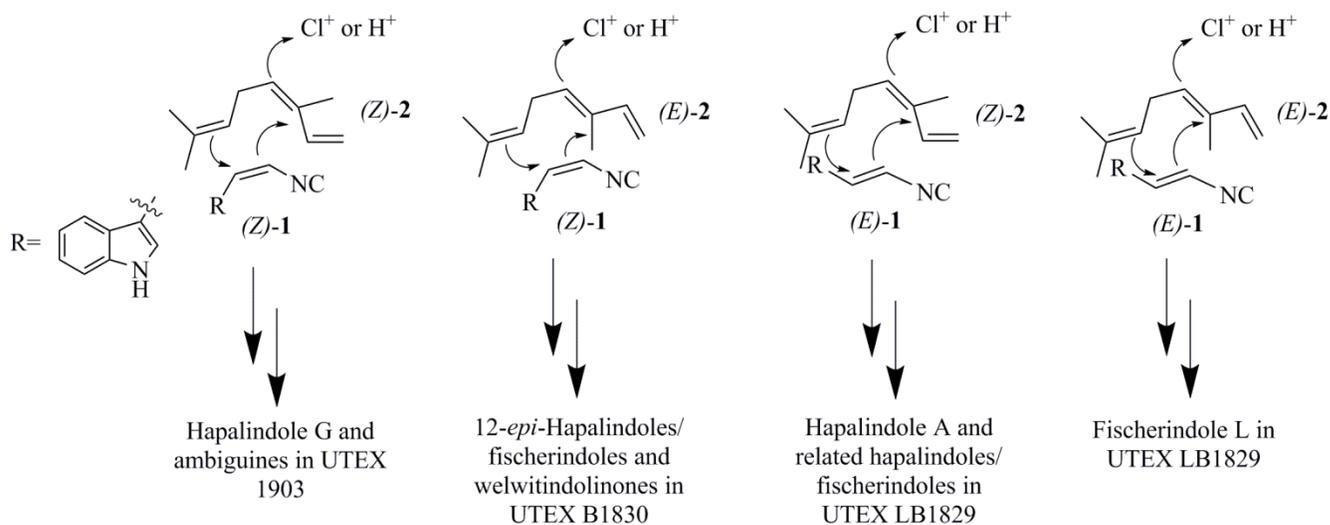
Figure 2. Existing hapalindole biosynthetic hypothesis suggests the stereochemical diversity of hapalindole-type

molecules arises from biosynthetic precursors

Since the knowledge regarding the biosynthesis of hapalindole-type alkaloid at both

genetic and molecular levels was absent, our group started a collective effort to address these

fundamental knowledge gaps. These efforts have culminated to the identification of the

biosynthetic gene clusters for hapalindole-type alkaloids ambiguines [11] and welwitindolinones

[21].

6

The first biosynthetic gene cluster identified was ambiguines produced by *F. ambigua* UTEX1903, where the identification was carried out by genome sequencing. There were a few considerations that ambiguines from UTEX1903 was chosen as the initial target for gene cluster identification. First, ambiguines are the only member of hapalindole-type molecules with a *tert*-prenyl group modification at C-2 of the indole ring, suggesting the involvement of an indole prenyltransferase. Second, the relative stereochemistry across C10-C11 of indole ring and isonitrile group in all ambiguines is *cis*, suggesting that they should all be derived from a common biosynthetic precursor, *(Z)*-**1**, indicating the involvement of isonitrile synthase superfamily [22-23]. These considerations provided bioinformatic clues to mine the target gene cluster. In addition, *F. ambigua* UTEX1903 was the only available axenic cyanobacterial producer of hapalindole-type molecules at the time. The fact that *F. ambigua* UTEX1903 is axenic would facilitate the efforts in assembling raw genomic data obtained from next generation sequencing. *In vitro* characterization was used to disclose the roles of key enzymes in the ambiguine biosynthesis pathway. This study confirmed the function of isonitrile synthases, namely AmbI1, AmbI2 and AmbI3, responsible for the stereoselective assembly of *(Z)*-**1** precursor. It also ruled out the involvement of *(Z)*-**2** and demonstrated the core hapalindole scaffold is not generated from the chloronium ion or proton-catalyzed polycyclization of **1** with **2** as proposed in the literature. Instead, the core biosynthetic precursors leading to the formation of core terpenoid indole scaffold are *(Z)*-**1** and GPP.
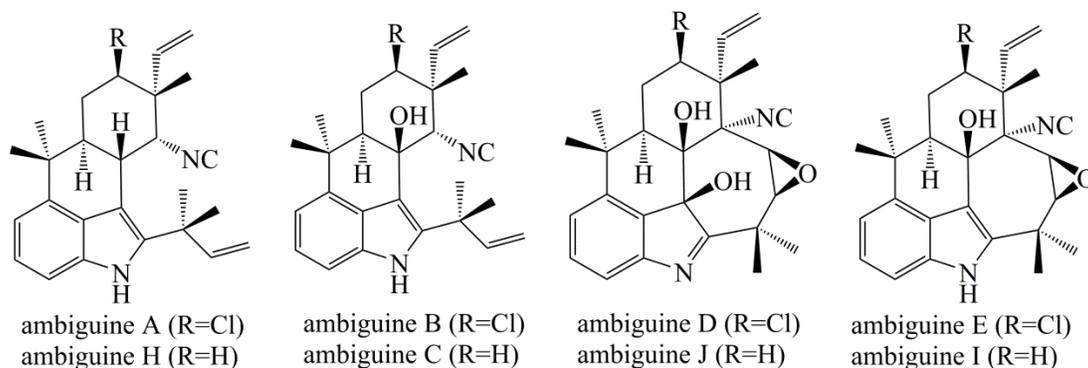
Figure 3. Structure of ambiguines produced by *F. ambigua* UTEX1903

The second hapalindole-type-molecule biosynthetic gene cluster identified by our group was welwitindolinones produced by *H. welwitschii* UTEX B1830. With identical terpenoid backbone stereochemistry as ambiguines at C10-C11, welwitindolinones, hapalindoles and fischerindoles isolated from *H. welwitschii* UTEX B1830 possess inverted stereochemistry at C12. A question then arises as to what causes the stereochemical diversity. According to Moore's original hypothesis, this stereochemical discrepancy should have been derived from *(E)-***2** and *(Z)-***1** precursors. However, *in vitro* characterization demonstrated, in agreement with ambiguine biosynthesis study, that **2** was not involved in the biosynthesis and that the biosynthetic precursors were geranyl pyrophosphate (GPP) and *(Z)-***1**, where *(Z)-***1** was assembled via isonitrile synthases, WelI1, WelI2 and WelI3 (Figure 6).
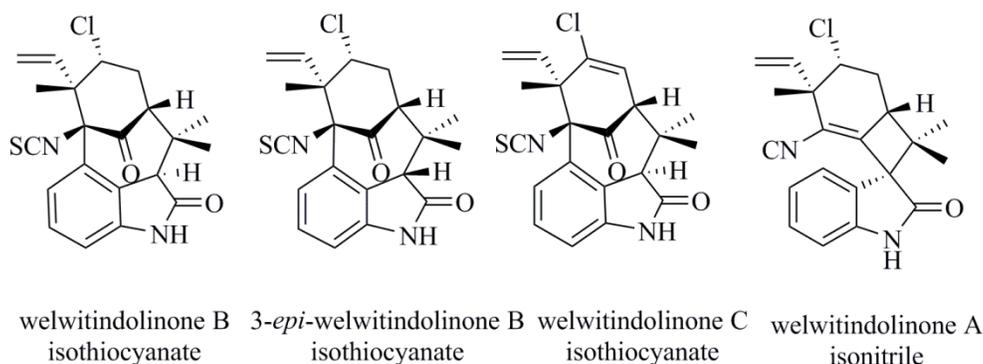
Figure 4. Structures of welwitindolinones produced by *H. welwitschii* UTEX B1830

8

The results from the *amb* and *wel* biosynthetic pathways studies revealed the common biosynthetic precursors in both pathways, GPP and *(Z)*-**1**, regardless of stereochemical diversity at C11-C12 (Figure 7). Further comparative biosynthetic analysis of ambiguine (*amb*) in *F. ambigua* UTEX 1903 and welwitindolinone (*wel*) in *H. welwitschii* UTEX B1830 demonstrated that the sequence homologies of the core precursor biosynthetic and regulatory proteins, including DMAPP synthase (D1-4), tryptophan synthase (T1-5), isonitrile synthase (I1-3), prenyltransferases (P1-2), and other core proteins, are more than 90% as shown in the gene cluster comparison in Figure 5. On the other hand, the sequence homologies of NHI-dependent oxygenases, such as Rieske oxygenases (O1-4), 2OG-Fe(II) oxygenase (O5) (Figure 5). This implied that all hapalindole-type compounds from each species all share the same proteins, D1-4 and P2, to produce GPP and, T1-5 and I1-3, to produce *(Z)*-**1** (Figure 8).
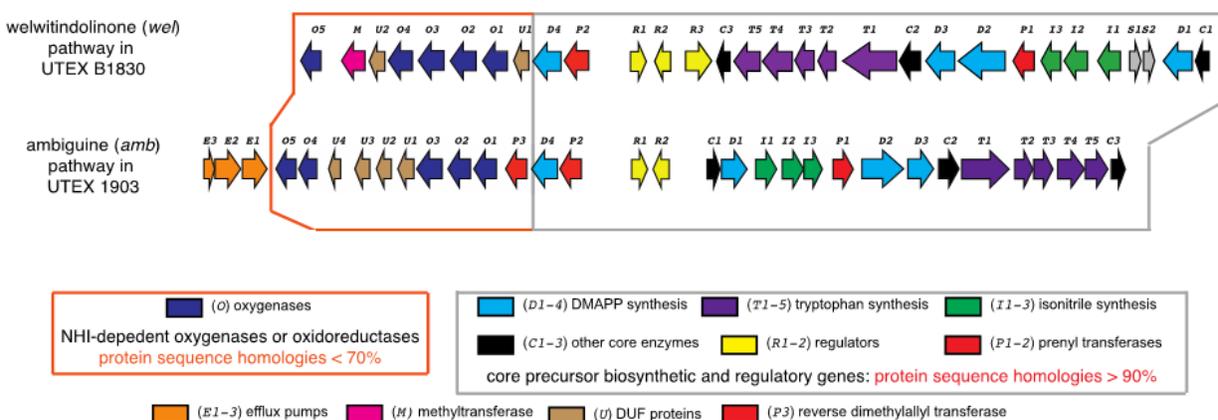
Figure 5. Comparative biosynthetic gene clusters of welwitindolinone (*wel*) in UTEX B1830 and ambiguine (*amb*) in UTEX 1903. Gene functions are grouped on the basis of their putative roles associated with biosyntheses.
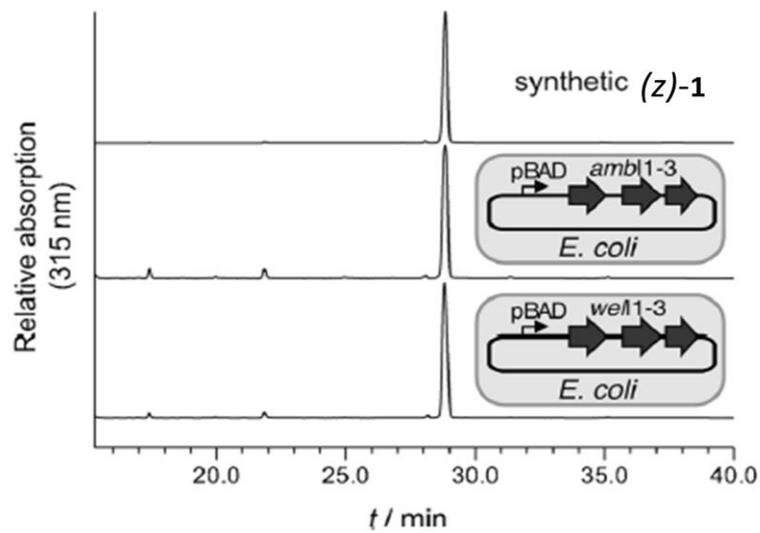
Figure 6. *In vivo* characterization of WelI1-I3 and AmbI1-I3 enzymatic product showing both give the same product

which is *(Z)*-**1** [21]
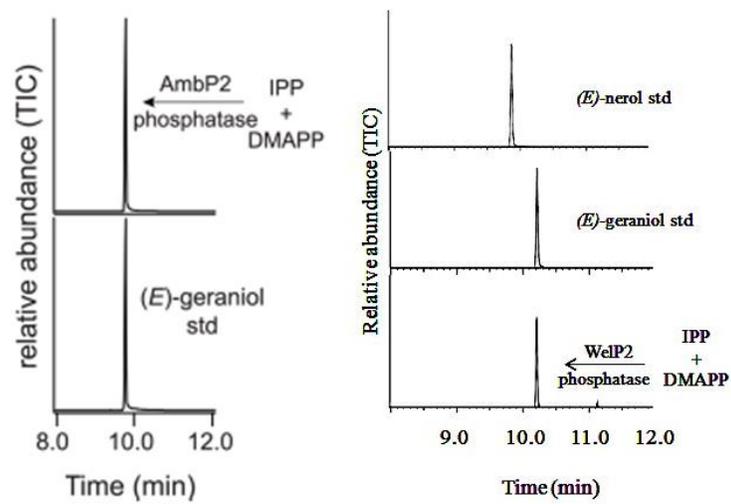
Figure 7. GC-MS chromatograph showing that AmbP2 and WelP2 are both a dedicated GPP synthase from IPP and

DMAPP
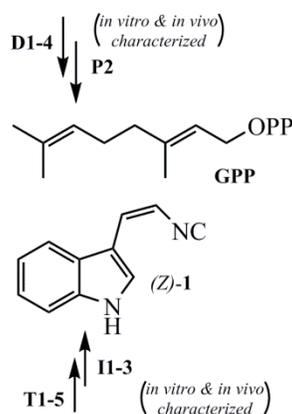
Figure 8. Current view of hapalindole-type alkaloids biosynthesis at the molecular level involving early assembly of GPP and *(Z)*-**1** as the biosynthetic genes, D1-4, P2, I1-3, T1-5 in ambiguine and welwitindolinone biosynthesis are functionally identical

For later stage modification of the precursors, GPP is then transferred to the aromatic backbone of *(Z)*-**1** by an aromatic prenyltransferase, AmbP1 (*amb* pathway) or WelP1 (*wel* pathway). Our early hypothesis regarding the function of AmbP1 and WelP1 is that they were responsible for directly fusing GPP and *(Z)*-**1** to generate hapalindole U in *amb* pathway or 12-*epi*-fischerindole U in *wel* pathway, respectively. However, because the sequences of WelP1 and AmbP1 are nearly identical, it is unlikely that they will perform different chemistry to generate hapalindole U and 12-*epi*-fischerindole U [21]. The actual function of P1 remains under investigation in our laboratory.

Based on the comparative analysis of the biosynthetic gene cluster for ambiguines, welwitindolinones, and fischerindoles (Figure 5), we have proposed that the aromatic prenyltransferase (P3) and nonheme iron (NHI)-dependent oxygenases, especially Rieske-type oxygenases would be responsible for the diversification to generate structurally and stereochemically different tri-, tetra-, or pentacyclic hapalindole-type molecules [11, 21]. Our lab has recently validated that AmbP3 is a dimethylallyltransferase that regioselectively converts

11

hapalindole G to ambiguine A [11], while AmbO5 is an α-ketoglutarate (α-KG)-dependent halogenase that transform ambiguine H to ambiguine A by monochlorination at C-13 position. Similarly in *wel* pathway, we have validated that WelO5 is an α-KG-dependent halogenase that transform 12-*epi*-fischerindole U to 12-*epi*-fischerindole G also by monochlorination at C-13 position [24] (Figure 9).
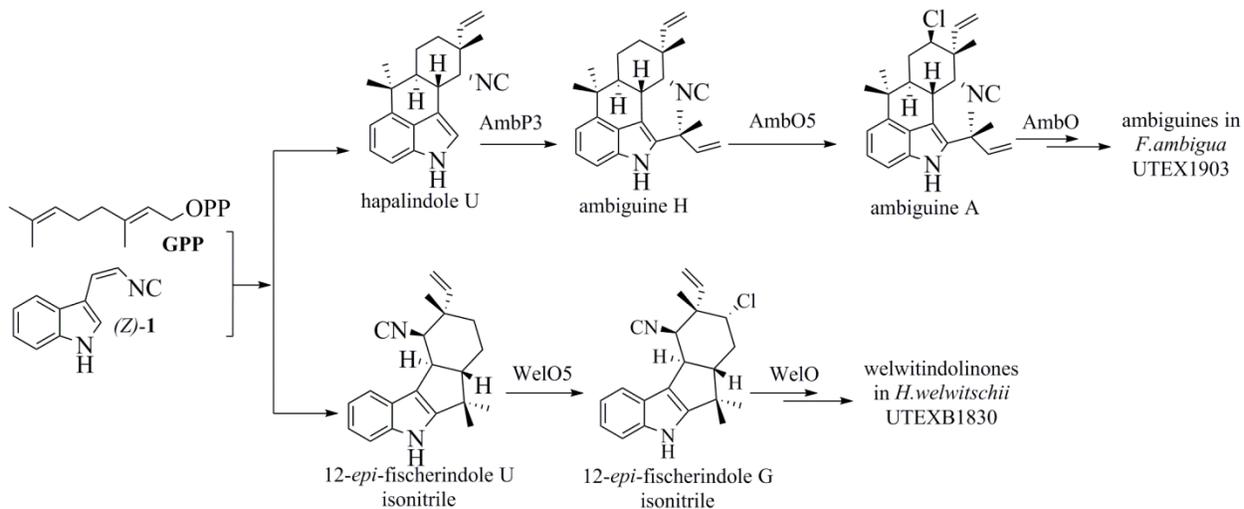
Figure 9. Current view on the late-stage diversification of hapalindole-type molecules. The function of AmbP3, AmbO5 and WelO5 have been validated

In summary, the previous works from our lab demonstrated that all hapalindole-type alkaloids produced by three different species of stigonematalean cyanobacteria, *F. ambigua* UTEX 1903, *H. welwitschii* UTEX B1830 and *F. musicola* UTEX LB1829, share the same biosynthetic precursors, GPP and *(Z)*-**1**. As *(Z)*-**1** biosynthesis was identified to involve a three enzyme pathway (I1, I2, I3), contrasting the earlier observations that only requires two enzymes (IsnA, IsnB), the initial question I try to address is the unique aspect of this three enzyme pathway. This portion of the work is summarized in Chapter 2.

In addition, since *(Z)*-**1** is identified as a universal precursor to the biosynthesis of hapalindole-type molecules, I also initiate an effort to study whether this pathway is promiscuous to allow the incorporation of unnatural tryptophan. This portion of the work is summarized in Chapter 3.

2.0 BIOCHEMICAL INVESTIGATION OF THREE ENZYME PATHWAYS FOR *(Z)*-INDOLE VINYL ISONITRILE BIOSYNTHESIS IN STIGONEMATALEAN CYANOBACTERIA

2.1 PREVIOUS WORKS ON *(Z)*-INDOLE VINYL ISONITRILE BIOSYNTHESIS

Shortly after our group disclosed the biosynthetic gene clusters of ambiguines and welwitindolinones that led to the discovery of both pathways sharing the same biosynthetic precursors, the effort was initiated to investigate the biosynthesis of one of the precursors, *(Z)*-indole vinyl isonitrile **1**. The three biosynthetic genes of *(Z)*-**1** are *amb*I1-I3 in the *amb* pathway (or *wel*I1-I3 in the *wel* pathway). The identification of a three enzyme pathway for *(Z)*-**1** biosynthesis contrasts the early study by Brady and Clardy [23, 25], who showed that proteins IsnA and IsnB are two core proteins for the biosynthesis of *(E)*-**1** from L-tryptophan and ribulose 5-phosphate (5P). By sequence identity, IsnA is closely related to PvcA of the pyoverdine chromophore biosynthesis (33%) and Dit1 of yeast-spore wall biosynthesis (21%), both of which are thought to be involved in the biosynthesis of C-N bonds (isonitrile synthase) [22]. IsnB shows the highest sequence identity to PvcB, which is a non-heme iron α-KG dependent oxygenase in the pyoverdine chromophore biosynthesis. By feeding experiment and systematic differential isotope labeling, they found that the source of nitrogens and isonitrile carbon in *(E)*-**1** are from tryptophan and C-2 of ribulose-5P, respectively. In addition, ribose-5P and arabinose-

14

5P were also found to be cognate substrates for IsnA/IsnB [25]. It was hypothesized that ribose-5P/ arabinose-5P can be nonenzymatically tautomerized to ribulose-5P [25].
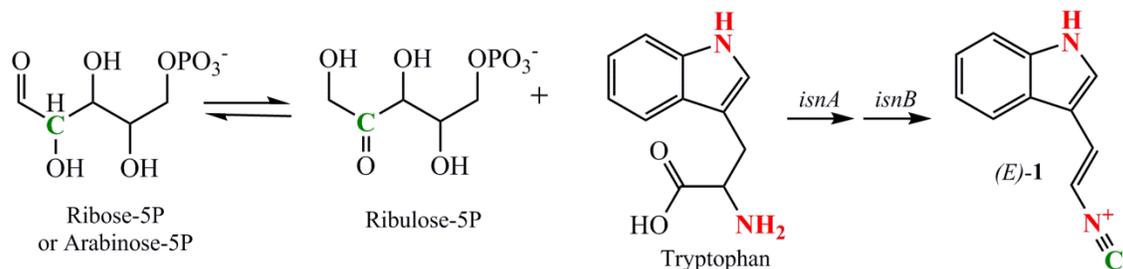
Figure 10. Substrates for *(E)*-**1** biosynthesis by IsnA/IsnB [25]
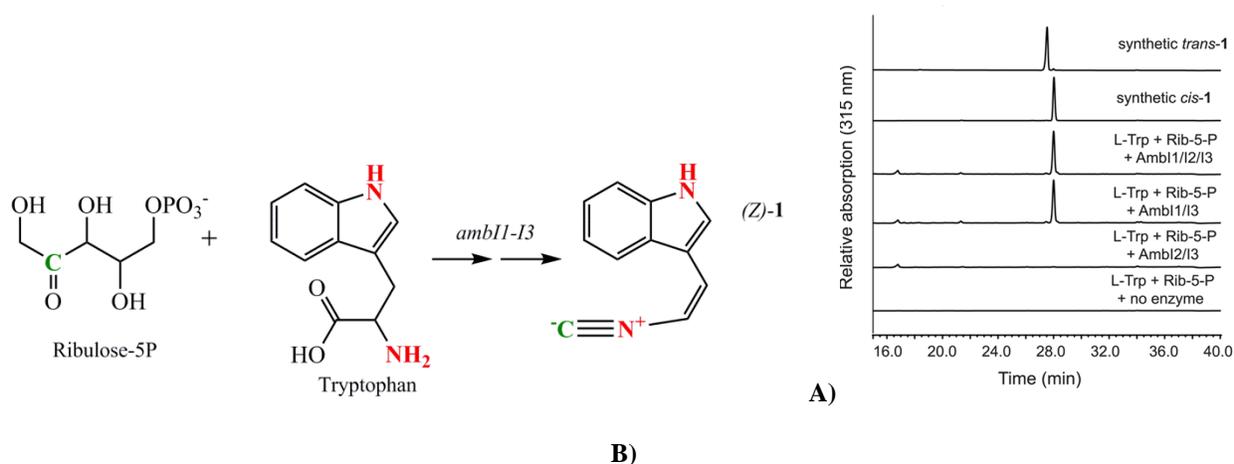
Figure 11. A) *(Z)*-**1** biosynthesis from ambiguine (*amb*) biosynthetic pathway, B) HPLC chromatograph showing the production of *(Z)*-**1** instead of *(E)*-**1**, and the redundancy of AmbI2 in *(Z)*-**1** biosynthesis from L-Trp and ribulose-5P *in vitro* [11]

By sequence homology, AmbI1-I3 are homologues of isonitrile synthase IsnA/IsnB, where both AmbI1 and AmbI2 share homology to IsnA (46%/67% and 47%/69% identity/positivity) while AmbI3 to IsnB (48%/68% identity/positivity) [11].

To confirm the production of *(Z)*-**1** by AmbI1-I3, we first carried out an *in vitro* reconstitution experiment. To reconstitute AmbI1-I3 enzymatic assay *in vitro,* individual

15

AmbI1, AmbI2 and AmbI3 enzymes were overexpressed and purified as N-His-tagged proteins in *E. coli* and subsequently treated with L-tryptophan, ribulose-5P in equal molar amounts in the presence of α-KG and Fe(II). As a result, it was found that *(Z)*-**1** was produced instead of *(E)*-**1** as shown in HPLC analyses in Figure 11. Additionally, we discovered that AmbI1 and AmbI3 alone were sufficient to generate *(Z)*-**1** when using L-Tryptophan and ribulose-5P as substrates, but AmbI2 and AmbI3 were not [11].

This discovery indicated the possible redundanct role of AmbI2 in *(Z)*-**1** biosynthesis. In order to verify the *in vitro* result, an *in vivo* experiment was similarly implemented. We found that, in contrast to the *in vitro* result, AmbI1, AmbI2, and AmbI3 were all required for the generation of *(Z)*-**1** *in vivo* (Figure 12). This has raised interesting questions about the functional role of AmbI2 in the context of *(Z)*-**1** biosynthesis. To this end, we initiated a study to investigate the role of individual AmbI2 as compared with AmbI1 and AmbI3 enzymes.

To address the initial conflicting results in the functional role of AmbI2 from *in vivo* and *in vitro* studies, we initiated a collective effort using *in vivo* heterologous expression in *E. coli* and *in vitro* reconstitution that would allow us to understand the underlying significance of AmbI2 in *(Z)*-**1** biosynthesis.
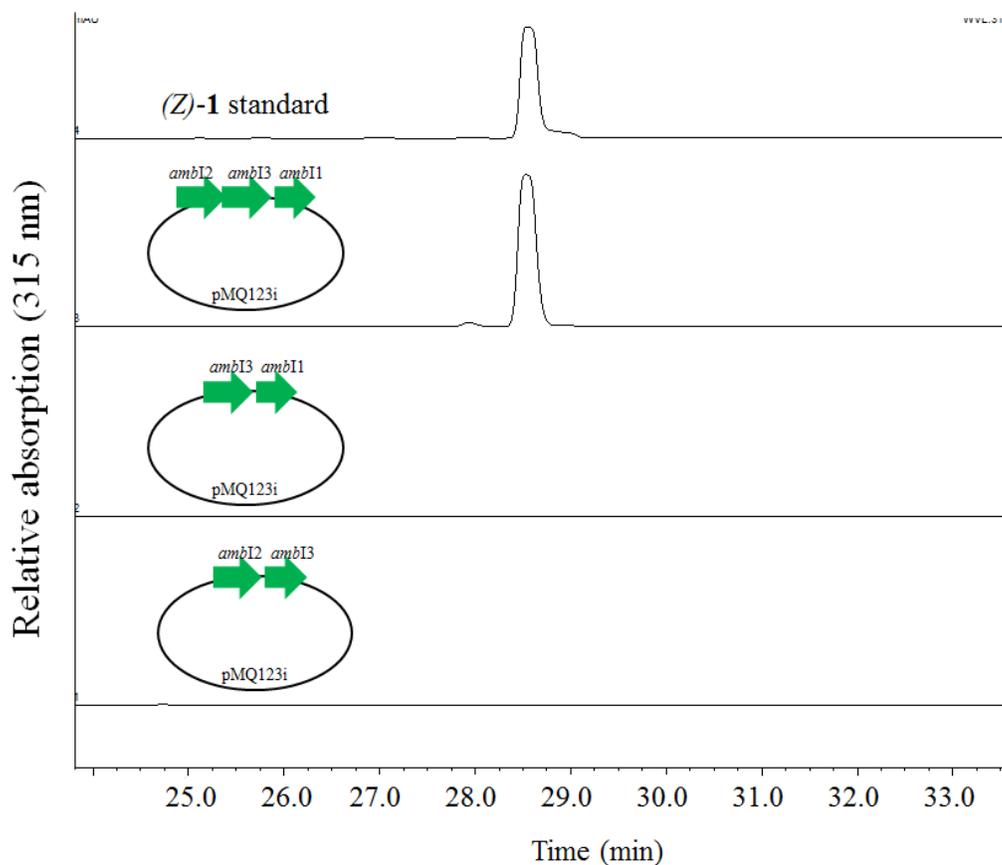
Figure 12. HPLC chromatograph showing **A)** the redundancy of AmbI2 in *(Z)-*1 biosynthesis from L-Trp and ribulose-5P *in vitro* [11] **B)** the generation of *(Z)-*1 *in vivo* when *amb*I1-I3 are all present. No trace of *(Z)-*1 observed when *amb*I2 or *amb*I1 is absent.

2.2   THE ROLE OF AMBI2 ENZYME IN THE BIOSYNTHESIS OF *(Z)-*INDOLE VINYL ISONITRILE *IN VIVO*

We first investigated the functional role of AmbI2 by using an established heterologous expression system in *E. coli*.

Three different *amb*I1-I2-I3 expression constructs were generated. The first construct was wild type *amb*I1-I3 with full-length *amb*I2. The second and third constructs were identical to the

first construct except *amb*I2 gene is truncated from 5' end by approximately 200 base pairs and 500 base pairs. *amb*I2-I3 was cloned as a single operon, while the ribosome binding site (RBS) was inserted upstream of *amb*I1 to allow polycistronic expression of three enzymes under the control of pTac promoter. The successfully constructed plasmids were then transformed into *E. coli*. and induced by IPTG to allow the overexpression of AmbI1-I3. The ethyl acetate extract from the supernatants were analyzed by HPLC for *(Z)*-**1** identification.
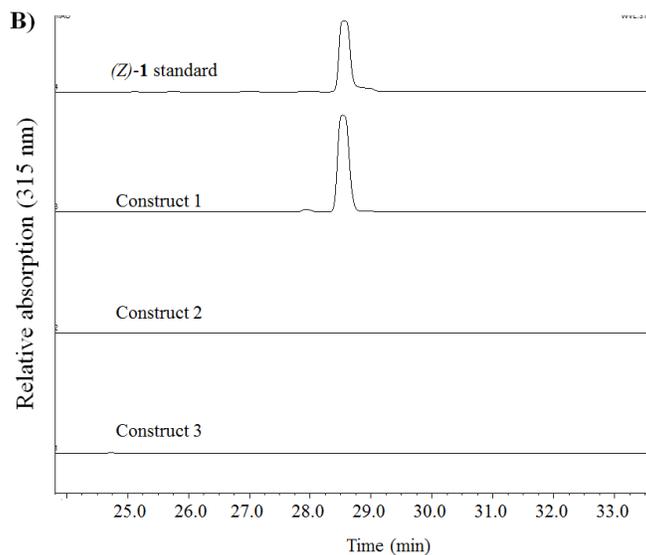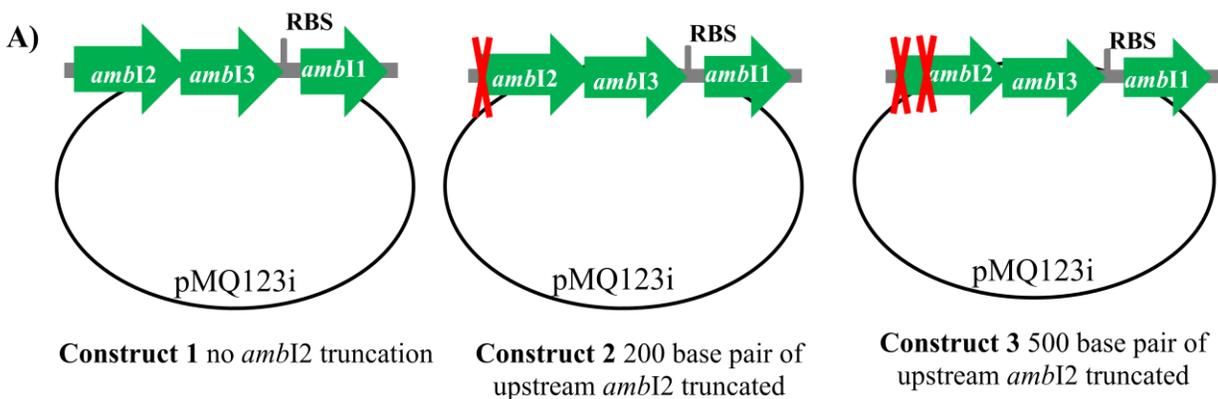
Figure 13. A) *amb*I2 truncation experiment where approximately 200 and 500 base pair of upstream *amb*I2 was truncated. Ribosome binding site (RBS) is located upstream of *amb*I1 B) Comparative HPLC analysis of extracts from construct 1-3.

The results from this study showed that, only construct 1 when full-length *amb*I1-I2-I3 are expressed, can produce *(Z)*-**1**. On the other hand, no *(Z)*-**1** can be generated in construct 2 and construct 3, where *amb*I2 gene is truncated, according to HPLC analyses (Figure 13B). This suggested that full-length *amb*I2 gene was necessary for the production of *(Z)*-**1** *in vivo*.

Since it became apparent from the *amb*I2 truncation experiment that full-length AmbI2 was required for the production of *(Z)*-**1** *in vivo*, the next goal was to investigate the effect of AmbI2 expression level in *E. coli*. To this end, we designed the complementation experiment where complete *amb*I2 and *amb*I1/*amb*I3 were cloned separately into two different types of plasmid vector. *amb*I2 was cloned into pMQ131, a low-copy pBBR1 origin of replication (5-20 copy per cell) with pLac promoter, and *amb*I1/I3 was cloned into pMQ123i, a high copy ColE1/RK2 origin of replication (200-300 copy per cell) with pTac promoter [26]. Both constructs were then co-expressed heterologously in *E. coli*. The use of different copy-number vectors facilitated the investigation regarding effect of AmbI2 expression level, as compared to AmbI1/AmbI3, to the production of *(Z)*-**1**. The scheme illustrating this experiment is shown in Figure 14. Two respective constructed plasmids were then co-transformed into *E. coli*. to express the genes followed by ethyl acetate extraction to recover *(Z)*-**1**.

The results from this study demonstrated that, by complementing *amb*I2 via a low copy plasmid, the quantity of *(Z)*-**1** generated is similar to that generated by the wild-type full *amb*I1-I3 in a high copy plasmid (construct 1 in Figure 13A). On the other hand, the quantity of *(Z)*-**1** generated is significantly higher than that produced by the wild-type full *amb*I1-I3 in a low copy plasmid (Figure 14).
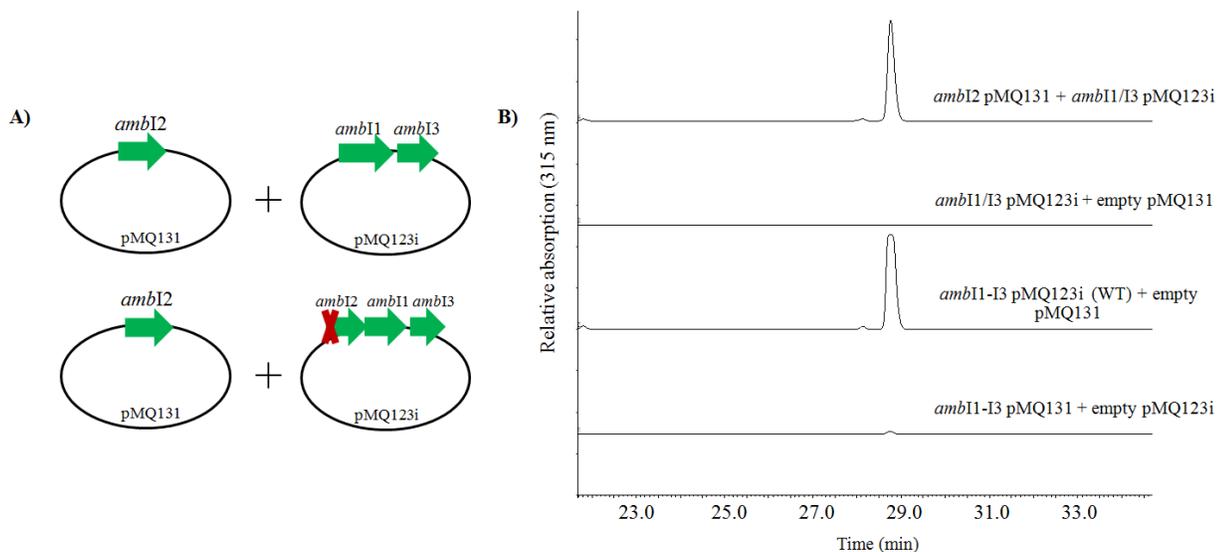
Figure 14. A) AmbI2 complementation experiment, B) Comparative HPLC analysis of extracts from *amb*I2 complementation experiment compared with wild- type (WT) full-length *amb*I1-I3 in both high copy (pMQ123i) and low copy (pMQ131) plasmids

To compare the role of AmbI2 with AmbI1 and AmbI3, we carried out a similar comparative complementation experiment with Amb1 and AmbI3. *amb*I1 and *amb*I2/*amb*I3 or *amb*I3 and *amb*I1/*amb*I2 were cloned separately into two different types of expression vector. *amb*I1 or *amb*I3 was cloned into a low-copy pMQ131 vector while *amb*I2/*amb*I3 or *amb*I1/*amb*I2 was cloned into a high-copy pMQ123i. *amb*I1 in pMQ131 was co-expressed with *amb*I2/*amb*I3 in pMQ123i while *amb*I3 in pMQ131 was co-expressed with *amb*I1/*amb*I2 in pMQ123i (Figure 15A). For AmbI1 and AmbI3, the previous studies had shown that both must be present in order to generate *(Z)*-**1** both *in vitro* and *in vivo,* which suggested the vital roles of AmbI1 and AmbI3 in the biosynthesis of *(Z)*-**1** [11, 23]. The results from this study supported the importance of the roles of AmbI1 and AmbI3 because it was found that the production of *(Z)*-**1** from *amb*I1 and *amb*I3 complementation was not as high as that of *amb*I2 complementation even though we observed that the production of *(Z)*-**1** from *amb*I3 complementation was less than that

20

of *amb*I1 (Figure 15B). These studies confirmed the vital roles of AmbI1 and AmbI3 in *(Z)*-**1** biosynthesis because the production of *(Z)*-**1** depends on the expression level of respective enzymes. On the other hand, only substoichiometric expression level of AmbI2, relative to AmbI1/I3, is required to robustly generate *(Z)*-**1** *in vivo*.
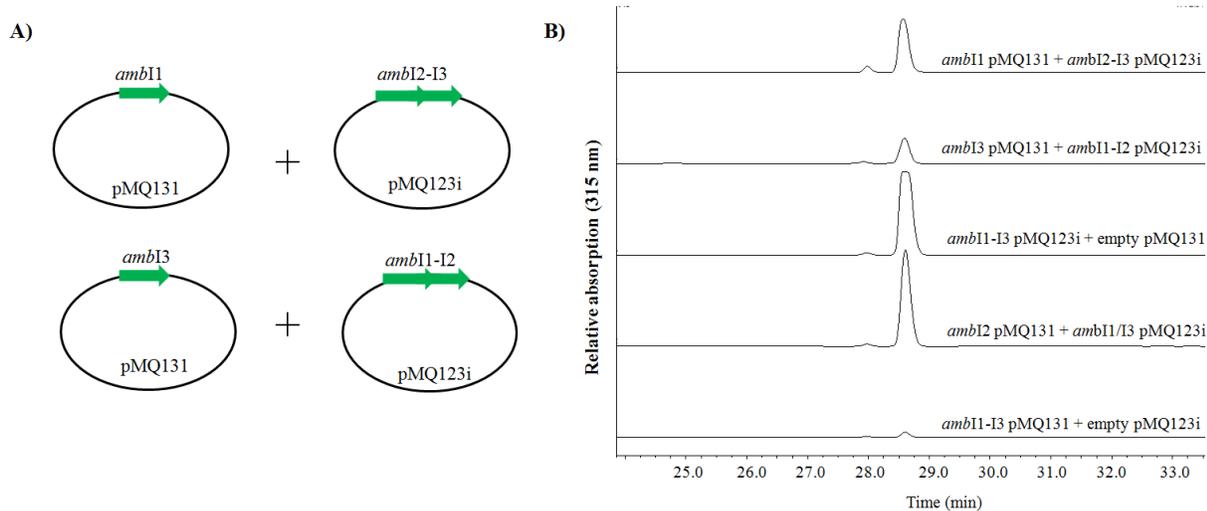
Figure 15. A) AmbI1 and AmbI3 complementation experiment to compare with AmbI2 complementation experiment, B) Comparative HPLC analysis of extracts from *amb*I1 and *amb*I3 complementation experiment compared with *amb*I2 complementation and wild- type (WT) full-length *amb*I1-I3 in both high copy (pMQ123i) and low copy (pMQ131) plasmids.

	Quantity (µg/L)
*amb*I1 pMQ131 + ambI2-I3 pMQ123i	225.97
*amb*I3 pMQ131 + *amb*I1-I2 pMQ123i	61.2
*amb*I1-I3 pMQ123i + empty pMQ131	434.65
*amb*I2 pMQ131 + ambI1/I3 pMQ123i	373.93
*amb*I1-I3 pMQ131 + empty pMQ123i	21.86

Table 2. Quantity of *(Z)*-**1** produced in comparative complementation experiment based on standard curve

2.3 THE ROLE OF AMBI2 ENZYME IN THE BIOSYNTHESIS OF *(Z)*-INDOLE VINYL ISONITRILE *IN VITRO*

After we disclosed that only substoichiometric amount of AmbI2, relative to AmbI1 and AmbI3, is required for the production of *(Z)*-**1** in *E. coli*, we further investigated the actual role of AmbI2 by *in vitro* study. First, we found that the AmbI1-I3 *in vitro* assay did not perform well in low pH condition as reported for the IsnA/IsnB study by Brady [25]. Therefore, we optimized the condition for AmbI1-I3 *in vitro* assay and found that pH = 6.5-7.4 is the optimized condition for the most effective production of *(Z)*-**1** by AmbI1-I3. We subsequently chose to perform *in vitro* assays at pH=7.4.
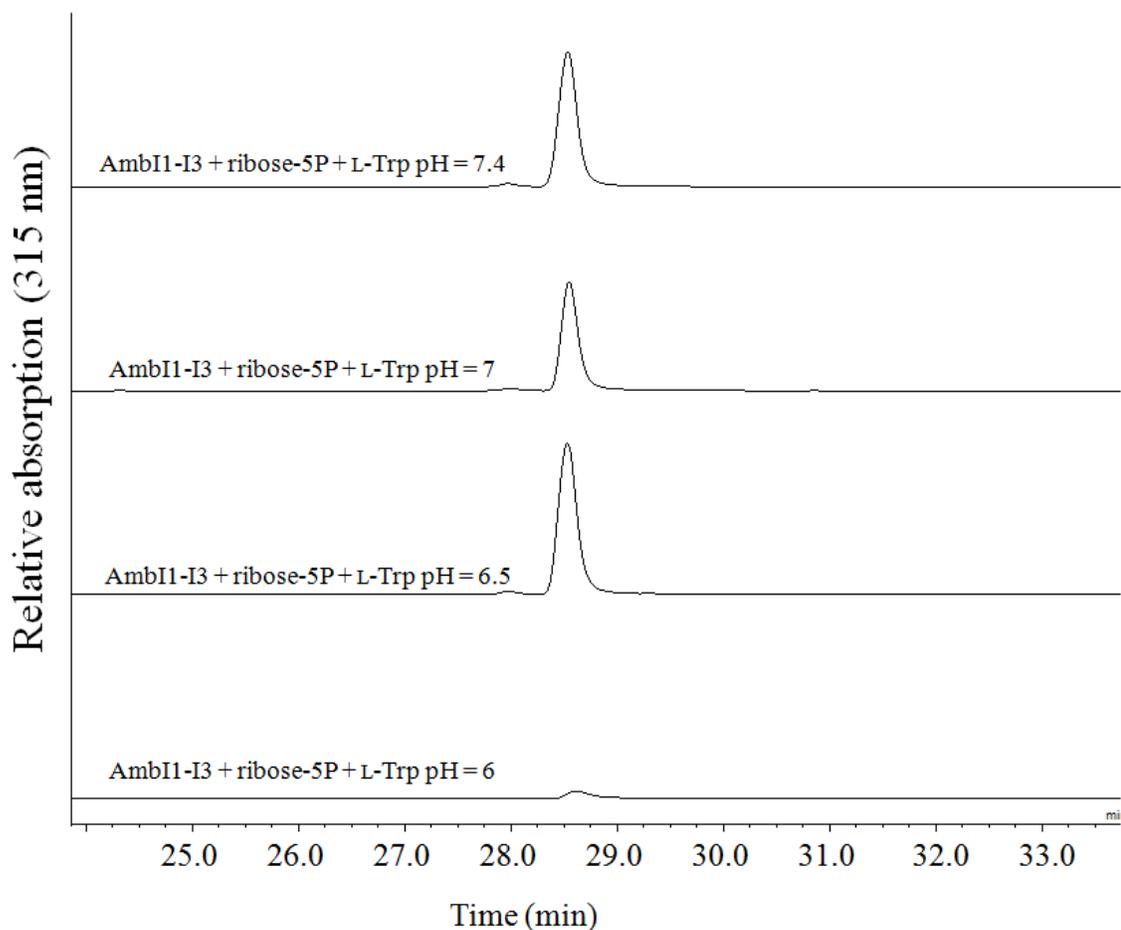
Figure 16. HPLC chromatograph analysis of AmbI1-I3 assay showing that the optimized condition for the most effective production of *(Z)*-**1** is at neutral pH. The elution time of *(Z)*-**1** is at 28.8 minutes.

The initial result of the *in vitro* investigation of the role of AmbI2 revealed some interesting evidence that can be correlated with the function of AmbI2 *in vivo*. This finding was due to the effort to reproduce the previous results in ambiguine biosynthetic study that AmbI1/I3 was adequate to generate *(Z)*-**1** by using L-tryptophan and ribulose-5P [11]. Since previous study demonstrated that either ribose-5P or ribulose-5P can be used as the substrate for the production of *(E)*-**1** by IsnA/B [25], ribose-5P was used to carry out the *in vitro* assays. However, we found that when ribose-5P was used as the substrate with AmbI1/I3, *(Z)*-**1** was not produced (Figure

17). On the other hand, in agreement with the previous ambiguine biosynthesis study, if ribulose-5P was used as the substrate with AmbI1/I3, *(Z)*-**1** was generated. This observation led us to hypothesize that AmbI2 could be responsible for the tautomerization of ribose-5P to ribulose-5P.
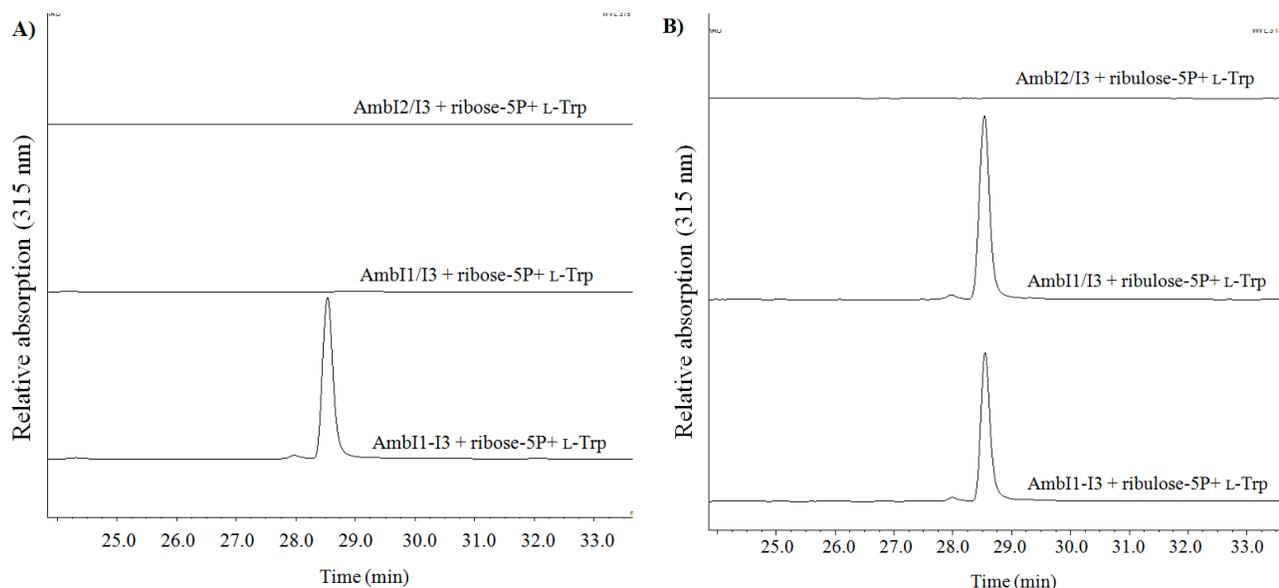
Figure 17. HPLC chromatograph with a UV detector showing the contradiction in using ribose-5P and ribulose-5P as the substrate in AmbI1−3 *in vitro* assay. AmbI1/I3 is suffice to produce *(Z)*-**1** if ribulose-5P was used whereas all three enzymes, AmbI1/I2/I3, were required to generate *(Z)*-**1** if ribose-5P was used. The elution time of *(Z)*-**1** is at 28.8 minute.
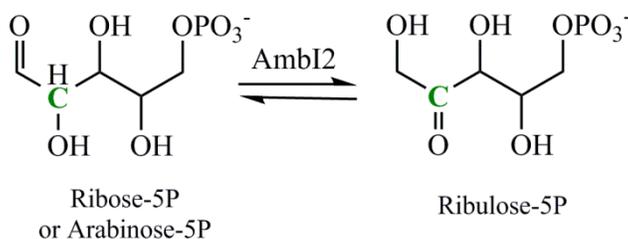
Figure 18. Proposed role of AmbI2 as ribose-5P isomerase to give ribulose-5P

24

To prove the hypothesis, given the fact that ribulose-5P has a UV absorbance peak at 280-290 nm whereas ribose-5P does not absorb UV light [27] (Figure 19), we monitored the reaction of AmbI2 with ribose-5P and observed the increase of the UV absorbance peak at 290 nm, indicating the possible role of AmbI2 as an isomerase (Figure 20).

As a further proof of the role of AmbI2 as an isomerase, we monitored the reaction of AmbI2 with ribose-5P by ^{31}P NMR, where the ^{31}P chemical shift of ribose-5P and ribulose-5P are 3.66 ppm and 4.16 ppm, respectively. This experiment indicated the conversion of ribose-5P to ribulose-5P as the NMR spectra showed that there was a new chemical shift occurred at 4.16 ppm in accordance with the ribulose-5P standard (Figure 21).
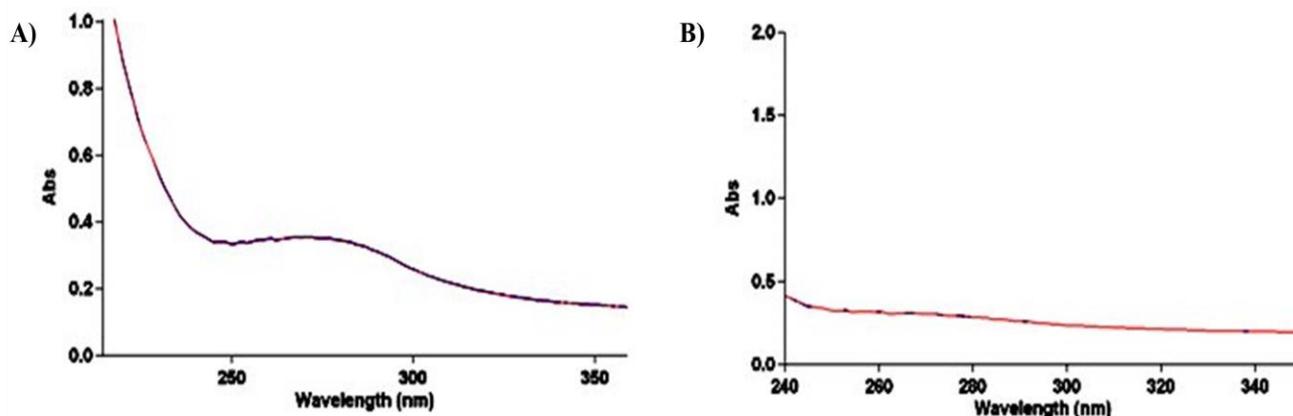
Figure 19. Absorbance spectral scan from 240-350 nm of A) 2.5 mM ribulose-5P and, B) 2.5 mM ribose-5P, showing the UV absorbance of ribulose-5P at 280-290 nm while no absorbance of ribose-5P
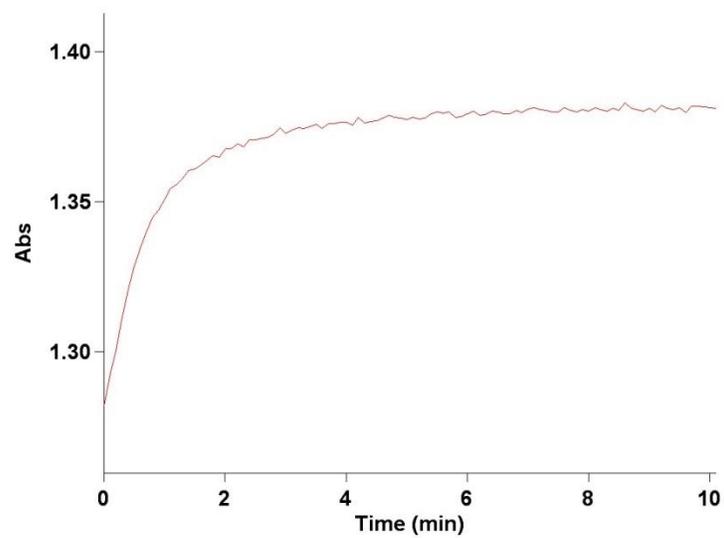
Figure 20. Real time absorbance at 290 nm monitoring AmbI2 (40 μM) reaction with ribose-5P (5 mM) in reaction

buffer (25 mM HEPES (pH=7.4), 150 mM NaCl, 5% glycerol) showing an increase of absorbance over time
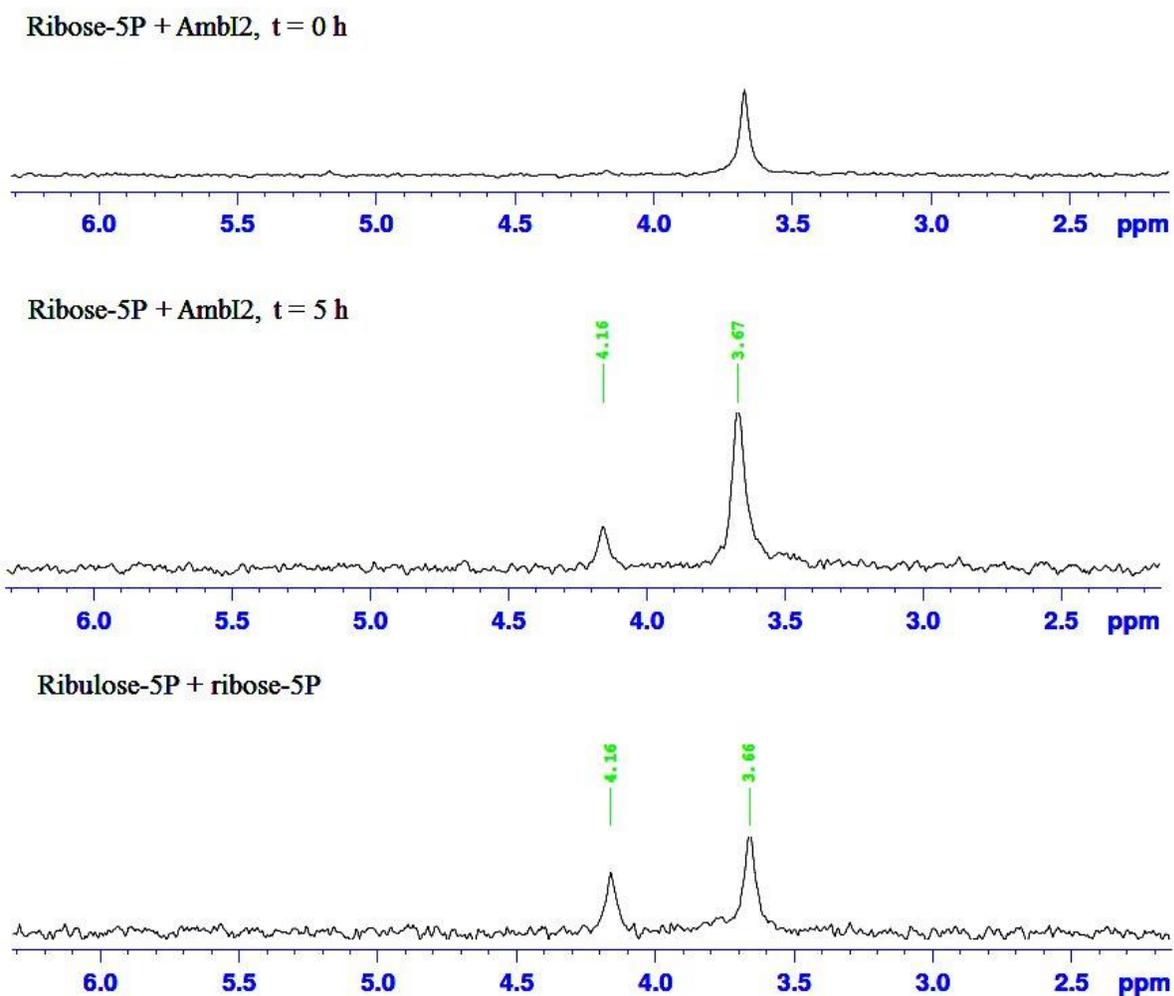
Figure 21. ^{31}P-NMR (600 MHz, D_2O) of the reaction between of ribose-5P (5 mM) and AmbI2 (40 μM) in reaction buffer (25 mM HEPES (pH=7.4), 150 mM NaCl, 5% glycerol) showing the conversion of ribose-5P to ribulose-5P after 5 hours. The chemical shift of ribulose-5P and ribose-5P at 4.16 ppm and 3.66 ppm, respectively, in accordance with the standard. The spectra were obtained from Dr. Qin Zhu.

2.4   CONCLUSIONS

From *in vivo* heterologous expression experiment, it can be concluded that full-length AmbI2 was necessary for the generation of *(Z)*-**1** *in vivo*. Additionally, the complementation experiment

27

showed that only substoichiometric quantity of AmbI2, relative to I1/I3, was required. Further *in vitro* reconstitution study revealed the possible role of AmbI2 as an isomerase that is responsible for converting ribose-5P to ribulose-5P as confirmed by UV spectrometry and ^{31}P-NMR. Collectively, these results implicate AmbI2 functions as a ribose-5P isomerase that can effectively recruit ribose-5P, an abundant metabolite in *E. coli*/cyanobacteria, to convert it to ribulose-5P as rare metabolite in *E. coli*/cyanobacteria essential for isonitrile biosynthesis.

3.0   PROMISCUITY OF *(Z)*-INDOLE VINYL ISONITRILE SYNTHASE IN

HAPALINDOLE-TYPE ALKALOID BIOGENESIS

3.1   INTRODUCTION

In the preceding chapters, we have described that all hapalindole-type indole alkaloids in all stigonematalean species share a common biosynthetic precursor, *(Z)*-**1**, regardless of the stereochemical discrepancy. During these investigations, it became clear I1-I3 is a universal three enzyme pathway responsible for the biosynthetic assembly of *(Z)*-**1.** Therefore, it would be interesting to study the promiscuity of this three enzyme pathway to produce halogenated *(Z)*-**1** as it can lead to the generation of halogenated analogs of hapalindole-type indole alkaloids. Although hapalindole-type indole alkaloids are extensively modified by chlorine and oxidation on their terpenoid backbones, their indole structure motifs are not naturally modified.

Modification of indole motif with halogens has been shown to enhance the biological activities and molecular properties of parental compounds [28-29]. One example is the study by Pauletti *et al.*, in which the biological activity of many indole alkaloids such as meridianins, psammopemmins, aplicyanins, and aplysinopsins were investigated. They concluded that the bromination of these indole alkaloid natural products could be associated with increased biological activity that they observed [29]. Given the fact that the family of hapalindole-type indole alkaloid compounds have been considered lead compounds for drug discovery, the

possible improvement of their biological activities of their halogenated analogs are of great interest [30-31].

The introduction of halogens to hapalindole-type alkaloids can be achieved by traditional approaches such as semi and total syntheses [32]. However, the drawback of this method is its impracticality for complex molecules [33]. An alternative approach involves the promiscuity of enzymatic pathway to incorporate unnatural halogenated substrates to the final products. An example of this approach is shown in the work of Roy *et al.* [33], in which *prn*A gene, the tryptophan 7-halogenase gene from pyrrolnitrin biosynthesis, was introduced to complement the pacidamycin biosynthetic gene in *Streptomyces coeruleorubidus* producer. As a result, chlorinated pacidamycins can be generated from halogenated tryptophan precursor. Another example is shown in the work of Runguphan *et al.* [34], in which they study the biosynthesis of plant monoterpene indole alkaloids halogenated analogs, including dihydroakuammicine, ajmalicine, catharanthine, and tabersonine, in *Catharanthus roseus* producer. Given that tryptophan is an early precursor in the biosynthesis of these plant alkaloids, they used *reb*H or *pyr*H genes, the tryptophan halogenase enzymes that selectively chlorinate tryptophan in the five and seven position, respectively, to complement the biosynthetic pathway. Introduction of these halogenase genes into *C. roseus* resulted in the regioselective production of halogenated analogs of plant alkaloid products.

These examples have shown the potential of early incorporation of halogenated tryptophans into the hapalindole-type indole alkaloid biosynthetic pathway as tryptophan is used at the beginning of the biosynthesis in early assembly of *(Z)*-**1** precursor. Therefore, the question I try to address is whether *(Z)*-indole vinyl isonitrile synthase is promiscuous as this will allow us

30

to explore the improvement of biological activity of halogenated analogs of hapalindole-type indole alkaloid natural products.

3.2   *IN VITRO* INVESTIGATION OF *(Z)*-INDOLE VINYL ISONITRILE SYNTHASE

We first investigated the promiscuity of AmbI1-I3 by using *in vitro* reconstitution. Hypothetically, if a halogenated tryptophan was used as a substrate for AmbI1-I3 instead of L-tryptophan, labeled *(Z)*-**1** could be produced (Figure 22).
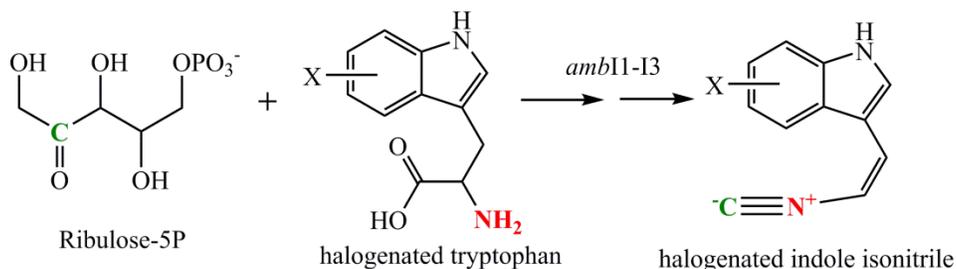
Figure 22. Halogenated indole issonitrile production from ribulose-5-phosphate and a halogenated tryptophan

To execute *in vitro* reconstitution, the optimized assay condition discussed in the previous chapter was used. The initial experiment was carried out using 7-fluoro tryptophan. The result indicated that 7-F tryptophan can be incorporated robustly to produce 7-F indole isonitrile. Competitive assay was also carried out to compare L-tryptophan with 7-F tryptophan where it showed that the effectiveness of both substrates were comparable (Figure 23).

31
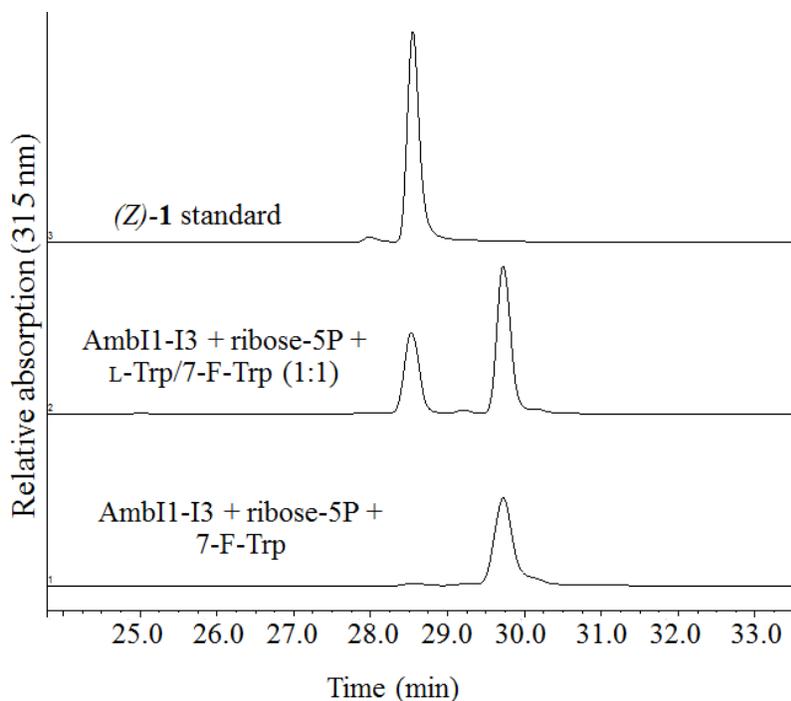
Figure 23. Initial assay showing the promiscuity of AmbI1-I3 by using 7-F tryptophan as a substrate

Now that we discovered the promiscuity of AmbI1-I3 with 7-F tryptophan, the next goal is to examine the enzymes' promiscuity with other halogenated substrates. A series of halogenated tryptophans were obtained from Rebeccca Goss' group [35]. The initial result indicated that the 4-, 5-, 6-, 7-F indole isonitrile can be robustly produced *in vitro* as shown in HPLC analysis (Figure 24). This suggests that the position of the halogens on the tryptophan indole ring had no significant effect.
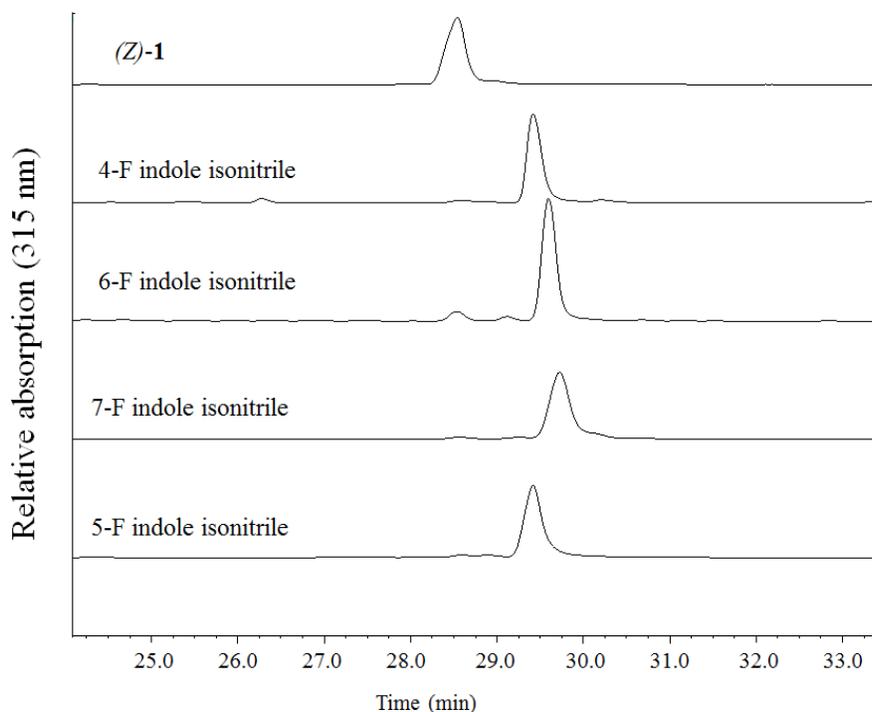
Figure 24. HPLC analysis of extracts from AmbI1-I3 *in vitro* reconstitution using unnatural substrates such as 4-, 5, 6- and 7- fluoro tryptophan. The ratio of *(Z)*-**1** to 4-F indole isonitrile to 6-F indole isonitrile to 7-F indole isonitrile to 5-F indole isonitrile is approximately 1:1.2:1.3:1:1

The next goal is to investigate the effect of the size of halogens on the indole ring on the effectiveness of AmbI1-I3 to generate halogenated *(Z)*-**1**. A series of halogenated tryptophans were used as a substrate for *in vitro* reconstitution compared to L-tryptophan. This included 5-fluoro, 5-chloro, 5-bromo, 7-fluoro, 7-chloro, 7-bromo, and 7-iodo tryptophans.

The results from this study indicated that the size of halogens on the indole ring had direct effect on the effectiveness of AmbI1-I3 to generate halogenated *(Z)*-**1**. The trend that we observed is that, the larger the size of the halogens on the tryptophan, the less ability of AmbI1-I3 to incorporate halogenated tryptophans to generate respective halogenated *(Z)*-**1** as seen in the comparative *in vitro* reconstitution using 5-F, 5-Cl, 5-Br tryptophans and 7-F, 7-Cl, 7-Br, 7-I tryptophans as the substrates. (Figure 25-26).
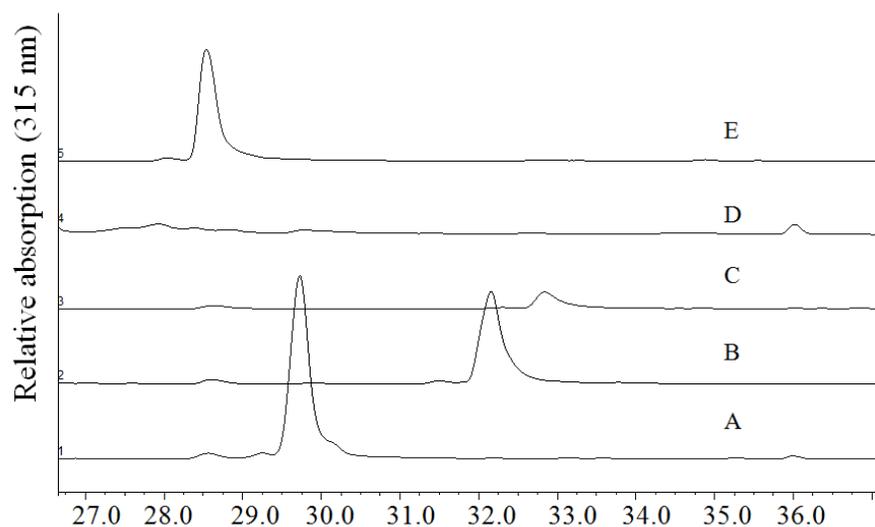
Figure 25. HPLC analysis of extracts from AmbI1-I3 *in vitro* reconstitution using unnatural substrates A) 7-fluoro, B) 7-chloro, C) 7-bromo, D) 7-iodo tryptophan and E) *(Z)*-**1**. The ratio of 7-F indole isonitrile to 7-Cl indole isonitrile to 7-Br indole isonitrile to 7-I indole isonitrile is approximately 4:3:1:0.1
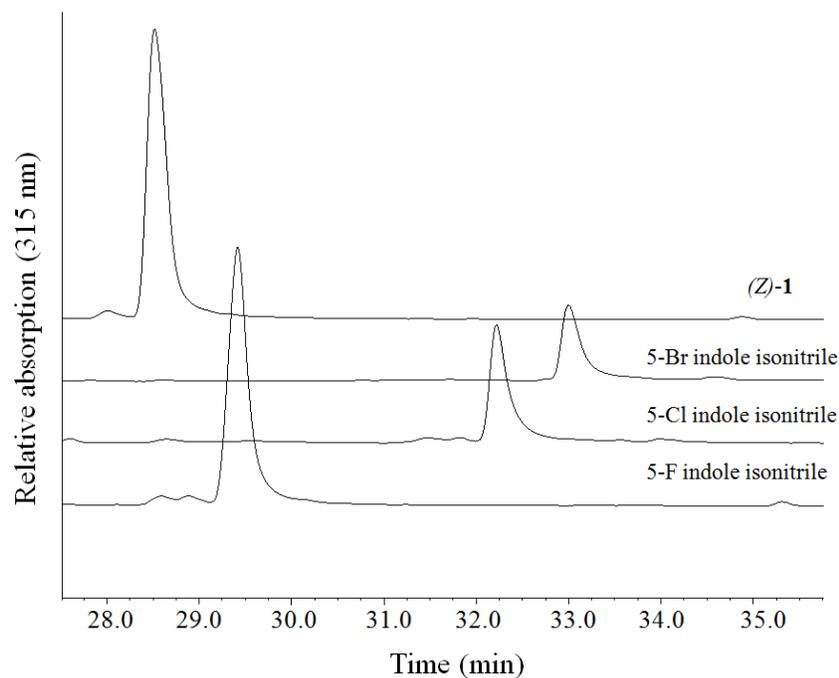
Figure 26. HPLC analysis of extracts from AmbI1-I3 *in vitro* reconstitution using unnatural substrates such as 5-fluoro, 5-chloro, and 5-bromo tryptophan compared with L-tryptophan. The ratio of 5-F indole isonitrile to 5-Cl indole isonitrile to 5-Br indole isonitrile is approximately 4:2:1

34

3.3   *IN VIVO* INVESTIGATION OF *(Z)*-INDOLE VINYL ISONITRILE SYNTHASE

AmbI1-I3 promiscuity investigation *in vivo* can be carried out by heterologously expressing AmbI1-I3 in *E. coli* in minimal media with the presence of a halogenated tryptophan. The result demonstrated, in agreement with *in vitro* result discussed previously, that 4-, 5-, 6-, 7-F indole isonitrile can be robustly produced *in vivo* as shown in HPLC analysis (Figure 27). This supported the earlier observation that the position of the halogens on the tryptophan indole ring had no significant effect.
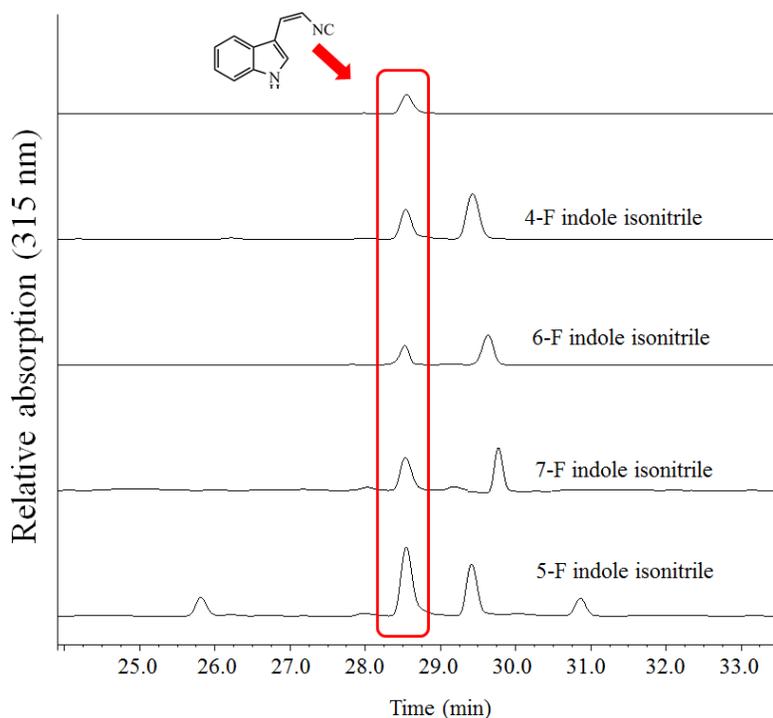
Figure 27. HPLC analysis of extracts from AmbI1-I3 *in vivo* overexpression using unnatural substrates such as 4-, 5-, 6- and 7- fluoro tryptophan.

To confirm the result in which the size of the substituted halogens contributed to the ability of AmbI1-I3 to incorporate unnatural tryptophans that we observed in competitive assay

35

in vitro, we implemented the competitive experiment *in vivo* using the equimolar mixture of L-tryptophan and 7-F or 7-Cl tryptophans. The result from this study correlated with the observation regarding the effect of the size of substituted halogens since significantly reduced quantity of Cl-indole isonitrile was produced as compared with F-indole isonitrile (Figure 28).
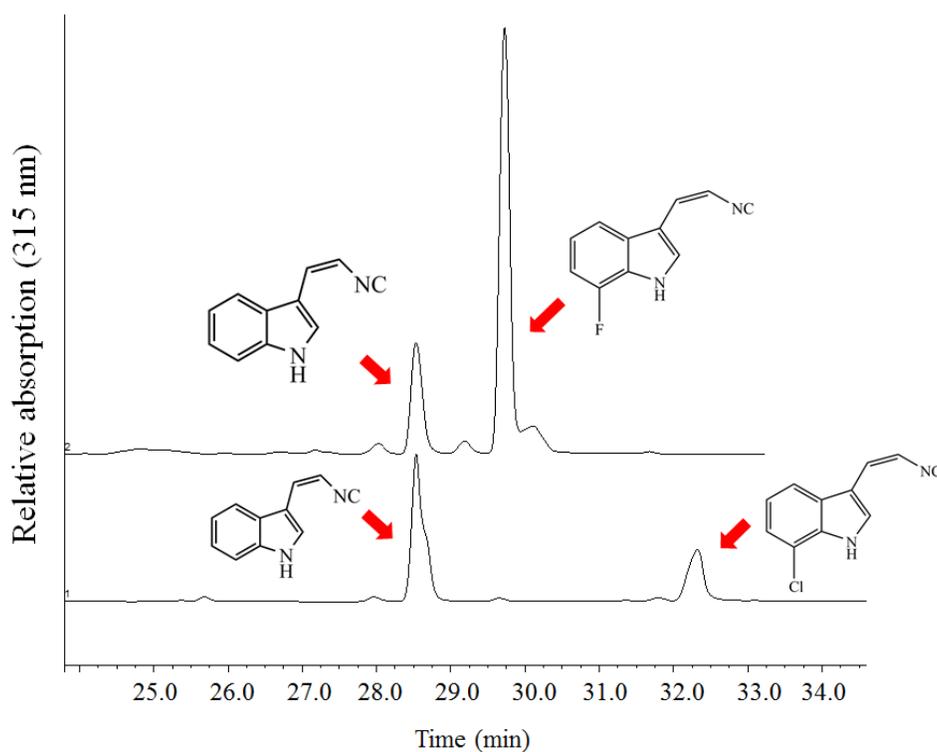
Figure 28. Competitive *in vivo* feeding experiment to compare the natural with halogenated substrates

3.4   CONCLUSIONS

In conclusion, AmbI1-I3 enzymes are considerably promiscuous as these three enzymes have the ability to incorporate unnatural halogenated tryptophans as the substrate as opposed to their natural L-tryptophan substrate. The result from the *in vitro* and *in vivo* investigation of AmbI1-I3 promiscuity demonstrated that as the size of the substituted halogens increased, the efficiency of

the enzymes to generate respective halogenated *(Z)*-**1** reduced. In contrast to the effect of the size of substituted halogens, the position of halogens on the indole ring of tryptophans had no significant effect.

Provided that *(Z)*-**1** is a universal precursor of all hapalindole-type indole alkaloids, the ability of AmbI1-I3 to incorporate halogenated tryptophans to provide halogenated analogs of *(Z)*-**1** can have profound effects on biological activities of this family of compounds such as enhanced pharmacological potentials. The future application of this part of the work is to investigate the ability of stigonematalean cyanobacteria to incorporate halogenated tryptophans to make hapalindole-type indole alkaloid analogs.

4.0   EXPERIMENTAL

General methods. All PCRs were carried out on a C1000 thermal cycler (Bio-Rad). DNA sequencing was performed by Elim BioPharm Inc. Preparative-scale reverse-phase HPLC was performed using a Dionex instrument equipped with a 21 × 250 mm Luna C18 column (Phenomenex). Analytical reverse-phase HPLC was performed using a Dionex UHPLC with a photo-diode array UV/Vis detector (Thermo Fisher Scientific) and a 4.6 × 250 mm Luna C18 column (Phenomenex). HRMS analysis was conducted using a Q Exactive Benchtop Quadrupole-Orbitrap mass spectrometer (Thermo Fisher Scientific) equipped with a Dionex RSLC (Thermo Fisher Scientific). The NMR spectrum was recorded on a Bruker Avance III 400 MHz and 600 MHz spectrometer equipped with ^{1}H/BBF observe probe. The concentration of DNA was measured by a Thermo Scientific NanoDrop 2000 Spectrophotometer. UV spectrometry was carried out using Varian Cary 50 Bio UV/ Visible Spectrophotometer.

Materials. Synthetic oligonucleotides for gene amplification by PCR were purchased from Life Technologies or Integrated DNA Technology. Kappa HiFi DNA polymerase was obtained from Kappa Biosystems. Restriction endonucleases, T4 DNA ligase and Antarctic phosphatase were purchased from New England BioLabs. LB broth and agar used for culturing *E. coli* were obtained from Teknova. YPD broth (1% Bacto-yeast extract, 2% Bacto-peptone and 2% dextrose) was purchased from USBiological. Yeast Nitrogen Base (without amino acids) was purchased from Research Products International Corp. SC-Ura drop-out mix was purchased from USBiological. All

other reagents, including inorganic salts and solvents, were purchased from Sigma-Aldrich or Fisher Scientific unless otherwise stated. All halotryptophans were obtained from Rebecca Goss' group.

Strains and plasmids. *E. coli* TOP10 cell (Life Technologies) was used for routine cloning and plasmid propagation. *E. coli* BL21(DE3) cell (Life Technologies) was used for protein expression. *Saccharomyces cerevisiae* (INVSc1) yeast strain obtained from Invitrogen was used in recombineering. pQTEV cloning plasmid was obtained from Addgene. pMQ123i and pMQ131 plasmids were obtained from Shanks' lab.

General methods for yeast recombineering. Vectors to be modified (pMQ123i and pMQ131) were cut with SmaI restriction enzyme and dephosphorylated by Antarctic phosphatase. Primers were designed to PCR amplify DNA to be joined with the linearized vectors. *S. cerevisiae* (INVSc1) was grown overnight in yeast extract-peptone-dextrose (YPD) at 30°C and transformed using a modified "lazy bones" transformation protocol where 20 to 200 ng of the dephosphorylated linearized vector and 50 to 500 ng of the amplicon were added [29]. Recombinants were selected on uracil dropout plates [26]. Plasmids were subsequently recovered from yeast using the "smash and grab" method. For screening, the isolated plasmids were transformed into *E. coli*, where kanamycin (50 μg/ml) and gentamicin (10 μg/ml) were used for selection and propagation of pMQ131 and pMQ123i constructs, respectively [26, 36-38].

Gene cloning for heterologous expression. All *amb*I1-I3 genes were PCR amplified by using primers listed in Table 4. The template containing amb*I1-I3* used for PCR was provided by Dr. Matthew Hillwig, which he previously used in ambiguine gene cluster studies. The genes were recombined into appropriate vectors as shown in Table 5 by yeast homologous recombination approach as discussed in general methods for yeast recombineering. Plasmid DNA was isolated

from selected colonies and digested with appropriate restriction enzymes to confirm positive

clones. The detail regarding the digestive verification is shown in Table 3 and the results of

digestive analyses are depicted in Figure 29-30.

Number	Construct	Enzyme	Expected products (bp)
1	*amb*I1-I3 pMQ123i	EcoRV	1248, 1481, 2833, 6156
2	*amb*I1-I3 pMQ123i (200 bp *amb*I2 truncated)	EcoRV	1248, 1481, 2833, 5956
3	*amb*I1-I3 pMQ123i (500 bp *amb*I2 truncated)	EcoRV	1248, 1481, 2833, 5656
4	*amb*I2 pMQ131	EcoRV + EcoRI	865, 1009, 2408, 2943
5	*amb*I1 pMQ131	EcoRV	862, 1035, 2367, 2951
6	*amb*I3 pMQ131	EcoRV + HindIII	121, 865, 1026, 2351, 2683
7	*amb*I1-I2 pMQ123i	EcoRV	1255, 1463, 2826, 5256
8	*amb*I2-I3 pMQ123i	EcoRV	1240, 2818, 6596
9	*amb*I1/I3 pMQ123i	EcoRV	1260, 1484, 2815, 5166

Table 3. Digestive verification of successful construction plasmids

Primer No.	Primer name	Sequence	Description
1	ambI1-5'-connect123i	GAA TTG GAT CCT CTA GAT TCT CCA TAC AGG AGG AAT AAT ATG ATT AGT GAA AAA ATT CTC	*ambI1*_5' to pMQ123i
2	ambI1-3'-connectI2	CAT TTA TAT CCT CCT ACG GGT ATG GAG AAC TAA CTC TTG TTG TCA AG	*ambI1*_3' to *ambI2*
3	ambI3-3'-connect123i	TCA GAC CGC TTC TGC GTT CTG ATT TAT TAT AAA ATA TGT ACC CGT TGC	*ambI3*_3' to pMQ123i
4	ambI2-993bp-5'-connectI1	TTC TCC ATA CCC GTA GGA GGA TAT AAA TGA CTC AAA TTA TCA ATA TCA C	*ambI2*_5' to *ambI1*
5	ambI2-753bp-5'-connectI1	TTC TCC ATA CCC GTA GGA GGA TAT AAA TGC CAG ATA TGG GAG AAC	Truncated *ambI2*_5' to *ambI1*
6	ambI2-414bp-5'-connectI1	TTC TCC ATA CCC GTA GGA GGA TAT AAA TGT TCA ACG GTA TTC ATC	Truncated *ambI2*_5' to *ambI1*
7	ambI2-5'-pMQ123i	GAA TTG GAT CCT CTA GAT TCT CCA TAC AGG AGG AAT AAT ATG ACT CAA ATT ATC AAT ATC	*ambI2*_5' to pMQ123i
8	ambI3-5'-131	GCG GAT AAC AAT TTC ACA CAG GAA ACA GCT ATG ATA GTT TCT ACT TCT GTT GAG	*ambI3*_5' to pMQ131
9	ambI1-3'-131	TTA TCA GAC CGC TTC TGC GTT CTG ATT TAC TAA CTC TTG TTG TCA AGG GTT AC	*ambI1*_3' to pMQ131
10	ambI1-5'-131	GCG GAT AAC AAT TTC ACA CAG GAA ACA GCT ATG ATT AGT GAA AAA ATT CTC	ambI1_5' to pMQ131
11	ambI2-3'-pMQ131	CAG ACC GCT TCT GCG TTC TGA TTT AAA TAA CCT CCT GAA C	ambI2_3' to pMQ131

Table 4. Primer sequences

Construct No.	Primer pairs	Description
1	1+2, 3+4	*amb*I1-I3 pMQ123i
2	1+2, 3+5	*amb*I1-I3 pMQ123i (200 bp *amb*I2 truncated)
3	1+2, 3+6	*amb*I1-I3 pMQ123i (500 bp *amb*I2 truncated)
4	7+11	*amb*I2 pMQ131
5	9+10	*amb*I1 pMQ131
6	3+8	*amb*I3 pMQ131
7	1+2, 4+11	*amb*I1-I2 pMQ123i
8	3+7	*amb*I2-I3 pMQ123i
9	3+8, 1+2	*amb*I1/I3 pMQ123i

Table 5. Cloned constructs used for truncation and complementation experiments
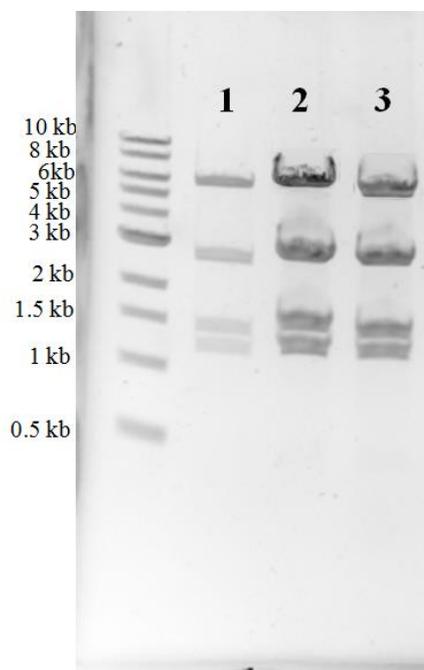
Figure 29. The successful construction plasmids for truncation experiment of construct 1, the expected bands are 1248, 1481, 2833, 6156 bp (lane 1), construct 2, the expected bands are 1248, 1481, 2833, 5956 bp (lane 2) and construct 3, the expected bands are 1248, 1481, 2833, 5656 bp (lane 3) as verified by digestive analysis by EcoRV. All plasmids were sequenced to verify the constructs.
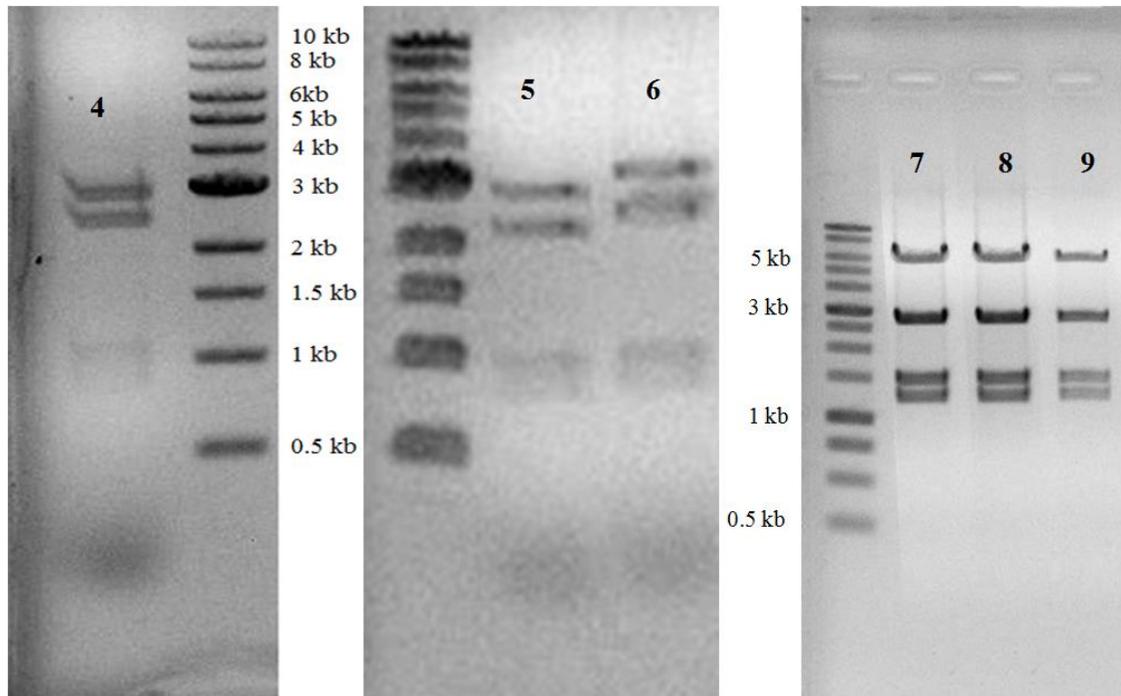
Figure 30. The successful construction plasmids for complementation experiment, lane 4 (construct 4): *amb*I2 in pMQ131 verified by EcoRV and EcoRI restriction enzyme digest analysis. The expected bands are 865, 1009, 2408, 2943 bp, lane 5 (construct 5): *amb*I1 in pMQ131 verified by EcoRV. The expected bands are 862, 1035, 2367, 2951 bp, lane 6 (construct 6): *amb*I3 in pMQ131 verified by EcoRV and HindIII. The expected bands are 121, 865, 1026, 2351, 2683 bp, lane 7 (construct 7): *amb*I1-I2 in pMQ123i, the expected bands are 1255, 1463, 2826, 5256 bp, lane 8 (construct 8): *amb*I2-I3 in pMQ123i, the expected bands are 1240, 2818, 6596 bp, lane 9 (construct 9): *amb*I1/I3 in pMQ123i, the expected bands are 1260, 1484, 2815, 5166 bp verified by EcoRV

Heterologous expression and extraction for AmbI2 role investigation. Successfully constructed plasmids (Table 5) were transformed into *E. coli* Top 10. Individual colony of *E. coli* Top 10 was picked and inoculated into 3 mL cultures overnight, which was grown in 50 mL LB at 37°C (where appropriate, antibiotics were added to the following final concentration: gentamycin 10 μg/ml; kanamycin 50 μg/ml) until OD_{600} reached 0.4-0.6. Then it was induced for 2 days at 25°C by the addition of isopropylthio-β-galactoside (IPTG) to a final concentration of 1 mM. To extract, the culture was centrifuged and the supernatant was transferred into a separating

44

funnel, mixed with an equal volume of ethyl acetate, and shaken vigorously. The mixture was then allowed to stand for the layers to be separated. The extraction was repeated twice. The organic layer was dried with Na_2SO_4, filtered and evaporated to dryness. The crude extract was centrifuged to remove particles, dissolved in methanol and analyzed by HPLC and LC/HRMS. A gradient from 5-70% acetonitrile in 35 min was used for the mobile phase. For quantification of the indole isonitrile, an indole isonitrile standard curve was constructed by using the synthesized compound of the following concentrations: 1, 4, 6, 10, 20 and 40 µg in relation with area under the peak (mAU*min).

Heterologous expression and extraction for unnatural substrates. Individual colony of *E. coli* Top 10 containing *amb*I1-I3 pMQ123i plasmid was picked and inoculated into a 3 mL LB culture overnight which was grown in 50-100 mL M9 minimal media (2 mM Mg_2SO_4, 0.4% glucose, 0.1 mM $CaCl_2$, 1x M9 salts containing 12.65 mM $Na_2HPO_4.7H_2O$, 22.04 mM KH_2PO_4, 8.56 mM NaCl, 18.6 mM NH_4Cl)) initially supplemented with 50 µM halogenated tryptophan at 37°C overnight until OD_{600} reached 0.7-0.9. Then it was supplemented with additional 200 µM halogenated tryptophan and induced for 2 days by the addition of isopropylthio-β-galactoside (IPTG) to a final concentration of 1 mM. The extraction, HPLC, LC/HRMS analyses and quantification were carried out in an identical fashion as described in heterologous expression and extraction for AmbI2 role investigation.

Protein expression. *amb*I1, *amb*I2 and *amb*I3 were previously cloned into pQTev [11] and transformed into the BL21(DE3) Star protein expression *E. coli* strain. Protein purification was carried out in an identical fashion as described in the previous publication [11]. The eluted protein was dialyzed (20 mM HEPES, 10 mM NaCl, 10% glycerol, 0.5 mM DTT) in a 10 kDa molecular weight cutoff membrane (Spectrum Laboratory Products, Inc., Gardena, CA) to

exchange buffer. The purified protein was flash-frozen in liquid nitrogen and stored in -80°C. The averaged protein yields purified from 1-liter pellet of AmbI1, AmbI2 and AmbI3 were 13 mg, 18 mg, and 20 mg, respectively. The purified protein was analysed by SDS-PAGE to ensure homogeneity (Figure 31).
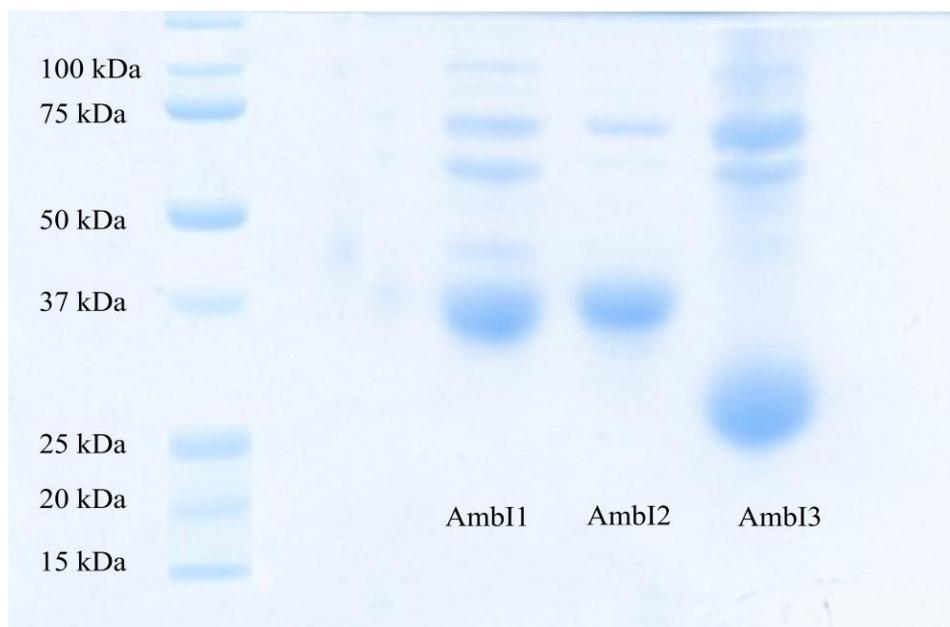
Figure 31. SDS-PAGE gel images of purfied NHis-tagged AmbI1-3 proteins

Isonitrile synthase assay. The assay was carried out in a similar fashion as described in the previous publication [11]. 1 mL assay was set up in a buffer containing 25 mM HEPES, 150 mM NaCl, 5% glycerol (pH 7.4). The substrates added included 2.5 mM L-tryptophan (or halogenated tryptophans), 2.5 mM ribose-5-phosphate, 2 mM α-KG and 20 μM of each enzyme. The assays were incubated at 30°C for 4 hours and extracted twice with an equal volume of ethyl acetate and the organic solvent was separated and evaporated to dryness with nitrogen gas, dissolved in 60 μL of methanol. For the analysis, 50 μL was injected into HPLC. The HPLC and LC/HRMS analyses were carried out in an identical fashion as described in heterologous

46

expression and extraction for AmbI2 role investigation. The UV spectra obtained from HPLC are shown in Figure 34. The HRMS data are shown in Figure 35. Synthetic standards, including *(Z)-* **1**, *(Z)*-7 chloro indole isonitrile and *(Z)*-6 fluoro indole isonitrile, used to compare with the assay product were obtained from Dr. Qin Zhu. ^{1}H-NMR (600 MHz, CDCl$_3$) of these synthetic standards are shown in Figure 33. ^{19}F-NMR (400 MHz, CDCl$_3$) of *(Z)*-6 fluoro indole isonitrile is shown in Figure 32. For quantification of the 6-F indole isonitrile, a 6-F indole isonitrile standard curve was constructed by using the synthesized compound of the following concentrations: 1, 4, 6, 10, 20 and 40 µg in relation with area under the peak (mAU*min).
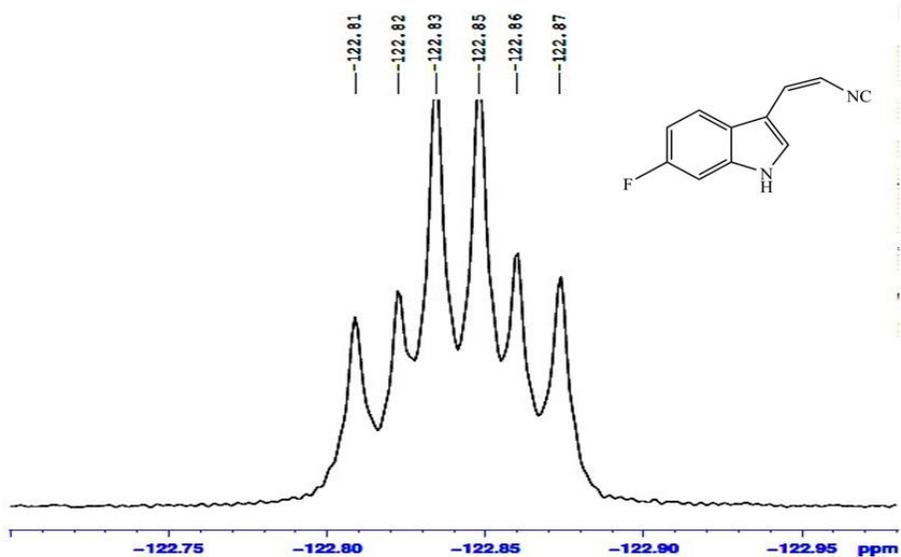
Figure 32. ^{19}F-NMR (400 MHz, CDCl$_3$) of synthetic *(Z)*-6-fluoro indole isonitrile (obtained from Dr. Qin Zhu)

47

(Z) indole isonitrile

6-F indole isonitrile

7-Cl indole isonitrile
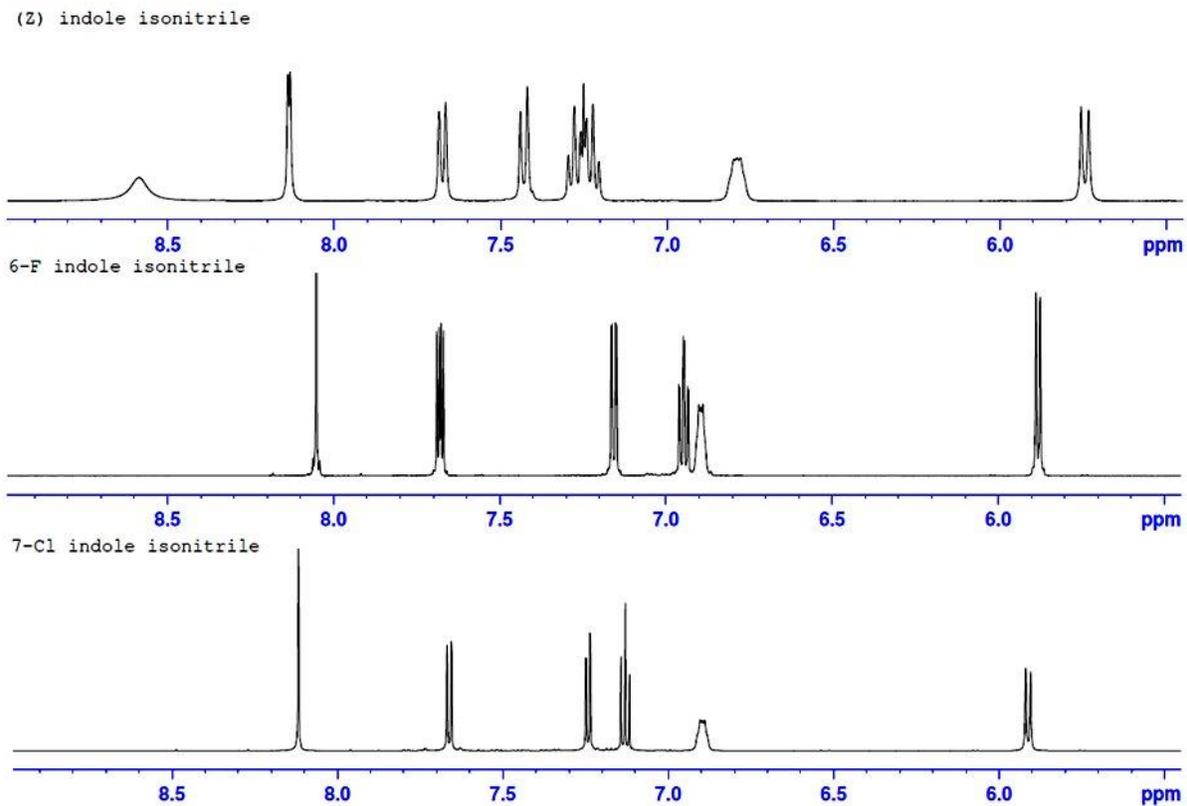
Figure 33. Overlay [^1]H-NMR (600 MHz, CDCl$_3$) of synthetic *(Z)*-**1**, *(Z)*-6-fluoro indole isonitrile, *(Z)*-7-chloro

indole isonitrile (obtained from Dr. Qin Zhu)

Peak #12 -50% at 28.65 min

224.6

315.1

39

Peak #15 -50% at 29.85 min

223.2

291.8 312.5

Peak #15   -50% at 29.53 min

223.5

303.9

Peak #11   -50% at 32.34 min

229.7

302.9

Peak #10   -50% at 33.13 min

229.2

301.1

Peak #11   -50% at 29.85 min

220.1

311.5

Peak #14   -50% at 32.30 min

228.9

309.3

Peak #17   -50% at 33.02 min

224.9

310.6

52
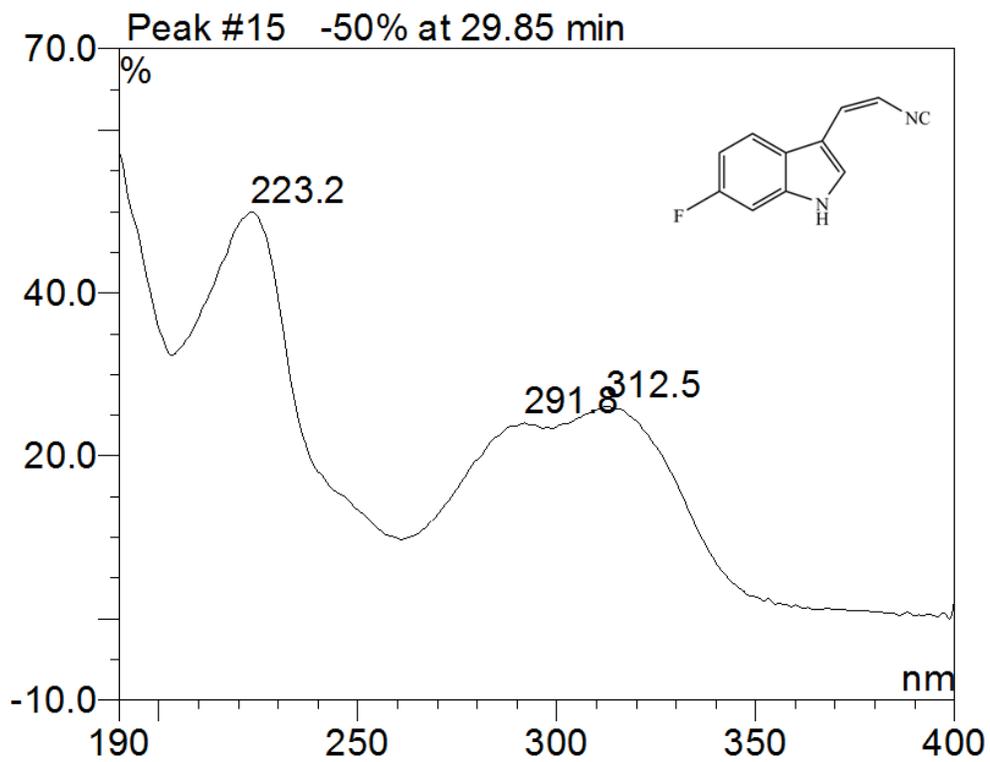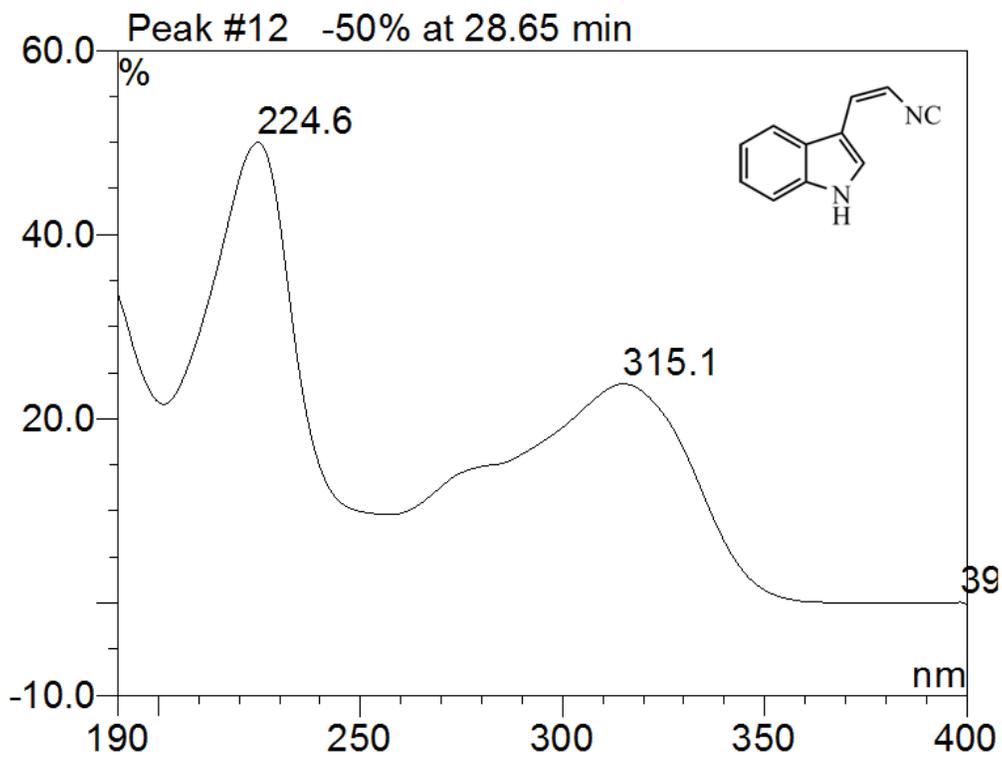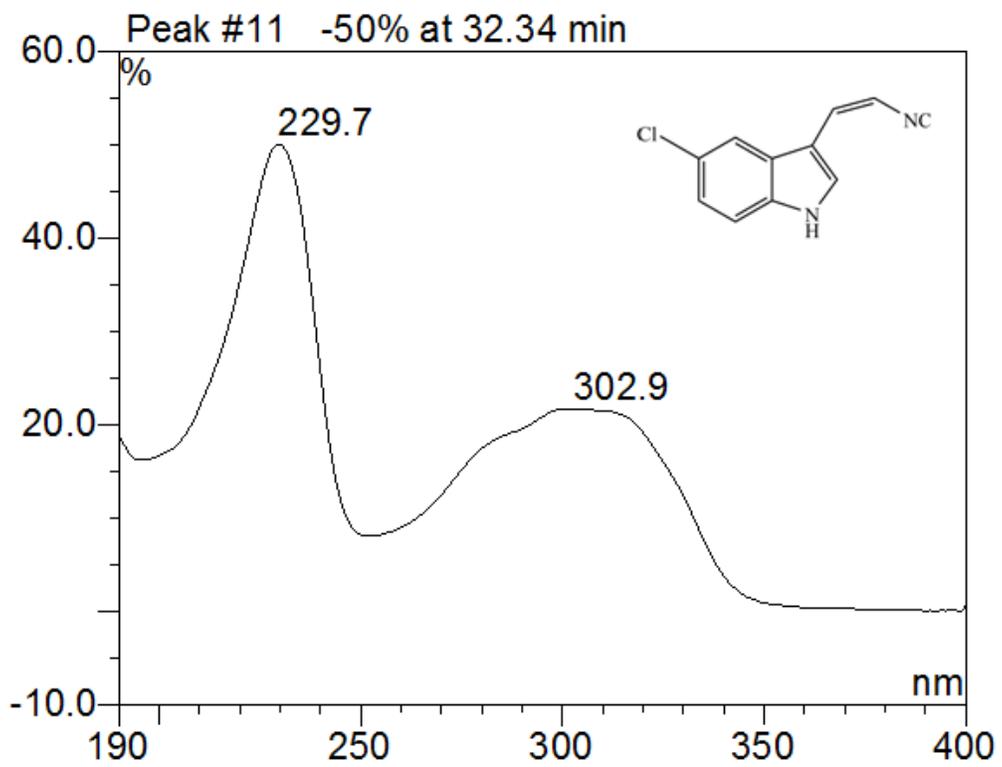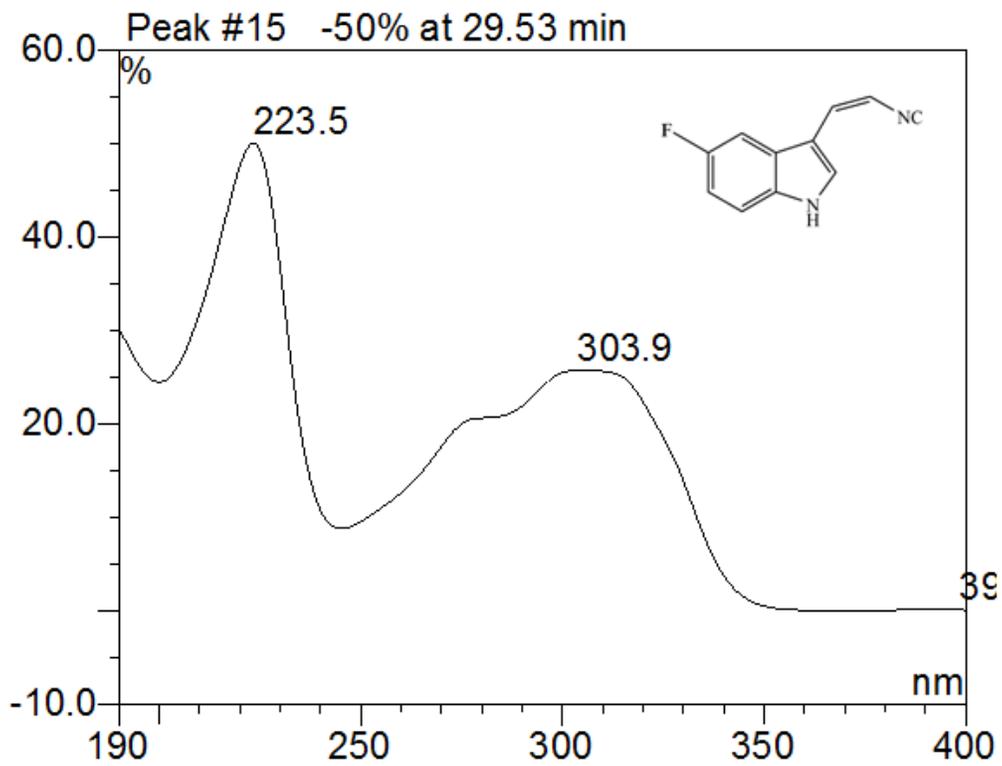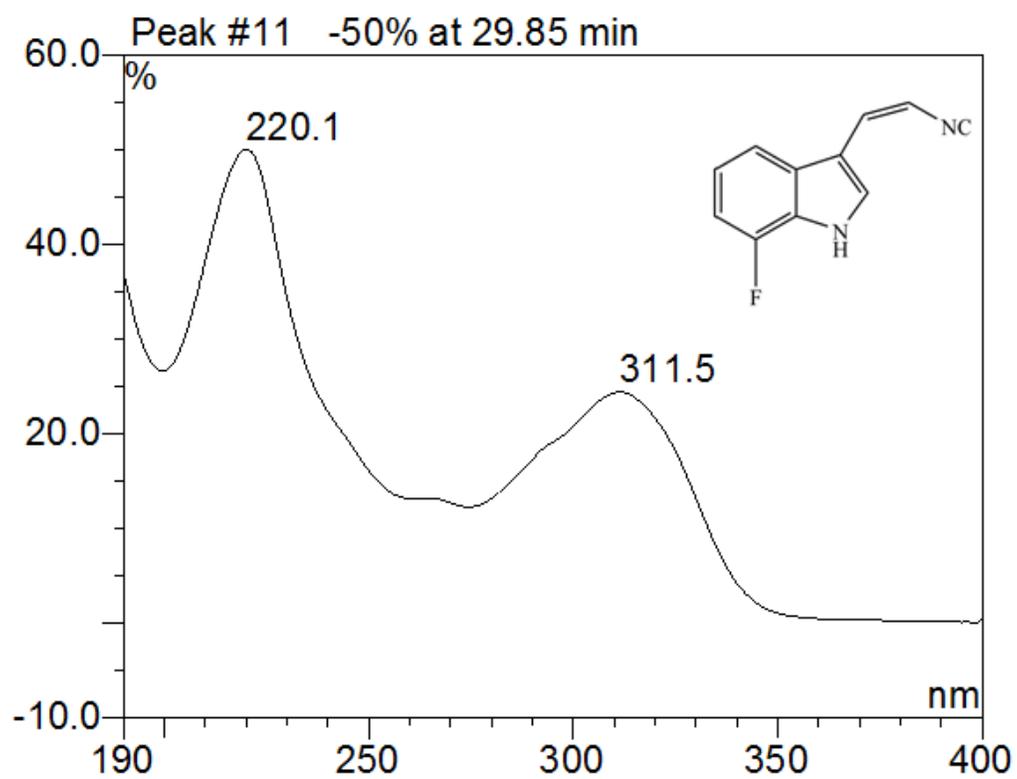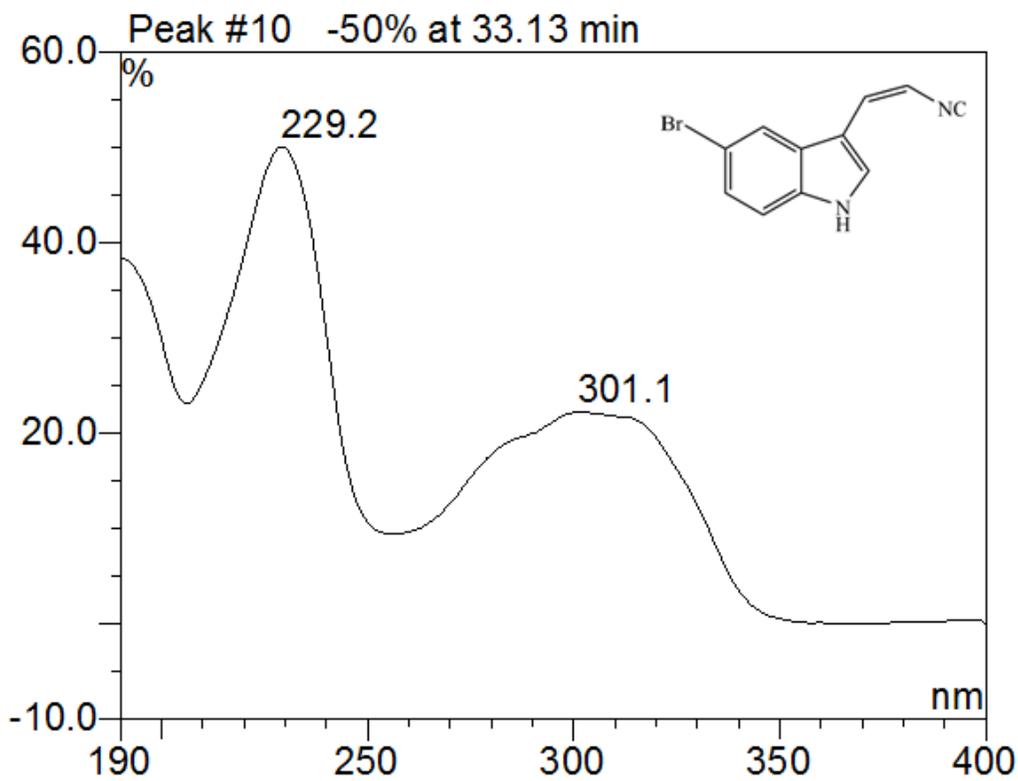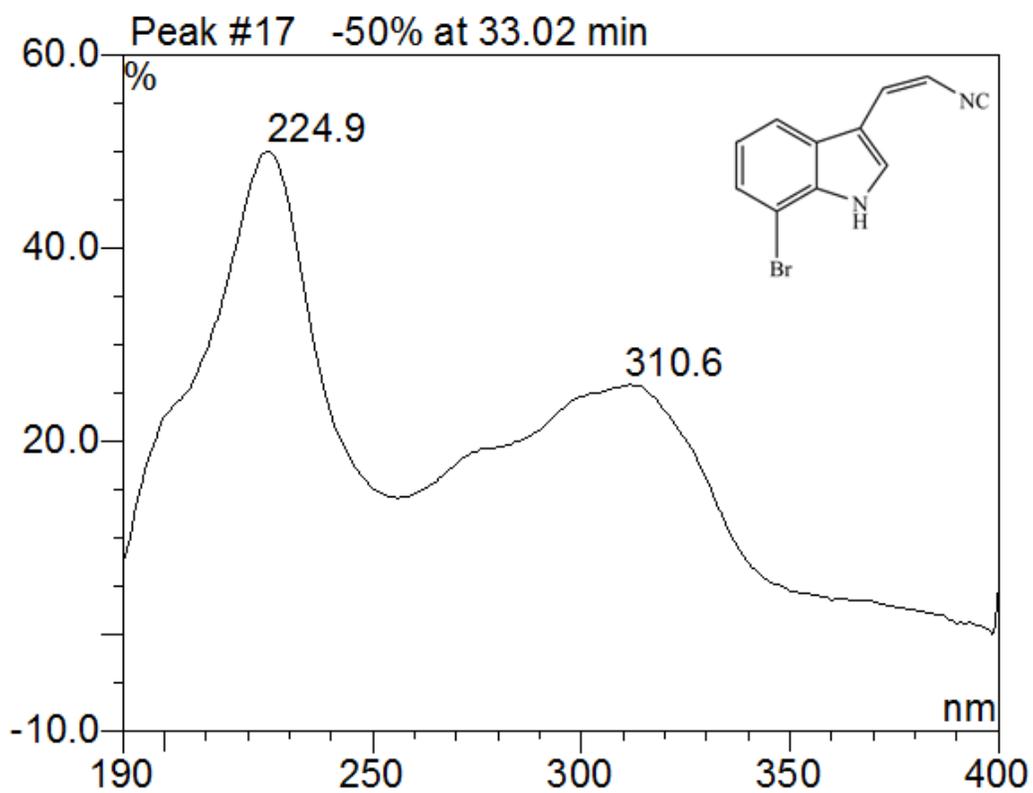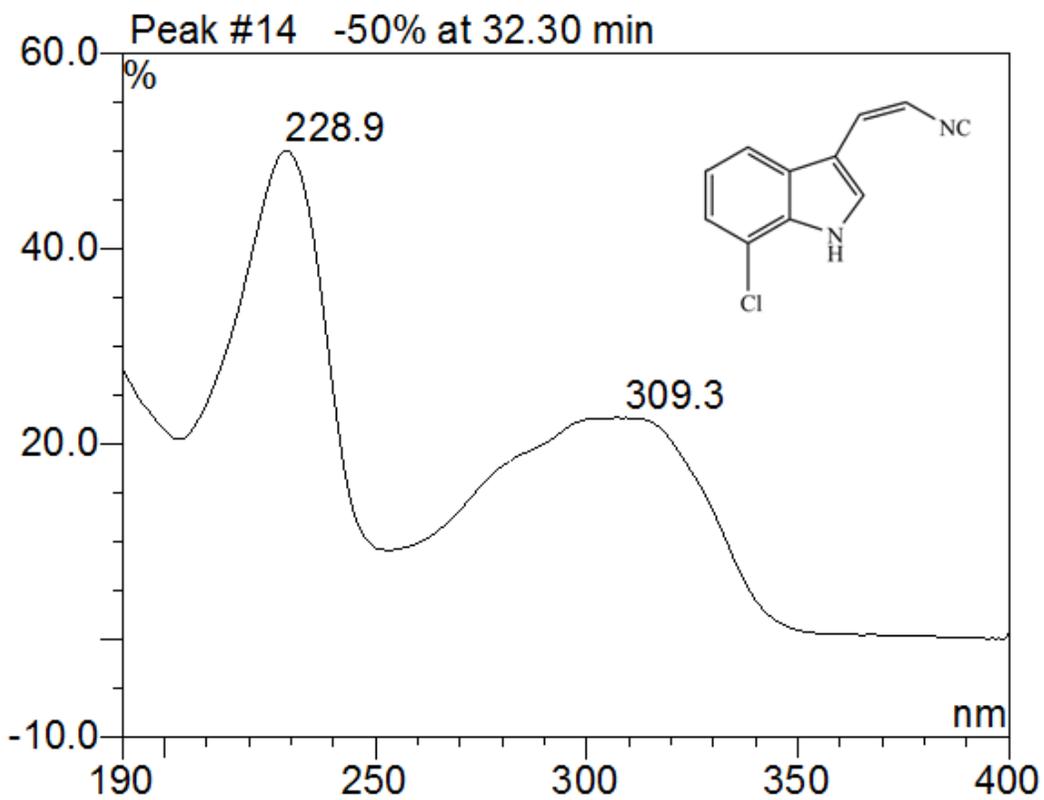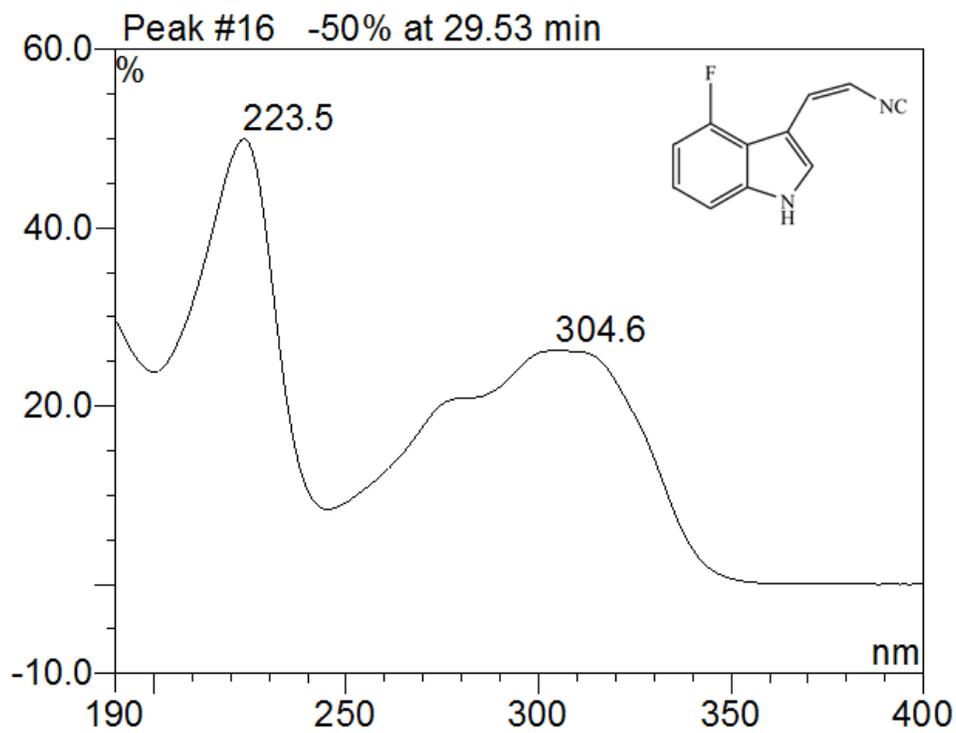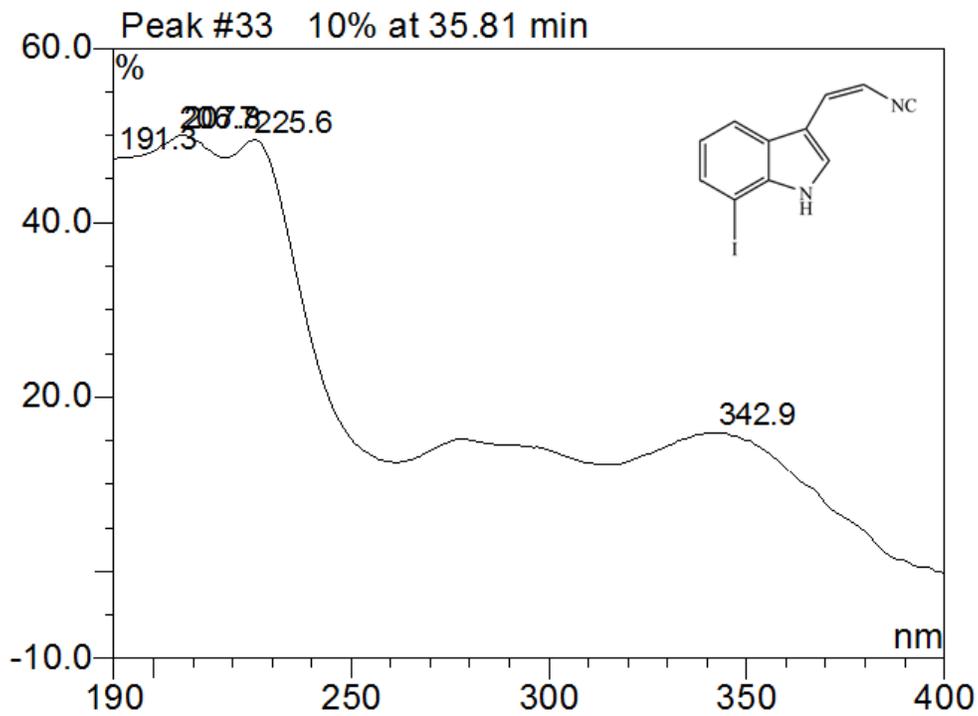
Figure 34. UV spectra from HPLC of *(Z)*-**1**, 6-F indole isonitrile, 5-F indole isonitrile, 5-Cl indole isonitrile, 5-Br indole isonitrile, 7-F indole isonitrile, 7-Cl indole isonitrile, 7-Br indole isonitrile, 7-I indole isonitrile, and 4-F indole isonitrile

72727LCMSP#4646-4770 RT: 24.99-25.66 AV: 125
T: FTMS + p ESI Full ms [100.00-1000.00]
m/z= 169.00000-170.00000

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
169.07564	199056016.0	100.00	169.07602	-0.38	$C_{11}H_9N_2$

(Z)-1
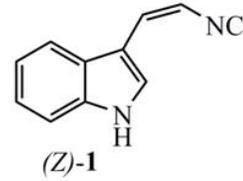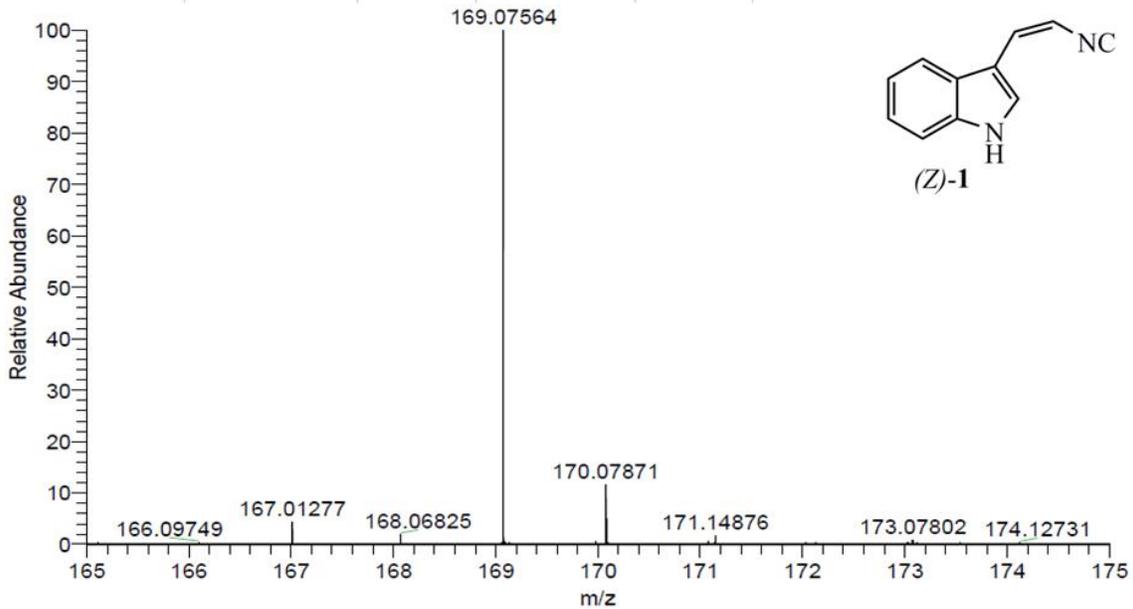
72675LCMSP#4845-4900 RT: 26.19-26.49 AV: 56
T: FTMS + p ESI Full ms [100.00-1000.00]
m/z= 187.00000-188.00000

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
187.06712	17003176.0	100.00	187.06660	0.51	$C_{11}H_8N_2F$
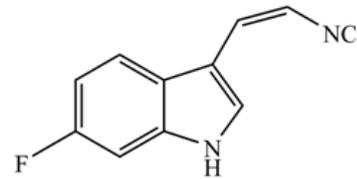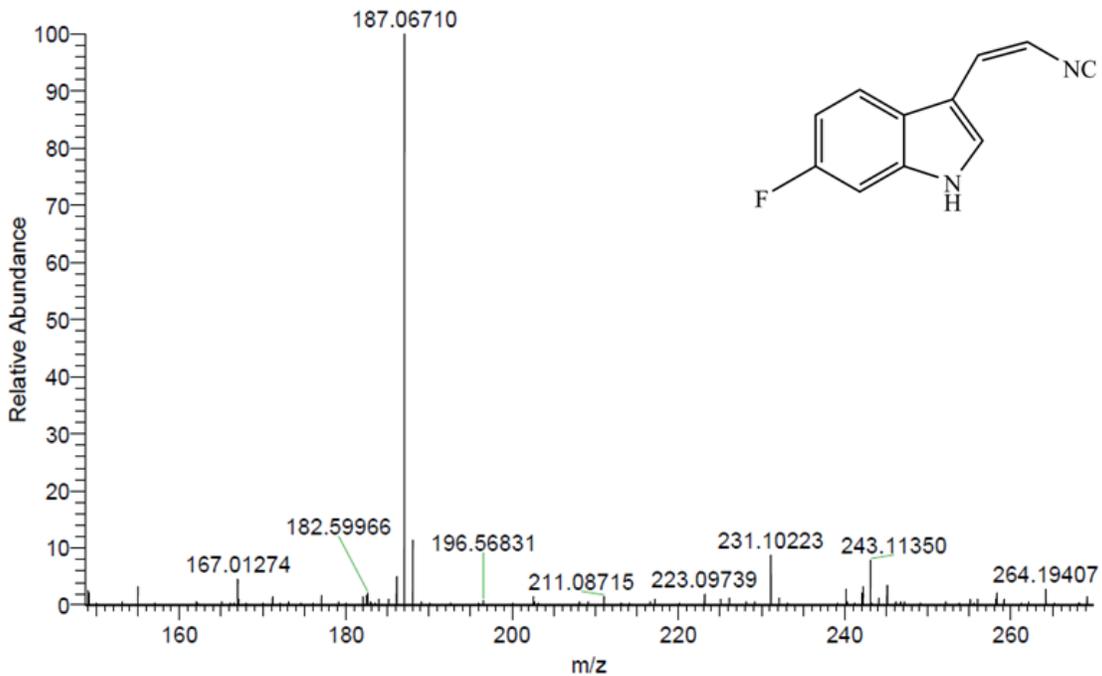
726651cmsp2#4775-4849 RT: 26.09-26.49 AV: 75
T: FTMS + p ESI Full ms [100.00-1000.00]
m/z= 187.00000-188.00000

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
187.06677	7089549.0	100.00	187.06660	0.91	$C_{11}H_8N_2F$
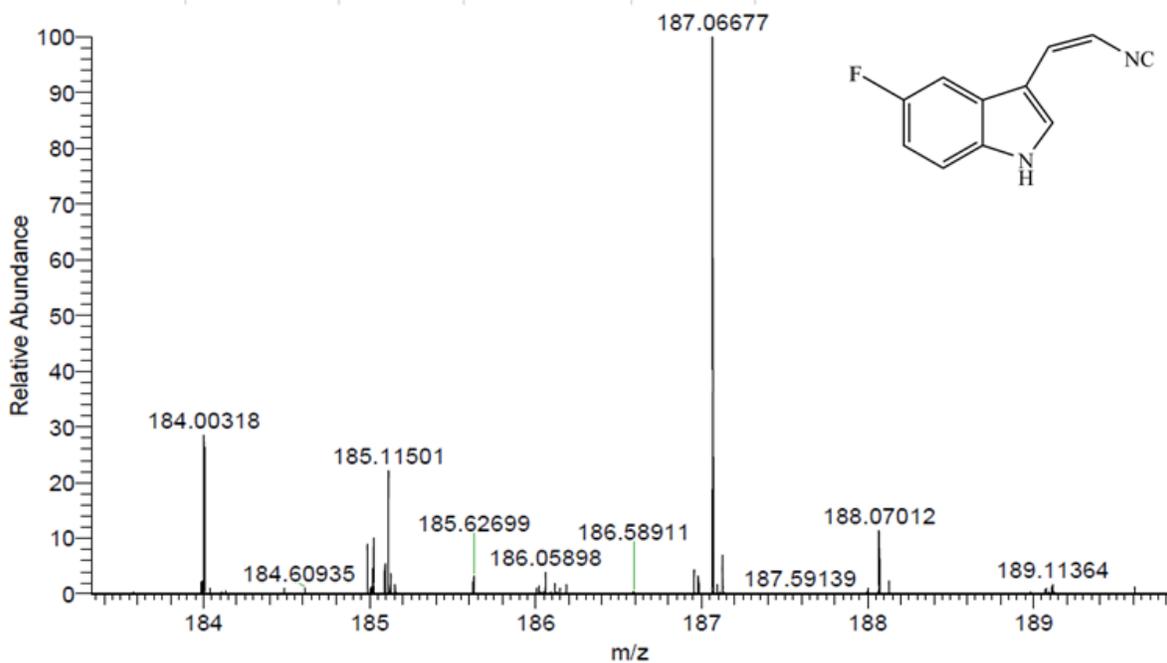
726661cmsp2#5199-5293 RT: 28.37-28.88 AV: 95
T: FTMS + p ESI Full ms [100.00-1000.00]
m/z= 203.00000-204.00000

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
203.03734	1817675.4	100.00	203.03705	1.41	$C_{11}H_8N_2Cl$
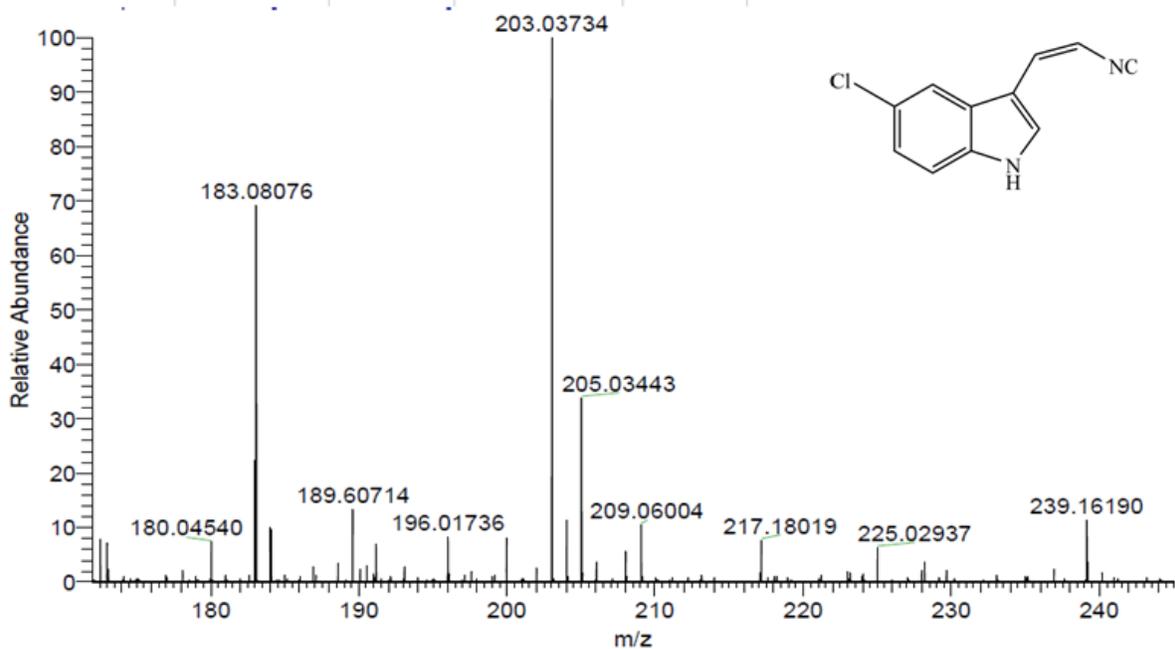
72667LCMSP2#5362-5397 RT: 29.27-29.46 AV: 36
T: FTMS + p ESI Full ms [100.00-1000.00]
m/z= 246.90000-247.00000

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
246.98691	2038514.5	100.00	246.98654	0.38	$C_{11}H_8N_2Br$
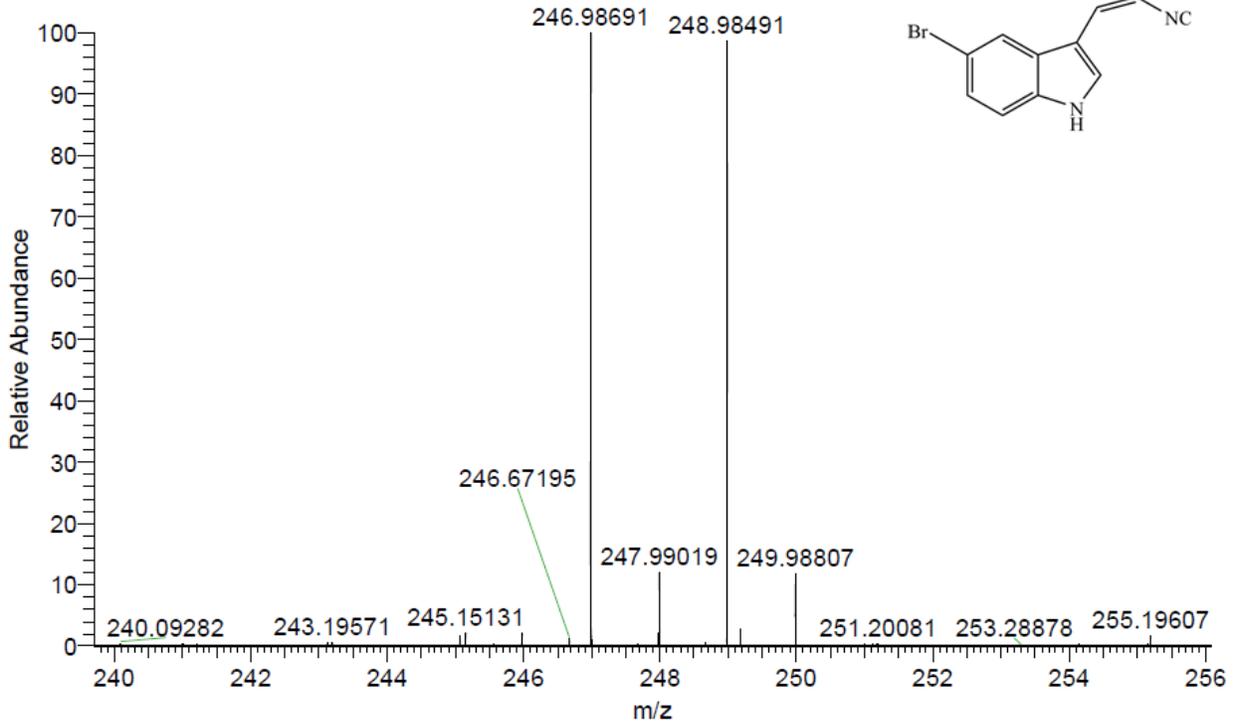
726681cmsp2#4826-4908 RT: 26.36-26.80 AV: 83
T: FTMS + p ESI Full ms [100.00-1000.00]
m/z= 187.00000-187.10000

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
187.06702	3516525.8	100.00	187.06660	2.23	$C_{11}H_8N_2F$

56

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
203.03760	1885703.3	100.00	203.03707	0.52	$C_{11}H_8N_2Cl$

m/z	Intensity	Relative	Theo. Mass	Delta (ppm)	Composition
246.98598	28958042.0	100.00	246.98654	-0.56	$C_{11}H_8N_2Br$

Figure 35. HRMS of *(Z)*-**1**, 6-F indole isonitrile, 5-F indole isonitrile, 5-Cl indole isonitrile, 5-Br indole isonitrile, 7-F indole isonitrile, 7-Cl indole isonitrile, 7-Br indole isonitrile, 7-I indole isonitrile, and 4-F indole isonitrile

APPENDIX A

IDENTIFICATION AND CHARACTERIZATION OF A WELWITINDOLINONEALKALOID BIOSYNTHETIC GENE CLUSTER IN THE STIGONEMATALEAN CYANOBACTERIUM HAPALOSIPHON WELWITSCHII

Matthew L. Hillwig, Heather A. Fuhrman, Kuljira Ittiamornkul, Tyler J. Sevco, Daniel H. Kwak, and Xinyu Liu*

Shortly after the ambiguine biosynthetic gene cluster in *F. ambigua* UTEX1903 had been disclosed, our group initiated an effort to study another hapalindole-type indole alkaloids, welwitindolinone gene cluster in *H. welwitschii* UTEX B1830. Welwitindolinone share identical terpenoid backbone stereochemistry with ambiguines at C10-C11, but inverted stereochemistry across C11-C12. This study resulted in characterization of the enzymes responsible for assembling the biosynthetic precursors including the assembly of geranyl pyrophosphate (GPP) by WelD1-4, WelP2 and *(Z)*-**1** by WelT1-5, WelI1-I3, which is in agreement with the ambiguine pathway as the protein sequence identities are higher than 90%. In addition, comparative analysis of the ambiguine and welwitindolinone biosynthetic pathways showed that WelP1 and AmbP1, an aromatic prenyltransferase, share nearly identical sequences which suggested the common

role of P1 as the enzyme responsible for monogeranylation of GPP to *(Z)*-**1** to generate a common intermediate for the hapalindole-type alkaloid biosynthesis.

Five non-heme iron (NHI)-dependent oxygenases, that are responsible for structure diversification, were identified in ambiguine pathway including four highly homologous Rieske-type oxygenases (WelO1-O4) and an iron(II)/α-ketoglutarate-dependent oxygenase (WelO5). WelO5 is responsible for stereoselective chlorination at the C-13 of hapalindole/fischerindole backbones. WelO1-4 are responsible for oxidative transformations of fischerindoles to welwitindolinones. Comparing with those in the *amb* pathway, the protein sequence homologies are 61-79%, which is significantly lower than the protein sequence homologies of the rest of *wel/amb* pathway [21].

These findings collectively addressed the question about the origin of stereochemical diversity that the structural diversification of hapalindole-type alkaloids does not arise from biosynthetic precursors as proposed in the literature, but from later-stage modification by non-heme iron oxygenases [21].

A.1 *H. WELWITSCHII* UTEX B1830 GENOME SEQUENCING AND ASSEMBLY

Hapalosiphon welwitschii UTEX B1830 used for welwitindolinone biosynthesis study is a xenic strain that is available in the public domain. To carry out genome sequencing, the genomic DNA was extracted and de novo sequenced and assembled by using a Roche 454 GS FLX+ Titanium sequencer system to obtain the draft assembly in nearly 10,000 contigs with the size of approximately 15 Mb which is accordance with its xenic status. Nucleotide BLAST search was implemented using genes in the *amb* pathway as bioinformatic leads which resulted in the

identification of 11 contigs, including a 21 kb contig that spanned a large portion of *wel* gene cluster from *welR3* to *welC1* genes, and ten small contigs with an average size of 1-2 kb which lack homologous end-joining sequences to accommodate the assembly of the entire *wel* gene cluster. Because this region of the cyanobacterial genome is highly repetitive, gap repair was used in order to determine the direction and sequence of each contig to ultimately map out the entire *wel* biosynthetic gene cluster.

A.1.1 Gap repair

Gap repair was carried out by two approaches—both of which applied the use of cross-contig PCR amplicons to avoid the sequence-repetitive regions. The first approach applied PCR amplification and cloning of 2-3 kb fragments to pMQ124 plasmids by yeast homologous recombination followed by Sanger sequencing.

The second approach, which I contributed, was for more complicated repetitive regions. It was implemented based on the use of the main conserved sequences outside of the highly homologous Rieske oxygenase (*wel*O1-4) to amplify 4 kb regions of the gene cluster by using sequences of *wel*P2, *wel*M and conserved nucleotide sequence exclusively present in *wel*O1. The obtained PCR products were subsequently ligated into a blunt-end cloning vector followed by Sanger sequencing.

Finally, the assembly of *wel* gene cluster was carried out manually by careful inspection of homologous end-joining sequences which can be found in gap-repairing plasmids. SeqBuilder (DNAStar, Madison, WI) was used to predict protein coding genes. The functions were defined by comparing to those that can be found in *amb* gene cluster or by protein BLAST analysis.

A.1.2 Experimental

Gene cloning for gap repair. Selected genes were amplified from *F. welwitschii* UTEX B1830 genomic DNA template by Polymerase Chain Reaction (PCR). The PCR products were gel purified and ligated into blunt cloning vector pJET1.2 using the Thermo CloneJet PCT cloning kit (Thermo Scientific) and transformed to Top10 *E. coli* cells (Life Technologies) for screening by colony PCR and digestion by restriction enzymes NcoI/NotI. Positive clones were further confirmed by Sanger sequencing by Elim Biopharm (Hayward, CA).
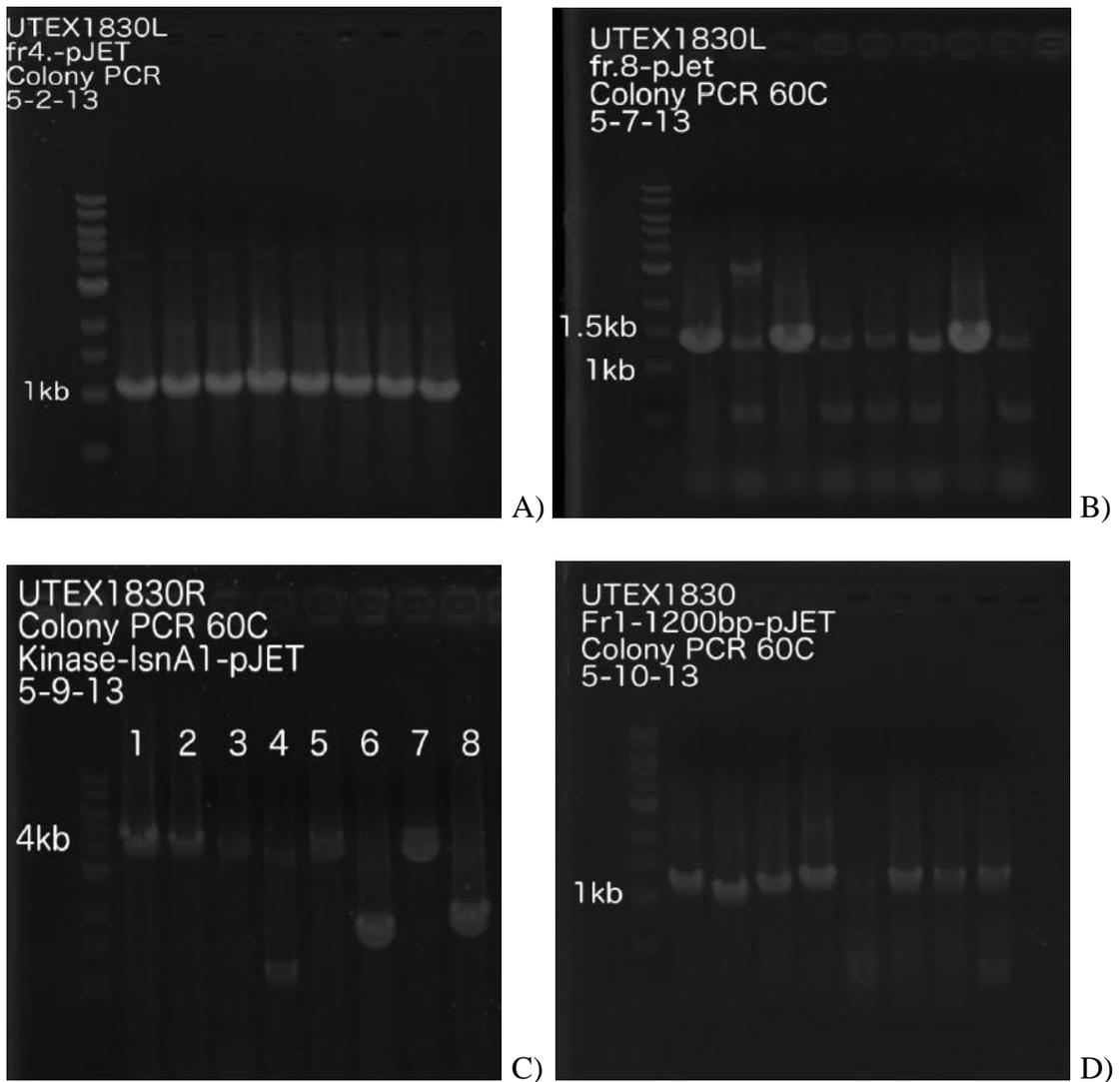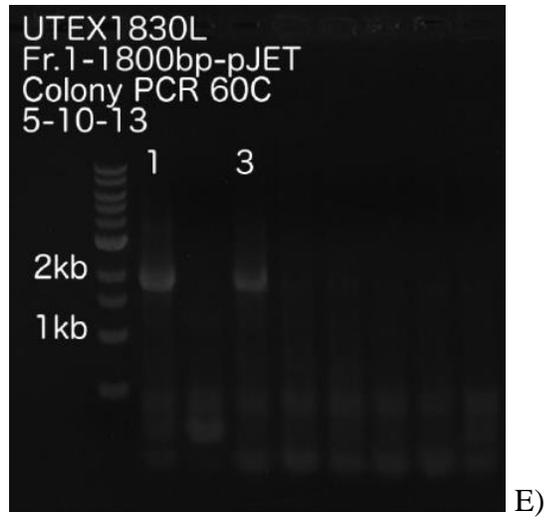
Figure 36. Gap repair pJET cloning screening for positive clones by colony PCR where positive clones have A) 1 kb insert, B) 1.5 kb insert, C) 3.5 kb insert, D) 1.2 kb insert, E) 1.8 kb insert

BIBLIOGRAPHY

1.   Kehr, J.C., D. Gatte Picchi, and E. Dittmann, *Natural product biosyntheses in cyanobacteria: A treasure trove of unique enzymes.* Beilstein J Org Chem, 2011. **7**: p. 1622-35.

2.   Chorus, I. and J. Bartram, *Toxic cyanobacteria in water: A guide to their public health consequences, monitoring and management*. 1999.

3.   Tomitani, A., et al., *The evolutionary diversification of cyanobacteria: Molecular–phylogenetic and paleontological perspectives.* Proceedings of the National Academy of Sciences, 2006. **103**(14): p. 5442-5447.

4.   Dagan, T., et al., *Genomes of Stigonematalean cyanobacteria (subsection V) and the evolution of oxygenic photosynthesis from prokaryotes to plastids.* Genome Biol Evol, 2013. **5**(1): p. 31-44.

5.   Gugger, M.F. and L. Hoffmann, *Polyphyly of true branching cyanobacteria (Stigonematales).* International Journal of Systematic and Evolutionary Microbiology, 2004. **54**(2): p. 349-357.

6.   Fran, et al., *Indole Alkaloids from Marine Sources as Potential Leads against Infectious Diseases.* BioMed Research International, 2014. **2014**: p. 12.

7.   Becher, P.G., et al., *Insecticidal activity of 12-epi-hapalindole J isonitrile.* Phytochemistry, 2007. **68**(19): p. 2493-7.

8.   Mo, S., et al., *Antimicrobial Ambiguine Isonitriles from the Cyanobacterium Fischerella ambigua.* Journal of Natural Products, 2009. **72**(5): p. 894-899.

9.   Smitka, T.A., et al., *Ambiguine isonitriles, fungicidal hapalindole-type alkaloids from three genera of blue-green algae belonging to the Stigonemataceae.* The Journal of Organic Chemistry, 1992. **57**(3): p. 857-861.

10.   Smith, C.D., et al., *Welwitindolinone analogues that reverse P-glycoprotein-mediated multiple drug resistance.* Mol Pharmacol, 1995. **47**(2): p. 241-7.

11.   Hillwig, M.L., Q. Zhu, and X. Liu, *Biosynthesis of Ambiguine Indole Alkaloids in Cyanobacterium Fischerella ambigua.* ACS Chemical Biology, 2013. **9**(2): p. 372-377.

12. Bhat, V., et al., *Chapter Two - The Chemistry of Hapalindoles, Fischerindoles, Ambiguines, and Welwitindolinones*, in *The Alkaloids: Chemistry and Biology*, K. Hans-Joachim, Editor. 2014, Academic Press. p. 65-160.

13. Hagmann, L. and F. Jüttner, *Fischerellin A, a novel photosystem-II-inhibiting allelochemical of the cyanobacterium Fischerella muscicola with antifungal and herbicidal activity.* Tetrahedron Letters, 1996. **37**(36): p. 6539-6542.

14. Wright, A.D., O. Papendorf, and G.M. König, *Ambigol C and 2,4-Dichlorobenzoic Acid, Natural Products Produced by the Terrestrial Cyanobacterium Fischerella ambigua.* Journal of Natural Products, 2005. **68**(3): p. 459-461.

15. Singh, R.K., et al., *Cyanobacteria: an emerging source for drug discovery.* J Antibiot, 2011. **64**(6): p. 401-412.

16. Shunyan Mo, a.A.K., a Bernard D. Santarsiero,a,b Scott G. Franzblau,c and Jimmy Orjalaa,*, *Hapalindole-related Alkaloids from the Cultured Cyanobacterium Fischerella ambigua.* Phytochemistry, 2010. **71**(17-18): p. 2116-2123.

17. Moore, R.E., C. Cheuk, and G.M.L. Patterson, *Hapalindoles: new alkaloids from the blue-green alga Hapalosiphon fontinalis.* Journal of the American Chemical Society, 1984. **106**(21): p. 6456-6457.

18. Stratmann, K., et al., *Welwitindolinones, Unusual Alkaloids from the Blue-Green Algae Hapalosiphon welwitschii and Westiella intricata. Relationship to Fischerindoles and Hapalinodoles.* Journal of the American Chemical Society, 1994. **116**(22): p. 9935-9942.

19. Raveh, A. and S. Carmeli, *Antimicrobial Ambiguines from the Cyanobacterium Fischerella sp. Collected in Israel.* Journal of Natural Products, 2007. **70**(2): p. 196-201.

20. Park, A., R.E. Moore, and G.M.L. Patterson, *Fischerindole L, a new isonitrile from the terrestrial blue-green alga fischerella muscicola.* Tetrahedron Letters, 1992. **33**(23): p. 3257-3260.

21. Hillwig, M.L., et al., *Identification and Characterization of a Welwitindolinone Alkaloid Biosynthetic Gene Cluster in the Stigonematalean Cyanobacterium Hapalosiphon welwitschii.* ChemBioChem, 2014. **15**(5): p. 665-669.

22. Clarke-Pearson, M.F. and S.F. Brady, *Paerucumarin, a new metabolite produced by the pvc gene cluster from Pseudomonas aeruginosa.* J Bacteriol, 2008. **190**(20): p. 6927-30.

23. Brady, S.F. and J. Clardy, *Cloning and heterologous expression of isocyanide biosynthetic genes from environmental DNA.* Angew Chem Int Ed Engl, 2005. **44**(43): p. 7063-5.

24. Hillwig, M.L. and X. Liu, *A new family of iron-dependent halogenases acts on freestanding substrates.* Nat Chem Biol, 2014. **advance online publication**.

25.   Brady, S.F. and J. Clardy, *Systematic Investigation of the Escherichia coli Metabolome for the Biosynthetic Origin of an Isocyanide Carbon Atom.* Angewandte Chemie International Edition, 2005. **44**(43): p. 7045-7048.

26.   Shanks, R.M.Q., et al., *New yeast recombineering tools for bacteria.* Plasmid, 2009. **62**(2): p. 88-97.

27.   Wood, T., *The Pentose Phosphate Pathway.* 1985, Orlando, Florida: Academic Press, Inc. 216.

28.   Jeschke, P., *The unique role of halogen substituents in the design of modern agrochemicals.* Pest Management Science, 2010. **66**(1): p. 10-27.

29.   Pauletti, P.M., et al., *Halogenated Indole Alkaloids from Marine Invertebrates.* Marine Drugs, 2010. **8**(5): p. 1526-1549.

30.   Gul, W. and M.T. Hamann, *Indole alkaloid marine natural products: An established source of cancer drug leads with considerable promise for the control of parasitic, neurological and other diseases.* Life Sciences, 2005. **78**(5): p. 442-453.

31.   Frederich, M., M. Tits, and L. Angenot, *Potential antimalarial activity of indole alkaloids.* Trans R Soc Trop Med Hyg, 2008. **102**(1): p. 11-9.

32.   Runguphan, W. and S.E. O'Connor, *Diversification of Monoterpene Indole Alkaloid Analogs through Cross-Coupling.* Organic Letters, 2013. **15**(11): p. 2850-2853.

33.   Roy, A.D., et al., *Gene Expression Enabling Synthetic Diversification of Natural Products: Chemogenetic Generation of Pacidamycin Analogs.* Journal of the American Chemical Society, 2010. **132**(35): p. 12243-12245.

34.   Runguphan, W., X. Qu, and S.E. O/'Connor, *Integrating carbon-halogen bond formation into medicinal plant metabolism.* Nature, 2010. **468**(7322): p. 461-464.

35.   Smith, D.R., et al., *The first one-pot synthesis of L-7-iodotryptophan from 7-iodoindole and serine, and an improved synthesis of other L-7-halotryptophans.* Org Lett, 2014. **16**(10): p. 2622-5.

36.   Shanks, R.M., et al., *Saccharomyces cerevisiae-based molecular tool kit for manipulation of genes from gram-negative bacteria.* Appl Environ Microbiol, 2006. **72**(7): p. 5027-36.

37.   Hoffman, C.S. and F. Winston, *A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of Escherichia coli.* Gene, 1987. **57**(2-3): p. 267-72.

38.   Sharan, S.K., et al., *Recombineering: a homologous recombination-based method of genetic engineering.* Nat. Protocols, 2009. **4**(2): p. 206-223.