# Biosynthesis of fungal indole alkaloids

<table>
<thead>
<tr>
<th>Journal:</th>
<th>Natural Product Reports</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID:</td>
<td>NP-REV-05-2014-000073.R1</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Review Article</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>02-Aug-2014</td>
</tr>
</tbody>
</table>
| Complete List of Authors: | Tang, Yi; University of California, Department of Chemical and Biomolecular Engineering  
Xu, Wei; University of California, Los Angeles, Department of Chemical and Biomolecular Engineering  
Gavia, Diego; University of California, Los Angeles, Department of Chemistry and Biochemistry |
Biosynthesis of Fungal Indole Alkaloids

Wei Xu, Diego J. Gavia, and Yi Tang*
a Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, 420 Westwood Plaza, Los Angeles, CA, 90096, USA. Email: yitang@ucla.edu
b Department of Chemistry and Biochemistry, University of California, Los Angeles, 615 Charles E Young Dr. Los Angeles, CA, 90095, USA.

Received 00th January 2012, Accepted 00th January 2012
DOI: 10.1039/x0xx00000x
www.rsc.org/npr

This review provides a summary of recent research advances in elucidating the biosynthesis of fungal indole alkaloids. Different strategies used to incorporate and derivatize the indole/indoline moieties in various families of fungal indole alkaloids will be discussed, including tryptophan-containing nonribosomal peptides and polyketide-nonribosomal peptide hybrids; and alkaloids derived from other indole building blocks. This review also includes discussion regarding the downstream modifications that generate chemical and structural diversity among indole alkaloids.

1 Introduction

Indole alkaloids, one of the largest classes of nitrogen-containing secondary metabolites, contain one or more indole/indoline moieties. They are widely found in plants, bacteria, fungi and animals. Fungi, especially Ascomycota, have been reported as prolific producers of indole alkaloids, many of which display potent biological activities. In recent years, the availability of fungal genome sequences has significantly accelerated the identification of biosynthetic genes involved in the biosynthesis of secondary metabolites from fungi. Heterologous expression of biosynthetic genes has also provided detailed biochemical understanding of the enzymatic steps, including those involved in the biosynthesis of indole alkaloids.

Different strategies to incorporate indole into the final alkaloid structures are found in fungal secondary metabolism. Not surprisingly, most of the indole precursors are related to L-tryptophan (1), the most abundant indole-containing species in the cell. The biosynthesis of 1 starts from chorismate in the shikimic acid pathway and involves the intermediates anthranilate (2) and indole-3-glycerolphosphate (3). Intermediate 3 is transformed into indole, which can be coupled with serine to form 1. Tryptophan 1 can be decarboxylated and converted into tryptamine (4), or be prenylated at C4 to yield 4-dimethylallyl tryptophan (4-DMAT) (5). Feeding experiments with isotope-labeled precursors showed that 1, as well as 3, 4 and 5 can each serve as the biosynthetic precursor for the indole/indoline moieties in fungal indole alkaloids as summarized below and in Scheme 1:

i) 1 can be directly activated by the adenylation domains of nonribosomal peptide synthetases (NRPSs) and be incorporated into NRPs and polyketides-NRP hybrid compounds. Indole alkaloids derived from cyclic dipeptides such as meleagrin (6) from Penicillium chrysogenum, quinoline-containing tripeptides tryptoquialanine (7) from Penicillium aethiopicum and polyketide-containing tetracids are examples that incorporate 1 using this mechanism.

ii) The tryptophan precursor 3 can be an efficient indole donor as the glycerol-phosphate moiety at C3 can be substituted by an electrophile and leave as glyceraldehyde 3-phosphate. 3 serves as the indole donor during the biosynthesis of 1. The same chemistry is employed in the synthesis of indole diterpene family of fungal natural products, such as paxilline (8) from Penicillium paxilli, in which 3-geranylgeranyl indole is formed upon C3 prenylation of 3.
iii) 1 can be prenylated at C4 by 4-dimethylallyl tryptophan synthase (4-DMATS) to yield 5,16 which can undergo several oxidative cyclization steps that lead to the formation of the characteristic ergoline ring (a fused tetracyclic system) found in ergot alkaloids.17

iv) Decarboxylation of 1 leads to the formation of tryptamine 4, which is a widely used building block in the biosynthesis of plant indole alkaloids. For example, the biosynthesis of monoterpen indole alkaloids from plant shares the common precursor strictosidine, which is the product of the enzymatic Pictet–Spengler condensation18 between tryptamine and an iridoid monoterpene secologanin catalyzed by strictosidine synthase.19 A few fungal indole alkaloids, such as communesin F (9) from the communesin family of natural products, also have one of the indole moieties derived from 4.10 However, biochemical evidences that confirms the direct incorporation of 4 as a precursor has not been reported. Interestingly in 9, the second indole moiety is derived from 5, illustrating a differential use of indole precursors in the same molecule.10

2. Biosynthetic Reactions on the Indole Ring

Incorporation of the indole ring into alkaloid scaffolds endows powerful reactivity that can lead to dramatic morphing of the compound into complex, polycyclic structures. This is due to the unique chemistry of the benzopyrrole bicyclic ring system to undergo reaction at three sites, N1, C2 and C3. Indole can undergo a variety of electrophilic substitution modifications at C2 and C3, and can be subject to epoxidation at C2-C3, which inverts the polarity on C2 and leads to nucleophilic addition at this position. Indeed, the rich reactivity of indole has rendered it a “privileged structure” in drug discovery.20, 21 Here a brief overview of biosynthetic transformations involving indole will be discussed.

2.1 Electrophilic Substitution/Addition

Indole is an electron-rich heterocycle because of the conjugated nitrogen lone pair. Electrophilic substitution occurs readily on indole in the five-membered pyrrole ring since it is more electron rich compared to the benzene ring.22 During the biosynthesis of indole alkaloids, the most commonly observed electrophile added to the indole ring is the dimethylallyl carbocation generated from dimethylallyl-phosphoryl (DMAPP). Of the three positions on the pyrrole ring, C3 is the preferred site to electrophilic addition. Although in most cases C3 is substituted, there is still a strong preference for attack at this position to give an iminium ion (Scheme 2A). The iminium species can be further subjected to C2 nucleophilic addition, frequently by an intramolecular nucleophile, to yield a fused indoline structure (Scheme 2A). The iminium ion can also undergo Plancher rearrangement in which alkyl migration yields a C3 arenium cation intermediate, that can be aromatized through C2 deprotonation to afford a 3-substituted 2-alkylindole (Scheme 2A).23 Although there is a lack of evidence for the existence of the C3 arenium cation intermediate during indole prenylation, it is likely that C2 prenylation of a C3-substituted indole proceeds in this route.24, 25 This is also partially supported by the observation of both C2 and C3 prenylated products during the biochemical characterization of an indole prenyltransferase (PTase).26 An analogous prenyl migration was also proposed to take place between C3 and C4 positions via Cope rearrangement in the reaction catalyzed by fungal 4-DMATS.27 In addition to prenylation, S-adenosylmethionine (SAM)-dependent C3 methylation is another example of electrophilic addition, such as during the biosynthesis of physostigmine, a pyrroloindole drug found in Streptomyces species.28
2.2 Indole Epoxidation

The C2-C3 double bond in indole can also be epoxidized by a flavin-dependent monoxygenase (FMO). The resulting indole 2, 3-epoxide is in equilibrium with a hydroxyliminium ion due to the epoxide ring-opening reaction initiated from the nitrogen lone pair. The hydroxyliminium species is electrophilic and can be readily attacked at C2 by a nucleophile and form a C3-hydroxylated indoline. The nucleophilic attack can be frequently intramolecular, such as from the amide nitrogen as in the case during notoamide D biosynthesis, or be intermolecular, as from a free amino group during biosynthesis of 7. Interestingly, the presence of a C6 O-substitution in the benzyl ring can direct the 2, 3-epoxide intermediate to open at C3 position and undergo a semipinacol-like rearrangement, which results in the migration of the C2-substitution to C3 position to form a substituted 2-indolone, as in biosynthesis of notoamide C (Scheme 2B).

Scheme 2. Common Reactions on Indole Ring.

3 Biosynthetic Categories of Indole Alkaloids

3.1 Peptide Indole Alkaloids Synthesized from NRPS

NRPSs are large multimodular enzymes that assemble nonribosomal peptide (NRP) natural products through activating and joining of amino acid building blocks. Indole-containing NRPSs are produced when a NRPS module incorporates an L-tryptophan into the final compound. Adenylation (A) domains in NRPSs are responsible for activation of individual amino acids, and formation of aminoacyl thioesters attached to the thiolation (T) domains. Prediction of A domain substrate specificity has been successful for bacterial A-domains and has been developed into automated web-based resources. This method is based on the identities of residues within a radius of ~8 Å around the catalytic center (34-position signature sequence, as well as a concise 10 amino acid code) when mapped to the structure of the phenylalanine-activating A domain PheA. Application of this method towards predicting fungal A-domain specificity has been limited, with only few successful examples such as for anthramilic acid (2). Known bacterial NRPSs that utilize I as building block are relatively rare. A sequence logo representation of the 10 AA code for A-domains specific for I, from five bacterial and eight fungal NRPS examples, are shown in Figure 1. For comparison purpose, a comprehensive sequence logo of the 10 AA for >800 fungal A domains is shown. Among A domains that activate I, a valine residue at position 330 (numbering based on PheA) is highly conserved in addition to the general ‘invariant’ residues D235 and K517. Another moderately conserved residue is at position 236, which is occupied by less bulky hydrophobic residues (Val, Gly, Ala, etc.). Although the 10 AA code (or signature sequences) in fungal NRPSs is less informative due to the lack of sufficient sequence-specificity relationship studies, the sequence logo representation can be used to predict, confirm or exclude A domain specificity for I.

Most fungal indole alkaloids biosynthesized from NRPSs can be classified into either a dipeptide that involves formation of diketopiperazine; or an anthramilate-containing tripeptide that undergoes cyclization to contain a quinazoline core. There are larger NRP products that contain I, such as the tetrapeptide apicidin isolated from Fusarium semitectum that displays histone deacetylase inhibition activity. Due to space constrains, alkaloids originated from more than tripeptide NRPs will not be discussed here.
NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

### 3.1.1 Dipeptide Indole Alkaloids

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA.

Figure 1. A sequence logo representation of the 10 AA code for NRPS A-domains specificity. A) A domains that are specific for 1. B) All the fungal A domains in NRPSpredictor2 database. The residue numbering is based on PheA.

Dipeptidyl indole alkaloids are assembled by bimodular NRPSs, in which two amino acids are activated and condensed together. The second amino acid has been found to be L-proline, L-histidine, L-phenylalanine, L-alanine or another molecule of Trp. Aminoacyl-Trp is condensed with the second amino acid moiety to yield a linear dipeptidyl thiostreptamine intermediate brevianamide F (41). The residue numbering is based on PheA. 11 and its 6-methoxy version (tryprostatin A) can undergo oxidative coupling between the prenyl group and the diketopiperazine ring to form a fused 6-5-6-6-5 ring system as found in demethoxyfumitremorgin C (13) or fumitremorgin C (14). This reaction is proposed to be catalyzed by a cytochrome P450 monooxygenase (CYP) such as FtmE in the fumitremorgin pathway. Spirotryprostatins can be synthetized through further oxidative modifications via two distinct routes. One route is through a radical-mediated, two-step hydroxylation on 13 and 14 followed by dehydroxylation and semipina-acol-type rearrangement to form spirotryprostatin B (15) and G (16), which is catalysed by the CYP FtmG (Scheme 3A). Another route is through epoxidation of the indole 2,3-double bond by an FMO followed by a semipina-acol-type rearrangement. FtmG can also hydroxylate 14 twice and the resulting molecule can be converted to fumitremorgin B after N-prenylation by FtmH. Further oxidation by a Fe(II)-dependent α-ketoglutarate dioxygenase gives verruculogen (17).

During notoamide biosynthesis, 12 is modified by C6 hydroxylation and C7 normal prenylation to afford the common intermediate notoamide S (18). 18 serves as the branching point of subsequent biosynthetic steps towards either notoamide C (19) or notoamide B (20). To form 19, the C6 hydroxyl and C7 prenyl group are first oxidatively coupled to give notoamide D; which can be epoxidized at the C2 and C3 of the indole by NotB. Whereas a semipina-acol rearrangement on the hydroxymimium intermediate gives 19, a nucleophile attack from the α-amide of Trp (now part of diketopiperazine) forms notoamide D. Alternatively 18 can undergo dehydrogenation in the diketopiperazine ring to an achiral azadiene, which can form the bicyclo[2.2.2]diazocetane ring through intramolecular Diels-Alder reaction (IMDA). Whether the IMDA reaction is enzymatic or spontaneous is still unresolved because both stereoisomers of the cycloadduct are observed in nature. Additional oxidative steps, including a similar prenyl migration as proposed for 19, are required to produce (+) or (-)-20 in A. sp. MF297-2 (Scheme 3A).

Meleagrin (6) is an example of indole alkaloids derived from a histidyl-tryptophan diketopiperazine (HTD) (21). Biosynthetic studies of the meleagrin pathway have been performed in Penicillium chrysogenum. The gene cluster (roq) has been named after a well-known intermediate roquefortine C (22). In this seven-gene cluster, RoqA is the bimodular NRPS that synthesizes HTD. C3 reverse prenylation and intramolecular ring fusion catalyzed by the PTase RoqD yields a 6-5-5-6 ring system. α-β dehydrogenation of the histidyl component by the P450 oxidoeductase RoqR affords 22. Based on both gene silencing and gene deletion studies, the prenylation and dehydrogenation steps can take place independent of each other. 22 can be further converted into 6 and the O-methylated derivative oxaline that shows antimicrobial and anticancer activities. The conversion to 6 involves a proposed oxidative rearrangement by a Maackiaain monoxygenase RoqM, followed by indole N1 methylation by RoqO (cytochrome P450) and RoqN (methyltransferase).
The oxidative rearrangement that generates the unique 6-5-6-5 ring skeleton was proposed to initiate through hydroxylation on the diketopiperazine ring to form an acyloin, followed by ring opening and oxidative ring closure through a 9-membered ring β-ketoamide intermediate (Scheme 3B).²⁵

Scheme 3 Biosynthesis of Dipeptide Indole Alkaloids.

Acetylaszonalenin (23), a benzodiazepinedione isolated from Neosartorya fischeri NRRL181, is derived from the dipeptide R-benzodiazepinedione (24). The pharmacophore 24 is assembled by a bimodular NRPS (AnaPS) that combines 2 and 1 (Scheme 3C), and involves epimerization of the indole side chain.¹³, ²⁶ The release/cyclization of the dipeptide product is shown to take place spontaneously using model substrates.²⁷ Excision of the terminal epimerization domain in AnaPS led to the production of the S-stereoisomer in vitro, albeit with significantly reduced yield compared to that of 24.²⁸ C³ reverse
prenylation on 24 initiates intramolecular cyclization and forms the fused indoline-benzodiazepine azonalenin (25). Further acetylation at position N1 by an acetyltransferase (AnaAT) arrives at 23 (Scheme 3C).  

### 3.1.2 Tripeptide Indole Alkaloids

Fungal indole alkaloids containing three amino acid building blocks are synthesized by trimodular NRPSs (with the exception of asperlicin (26), see below). In addition to the incorporation of 1, a frequently used building block is the β-amino acid anthranilate 2. As a result of incorporation of 2, structurally complex tripeptidyl alkaloids can be synthesized from surprisingly short biosynthetic pathways. The amino and carboxylic acid functional groups in 2 are ortho to each other, which makes 2 a unique building block in the construction of multicyclic ring systems. One effective way of harvesting the reactivity and geometry of 2 is through a condensation like domain (C₁) positioned at the C-
terminal end of the NRPS assembly line. The C₇ domain catalyses the macrocyclization and the release of the tripeptide through nucleophilic attack of the anilide nitrogen of 2 on the peptidyl thioester (Scheme 4A). Depending on the sequence of the peptide, either a ten-membered (most frequent) or eleven-membered macrolactam can be formed, which can undergo additional intramolecular cyclization reactions to yield products such as a 6-6-6 tricyclic pyrazinoquinazolinedione as shown in Scheme 4A or an angular 6-6-7-6 tetracyclic pyrazinobenzodiazepinedione as asperlicin C (27).

The first biochemically characterized tripeptide alkaloid is fumiquinazoline F (FQF, 28), which is cyclized from Ant-d-Trp-Ala.⁶⁰ 28 serves as an intermediate in the biosynthesis of more complex alkaloids, such as fumiquinazoline D (29) in A. fumigatus; and tryptouquinaline (TQA, 7), a tremorgenic toxin from P. aethiopicum. The trimodular NRPS from the tqa biosynthetic pathway, TqaA, has been expressed from Saccharomyces cerevisiae and from Aspergillus clavatus.⁶¹ Aspergillus fischeri (FQC); or an aminal-containing, thermodynamic product asperlicin C (FQ); or an aminal-containing, thermodynamic product asperlicin C (FQ).

Removal of C₁ led to the hydrolisis of the linear tripeptide, which does not spontaneously cyclize under aqueous conditions. When the tripeptide is tethered to the last T domain of TqaA, the standalone C₇ domain catalyses the formation of 28. Substrates analogues of the tripeptide were used to probe the reaction catalysed by the C₇, which led to exclusion of a diketopiperazine cyclization route in favour of the macrocyclization step shown in Scheme 4A.⁵⁸

In the biosynthetic pathways of both 29 and 7, the indole ring in 28 is epoxidized by a FMO (Af12060 or TqaH).⁶⁰ Interestingly, Af12060 was shown to also epoxidize the diketopiperazine 14 en route to spirotryptrostatin A in A. fumigatus, hinting a shared indole epoxidase between unrelated pathways in fungi.⁴⁹ The resulting oxyz-fumiquinazoline F (30) is in equilibrium with the C₃ hydroxyiminium species (Scheme 4B). Although 30 is the last shared intermediate for the FQF and TQA pathways, the subsequent modifications are chemically parallel. In both pathways, a single-module NRPS activates either alanine (by Af12050) or aminoisobutyrate (by TqaB) to the corresponding aminoaoyl thioesters, and catalyses the attack of the free amino group on C₂ of 30 to yield an adduct.¹³,⁶⁰ The nucleophilic attack on the indole is catalysed by the terminal C domains in the monomodular NRPSs, which control the relative stereochemistry of the C₂-C₃ substituents (syn or anti).¹³ Subsequently, nucleophilic attack of the indole nitrogen on the NRPS thioester completes the 6-5-5 imidazolidinone ring substructure and yields either fumiquinazoline A (31) or 15-dimethyl-2'-epi-FQA (32). The last parallel step between the two pathways is the dehydrogenation of the pyrazinoquinazolinedione ring by a FMO Af12070 or TqaG, to yield a reactive imine. In the biosynthetic pathway of 29, the imine intermediate is attacked by nucleophiles from the imidazolidinone portion to form either a cyclic hemiaminal-containing fumiquinazoline C (FQC); or an aminal-containing, thermodynamic product 29. In contrast, in the TQA pathway, the imine intermediate undergoes ring-opening deamination, spirolactonization and additional modifications to yield the final product 7. The structurally related product tryptoquivaline found in A. clavatus is synthesized by the same overall pathway starting from the tripeptide Ant-d-Trp-Val.

Compared to 28, the biosynthesis of ardeemin (33) from Aspergillus fischeri involves the same three amino acids but with a different order of incorporation.⁶² The trimodular NRPS ArDa synthesizes the tripeptide Ant-d-Ala-Trp, which is cyclized into the pyrazinoquinazolinedione intermediate ardeemin FQ (34) by the C₇ domain of ArDa.⁶² Instead of epoxidation on the indole ring as observed in 28, the C₃ reverse prenylation by PTase ArdB generates the iminium cation that is attacked intramolecularly by α-amide of Trp (now as part of the pyrazinoquinazoline) to yield the hexacyclic (6-6-6-6-5-6) ring system in 33 (Scheme 4C).⁶² The prenylation/cyclization reactions performed by ArdB is analogous to those catalysed by AnaPT in the synthesis of 23,²⁶ and further illustrate the high degree of structural complexity that can be generated from alklylation of the indole moiety. Ardc, an AnaAT homolog, is proposed to acetyltye the indole N1 position to afford N-acetylardeemin (35).²⁶

The biosynthesis of asperlicin (26) involves an example of iterative function of a fungal NRPS module. Whereas three amino acids (two 2 and one 1) are proposed to constitute the peptidyl backbone, genome scanning of Aspergillus alliaceus revealed one bimodular NRPS AspA that may be involved in the biosynthesis.⁶³ Prediction of the specificities of the two A domains suggested that 2 and 1 should be activated sequentially. In vitro reconstitution of AspA purified from S. cerevisiae confirmed its function, as asperlicin C (27) was the dominant product formed accompanied by minor product asperlicin D.⁶³ Adenylation domain activation assays confirmed the predicted specificities and excluded dianthranilate being activated by the first module. Thus ArDa is proposed to first synthesize an Ant-Trp dipeptidyl thioester, followed by activation of a second molecule of 2 and formation of the Ant-Ant-Trp tripeptidyl thioester. The C₇ domain of AspA was confirmed to catalyse cyclization of the tripeptide into a proposed eleven-membered macrolactam intermediate.⁶³ The consequent transannular cyclization from attack of α-amide of Trp on the Ant carbonyl results in the formation of the minor product asperlicin D, while attack from the Ant amide on the Trp carbonyl leads to the on-pathway product 27.²⁶ Exopidation of the indole in 27 requires the FMO AspB that is functionally parallel to Af12060 and TqaH. Nucleophilic addition from the α-amide of Trp (now part of the benzodiazepinedione) affords the heptacyclic compound asperlicin E (36).²⁶ Hence, a remarkable two enzyme cascade (AspA and AspB) transforms two molecules of 2 and one molecule of 1 into a spectacularly complex multicyclic ring system. Alternatively a monomodular NRPS AspC can annulate the indole ring with one molecule of leucine to yield the isoindolone containing 26, similar to the function of TqaB in the synthesis of 32.²⁶
3.2 Hybrid Indole Alkaloids Synthesized from PKS-NRPS

A tryptophan-activating NRPS module can also be fused to fungal iterative polypeptide synthases (IPKs) to insert indole moieties. Fungal IPKs are highly programmed assembly-line machineries that can synthesize a large diversity of structures. A typical NRPS module domain arrangement is C-A-T-R, in which the C (Condensation) domain catalyses the transfer of the completed polypeptide acyl group to the α-amino of the aminoacyl thioester attached to the T domain. The terminal R (reductase) domain is proposed to reduce the final thioester product into an aldehyde followed by an intramolecular Knoevenagel condensation to yield a 3-pyrrolin-2-one core. Biochemical studies showed that the reductase-like domain is not always reducing: in some cases this domain functions as a Dieckmann cyclase (R* or DKC) that catalyses the attack of the α-carbon of the 1,3-diketo polypeptide portion on the thioester for tetramic acid formation and product release (Scheme 5).

Examples of indole alkaloids derived from characterized PKS-NRPS pathways include α-cyclopiazonic acid (CPA, 37) and chaetoglobosin A (38).

**Scheme 5. Biosynthesis of PKS-NRPS Derived Indole Alkaloids**

The pentacyclic 37 produced from *Aspergillus flavus* is a potent neurotoxin through inhibition of the sarcoplasmic Ca\(^{2+}\)-ATPase. The *cpa* gene cluster contains three genes, a PKS-NRPS (CpaS), a 4-DMATS (CpaD) and a flavoprotein oxidocyclase (CpaO). CpaS is proposed to synthesize the acetoacetyl-L-tryptophanyl thioester, which is cyclized by the terminal R* domain to yield cyclo-acetoacetyl-L-tryptophan (39) (Scheme 5A). Subsequently, 39 is regioselectively prenylated by CpaD at C4 to yield β-cyclopiazonic acid (β-CPA, 40). Unlike 4-DMATS, CpaD displays low activity against substrates lacking tetramic acid or similar functionalities. The final cascade of transformations that convert 40 to 37 is accomplished by CpaO, which most likely dehydrogenates β-CPA at the C₅ carbon of Trp, catalyses the electrophilic cyclization between the C3 and C4 substituents, and sets up a nucleophilic attack from the α-amide of Trp (now part of the tetramic acid) to afford the cis-pentacyclic 37 (Scheme 5A). Although the cyclized ring framework resembles that observed in ergot alkaloids, the cyclization mechanisms in these two pathways are drastically different.

Because of the reactivities associated with the indole ring, combinatorial matching of the CpaS NRPS module with PKSs that synthesize different polypeptide products is an attractive method of generating new indole compounds. Working with...
the aspyridone PKS-NPRs ApdA from *Aspergillus nidulans* that incorporates tyrosine, Xu et al. showed that the PKS and NRPS modules can function independently, and can be excised and paired in trans. Interestingly, heterologous pairing of the ApdA PKS module with the CpaS NRPS module led to functional crosstalk and synthesis of a new indole-containing tetramic acid (Scheme 5B).

Cytochalasans are a group of bioactive natural products that contain a highly substituted perhydroisoindolone core fused to a polyketide-derived macrocycle. Chaetoglobosins, including chaetoglobosin A (38) from *Chaetomium globosum*, is a subgroup of cytochalasans that use 1 as a building block. The biosynthetic pathway of 35 centers on the PKS-NRPS megasynthetase encoded by CHGG_01239, which synthesizes a pentaene-tryptophanyl product (Scheme 5C). The R domain in this PKS-NRPS is likely to be reductive and releases the acyclic product as an aldehyde, which can undergo spontaneous Knoevenagel condensation to form a doubly substituted pyrrolinone. A proposed intramolecular Diels-Alder (IMDA) reaction involving the pyrrolinone double bond and a specific diene in the polyketide portion of the acyclic molecule yields the key intermediate prochaetoglobosin (42), which can be furnished into 35 after a set of oxidative modifications that were recently worked out by Watanabe and coworkers.

![Scheme 6](image)

**Scheme 6. Biosynthesis of Fungal Indole Diterpenes.**

### 3.3 Indole Diterpenoids

Fungal indole diterpenes are indole-containing cyclic diterpenes found in numerous fungal species, including paxiline (8) from *Penicillium paxilli*; aflatem from *Aspergillus flavus*; emindoles from *Emericella desertorum*, and lolitrem from *Neotyphodium lolii*, etc. This group of structurally complex molecules has a variety of biological activities, including inhibition of ion channels and active against MRSA. Early feeding experiments suggested that the terpenoid structure is derived from geranylgeranyl pyrophosphate (GGPP) and the indole moiety is derived from indole-3-glycerol-phosphate (3). GGPP is synthesized from farnesy1 pyrophosphate (FPP) by an encoded GGPP synthase in the gene cluster, such as PaxG in pathway of 8. The acyclic 3-geranylgeranyl indole (43) was proposed to be a precursor common to all indole diterpenes. In vitro assays of PaxC, the geranylgeranyl transferase from the pax pathway, confirmed that 43 can be synthesized from 3 through geranylgeranyl addition at the C3 position followed by elimination of glyceraldehyde 3-phosphate. The differences in the oxidative cyclization steps from 43 generate structure diversity among the resulting cyclic products. Genetic studies showed that a flavin-dependent monoxygenase (PaxM) and an integral membrane protein (PaxB) are required for the transformation of 43 to paxiline (8), the fully cyclized precursor of 8. These two enzymes, along with PaxC and PaxG, are well-conserved in the genomes of indole diterpene producing fungi. Using a combination of heterologous expression and biotransformation in *Aspergillus oryzae*, Oikawa and coworkers showed that PaxM catalyses two epoxidation reactions on the diterpene
portion of 8, while PaxB serves as the terpene cyclase (Scheme 6). The electron-rich indole moiety is involved in the later steps of the carbocation-based cyclizations, resulting in the formation of the 6-5-5-6-6-6 fused ring system in 44. In vitro assays showed that PaxB can accept either the bis-epoxide containing diterpene 46 or mono-epoxide intermediate 45 to yield 44 and emindole SB (47) respectively. PaxB and its homologs therefore play an important role in dictating the regioselectivity and stereoselectivity of the final products. In forming 44 and 47, an anti-Markovnikov addition is required during the cyclization steps (Scheme 6), while a Markovnikov addition would lead to a structurally different subfamily including emindole DA (48). The regioselectivity and timing of the epoxidation reactions catalyzed by PaxM homologs also have a profound impact on the structure of the final products.

Additional enzymes present in the pathways can introduce further structural complexity and diversity to indole diterpenes, such as oxidative steps catalyzed by PaxP and PaxQ in transformation of 44 to 8. Recent findings by Liu et al. showed derivatives of 8 that are C5 mono-prenylated (paspalitrem A (49)) and C5/C6 diprenylated may be the final products in the pax pathway. Using recombinant PTase PaxD, the conversion of 8 to the prenylated derivatives was observed in vitro. Further search of the P. paxilli extract revealed the presence of these compounds, thereby suggesting the possibilities of 8 being a biosynthetic intermediate instead of a final product. A homologous PTase AtmD was found in the gene cluster of aflatem, which catalyses the C4 reverse prenylation. AtmD was shown to have considerable substrate flexibility towards the indole diterpene substrates, and can perform either normal or reverse prenylation when presented with different substrates.

Scheme 7. Biosynthesis of Indole Alkaloids from Other Indole Precursors
3.4 Indole Alkaloids Derived from Other Building Blocks

While 1 and 3 are the main sources of the indole moiety in fungi, other tryptophan derived metabolites are also used in the construction of indole alkaloids. These include indole pyruvate (50), 4-DMAT (5) and tryptamine (4).

Asterriquinones are a group of prenylated indole-containing compounds derived from two molecules of 50. Members of this family share the common dihydroxybenzoquinone core and vary mainly in prenylation pattern on the indole portion. The genes involved in the biosynthesis of terrequinone A (51) in A. nidulans was identified using a genome-wide survey of transcriptionally activated natural product gene clusters. Subsequent biochemical study reconstituted the entire five enzyme pathway required for formation of 51. As expected, 1 is first converted into 50 in a pyridoxal-5’-phosphate (PLP)-dependent fashion by the aminotransferase TdiD. A single-module NRPS TdiA with the domain organization of A-T-TE (TE: thioesterase) activates and couples two molecules of 50 to form dimethylasterriquinone D (52). Different from tryptophan-specific A domains, the A domain in TdiA is specific for the α-keto acid 50. The TE domain first serves as a waiting station in which the first molecule of 50 is covalently attached as an oxyster while the second molecule of 50 is activated as a thioester on the T domain; this is then followed by a TE-catalyzed head-to-tail dual Claisen condensation to form 52. The benzoquinone 52 can then be converted to 51 when incubated with TdiC (a zinc- and NAD(P)H-dependent quinone reductase), TdiB (an indole PTase) and TdiE (a methyltransferase like protein). 52 is proposed to be reduced by TdiC to the corresponding bisindolylhydroquinone, which is electron-rich and can be prenylated on the hydroquinone core by TdiB. The regioselectivity of this prenylation step is directed by TdiE, which was proposed to be a chaperone protein. In contrast, indole prenylation by TdiB is not influenced by TdiE (Scheme 7).

The biosynthesis of ergot alkaloids such as fumigaclavine C (53) goes through the intermediate 5, which can be synthesized from 1 by the action of dedicated 4-DMATS encoded in the individual gene clusters. The key intermediate in the biosynthesis of 53 is chanoclavine-I aldehyde (54), which is converted from 5 following a series of transformations including N-methylation, oxidative cyclizations and decarboxylation (Scheme 7). A detailed discussion of ergot biosynthesis can be found in a separate review article in this issue.

Decarboxylation of 1 by the PLP-dependent tryptophan decarboxylase results in the formation of 4, which is major constituent of plant alkaloids. A limited number of fungal alkaloids incorporating 4 have been reported, with communesins synthesized from Penicillium species and perophoramidine from Perophora namei being the most notable. Communesins, such as communesin F (9), have been reported to disrupt microfilament and lead to cytotoxicity to mammalian cells. The structural complexities of communesins, especially the cage-like polycyclic scaffold with four contiguous stereocenters and vicinal quaternary carbons, have attracted significant efforts from the total synthesis community. Both perophoramidine and communesins contain two connected, indole-derived halves. Isotope feeding experiments showed that both indoles originated from 1. One of the incorporated indole moiety is prenylated at C4 and has remained intact. Stoltz et al. proposed that (-)-aurantioclavine (55), an ergot-like molecule derived from 5, is a building block that serves as the prenylated half of the communesin structure (Scheme 7). The other incorporated indole in 9 is ring-opened and morphed into part of a benzylpyrrolidine. Feeding of 6-fluoro-tryptamine showed that the fluorine substitution was only found in the non-prenylated half of communesins, there by suggesting this half, which includes the pyrrolidine and aniline, are derived from 4. The enzymatic coupling of the two indole containing precursor (possibly 55 and 4) in such a regioselective and stereoselective fashion is unprecedented. At this point, no genetic and biochemical information of communesin biosynthesis has been reported.

4 Conclusions

The discussion above described the different biosynthetic strategies to incorporate and derivatize indole moiety involved in biosynthetic pathways of fungal indole alkaloids. The characteristic enzymes that catalyze key modifications for each family of indole alkaloids, including tryptophan-activating A domain in NRPSs or PKS-NRPSs, TqaH-like FAD-dependent indole epoxidases, 4-DMATS-like indole PTases, PaxB/PaxM homologs, and PLP-dependent tryptophan decarboxylases, can be fruitful leads for exploring new indole alkaloid pathways from sequenced fungal genomes. Such mining strategy combined with progressing techniques in molecular biology and microbiology, will accelerate the discovery of new fungal alkaloid natural products.

5 Acknowledgements

Due to space constraints, we apologize for any work that we were not able to cite. Related work in our lab is supported by the National Institute of Health (1R56AI101141 and 1DP1GM106413). We thank Prof. Neil K. Garg and Prof. Christopher T. Walsh for helpful discussions. We also thank Ralph A. Cacho for proof-reading the manuscript.

6 References