Published in final edited form as:
J Am Chem Soc. 2013 May 15; 135(19): 7205–7213. doi:10.1021/ja3123653.

Biosynthetic Pathway for Epipolythiodioxopiperazine Acetylaranotin in *Aspergillus terreus* Revealed by Genome-Based Deletion Analysis

Chun-Jun Guo[1], **Hsu-Hua Yeh**[1], **Yi-Ming Chiang**[1,3], **James F. Sanchez**[1], **Shu-Ling Chang**[1,4], **Kenneth S. Bruno**[2], and **Clay C.C. Wang**[1,5,*]

[1]Department of Pharmacology and Pharmaceutical Sciences, School of Pharmacy, University of Southern California, Los Angeles, CA 90089, USA

[2]Chemical and Biological Process Development Group, Energy and Environment Directorate, Pacific Northwest National Laboratory, Richland, WA 99352, USA

[3]Graduate Institute of Pharmaceutical Science, Chia Nan University of Pharmacy and Science, Tainan 71710, Taiwan

[4]Department of Biotechnology, Chia Nan University of Pharmacy and Science, Tainan 71710, Taiwan

[5]Department of Chemistry, College of Letters, Arts, and Sciences, University of Southern California, Los Angeles, CA 90089, USA

Abstract

Epipolythiodioxopiperazines (ETPs) are a class of fungal secondary metabolites derived from diketopiperazines. Acetylaranotin belongs to one structural subgroup of ETPs characterized by the presence of a seven-membered 4,5-dihydrooxepine ring. Defining the genes involved in acetylaranotin biosynthesis should provide a means to increase production of these compounds and facilitate the engineering of second-generation molecules. The filamentous fungus *Aspergillus terreus* produces acetylaranotin and related natural products. Using targeted gene deletions, we have identified a cluster of nine genes including one nonribosomal peptide synthetase gene, *ataP*, which is required for acetylaranotin biosynthesis. Chemical analysis of the wild type and mutant strains enabled us to isolate seventeen natural products from the acetylaranotin biosynthesis pathway. Nine of the compounds identified in this study are previously not reported natural products. Our data allow us to propose a biosynthetic pathway for acetylaranotin and related natural products.

Introduction

Epipolythiodioxopiperazines (ETPs) are a class of secondary metabolite toxins originating from diketopiperazines (DKPs).[1] One distinct feature of ETPs is the presence of unique di- or polysulfide bridges. This rare structural motif is hypothesized to mediate molecular toxicity in at least two ways: (1) by cross-linking vital proteins via cysteine bonds or (2) by forming reactive oxygen species (ROS) through redox cycling which can cause severe damage to the host cells.[1] The taxonomic distribution of ETPs amongst fungi is

*Correspondence should be addressed to C.C.C.W. (clayw@usc.edu).

Supporting Information Available: *A. terreus* strains and primers used in this study, detailed structure characterization, purification methods, RT-PCR and diagnostic PCR results, feeding experiment results, and NMR spectral data. This material is available free of charge via the Internet at http://pubs.acs.org.

discontinuous but at least 14 ETPs have been characterized from a diverse range of filamentous fungi.[1,2] The best characterized ETP, gliotoxin (Figure 1), was identified in the opportunistic pathogen *Aspergillus fumigatus*. A recent study has shown that gliotoxin induces apoptosis in mammalian cells by activating the proapoptotic Bcl-2 family member Bak.[3] Activation of the Bak protein elicits the generation of ROS and the release of apoptogenic factors by the mitochondria.[3] Gliotoxin has also been found to be an inhibitor of the transcription factor NF-κB.[4] Since this factor plays an integral role in the inflammatory immune response, the inhibitory effect of gliotoxin may account for the immunosuppressive properties of some ETPs.[4] Another ETP, acetylaranotin (**1**) (Figure 1), was first isolated from *Arachniotus aureus* as an antiviral agent.[5–7] The compound was later identified in *Aspergillus terreus*, the producer of lovastatin.[8] Subsequent bioactivity studies have shown that acetylaranotin (**1**) and its derivatives display an array of interesting biological activities including the inhibition of viral RNA polymerase,[6] the induction of apoptosis in cancer cell lines[9] and antifungal activity[10].

The cytotoxicity of ETPs has made them therapeutically important as potential anticancer agents.[11] With the dramatic increase of fungal genome sequence information, it is now possible to identify the specific genes that are responsible for ETP production. For instance, the gene clusters that are responsible for the biosynthesis of sirodesmin PL (Figure 1) in *Leptosphaeria maculans* (*sir* cluster) and gliotoxin (Figure 1) in *Aspergillus fumigatus* (*gli* cluster) have been identified and analyzed using bioinformatics.[12,13] Disruption of the key non-ribosomal peptide synthetase (NRPS) genes *sirP* and *gliP* confirmed their involvement in the biosynthesis of the two secondary metabolites.[12,14] The function of *gliP* was further elucidated by overexpressing the gene in *Escherichia coli*. GliP is a 236k Da three module (A_1-T_1-C_1-A_2-T_2-C_2-T_3) NRPS. The study revealed that the A_1 of GliP activates L-Phe and loads it onto T_1 while the A_2 activates L-Ser and tethers it onto T_2. The intramolecular cyclization is presumably catalyzed by C_1 to yield the DKP scaffold cyclo-L-Phe-L-Ser.[15] Previous studies have focused on the transannular disulfide bridge which is the main contributor to gliotoxin's deleterious effect. At least five genes, *gliC*, *gliG*, *gliJ*, *gliI* and *gliT* are essential for the formation of this chemical moiety. GliC is believed to be a P450 monoxygenase that hydroxylates the DKP scaffold followed by sulfurization by the glutathione S-transferase GliG. In the next step, the putative dipeptidase GliJ removes the Glu residues and GliI acts as a C-S lyase to yield the epidithiol moiety, which is oxidized to the disulfide bridge by the sulfhydryl oxidase GliT.[16–20]

Far less is known about acetylaranotin (**1**) biosynthesis (Figure 1) than that of gliotoxin prior to this study. Acetylaranotin (**1**) is characterized by the presence of two seven-membered dihydrooxepine rings. Total synthesis of acetylaranotin (**1**) was recently accomplished by the Reisman group and the Tokuyama group.[21,22] Yet, several obstacles have hindered an elaborate study of the biosynthesis of acetylaranotin (**1**) and the potential genes involved. First, the production yield of acetylaranotin (**1**) in *A. terreus* is low.[8,23–25] Second, although two candidate biosynthetic clusters for ETPs were predicted using bioinformatic analysis in *A. terreus*,[26] the lack of an efficient gene targeting system of this fungal species has made it difficult to determine the biosynthetic pathway of acetylaranotin (**1**).

Herein we report our efforts toward the identification and isolation of acetylaranotin (**1**) and its related natural products (**2** to **5**, **16**) from the wild type *A. terreus* NIH 2624 (Figures 1 and 2). One important step in identification of acetylaranotin (**1**) and related metabolites was examining a variety of culture conditions in an effort to produce the compounds at a relatively high yield. We have developed a gene deletion procedure in strain NIH 2624 of *A. terreus*. Targeted mutagenesis combined with metabolite analysis has allowed us to identify and characterize a gene cluster containing nine contiguous genes that code for the biosynthetic components responsible for the production of acetylaranotin (**1**). Targeted

deletions of each of the individual genes in the cluster generated several mutants that accumulated chemically stable intermediates or shunt products in sufficient amounts for characterization. Large scale culture of the wild type and mutant strains enabled us to isolate a total of 17 related compounds. Nine of the compounds were identified as previously not reported compounds. Subsequent bioinformatic analysis enabled us to propose a biosynthetic pathway for this ETP and the function of the enzymes involved in acetylaranotin (**1**) biosynthesis.

Results

Isolation and characterization of acetylaranotin (1) and biosynthetically related metabolites from *A. terreus* NIH 2624

Aspergillus species are known to produce different secondary metabolites when cultivated under different culture conditions. By screening different culture media we found that acetylaranotin (**1**) and its related secondary metabolites were produced at a quantifiable yield from Czapek's medium (Figures 2 and 3). The crude organic extract of *A. terreus* growing under acetylaranotin (**1**) producing conditions was analyzed by reverse phase liquid chromatography-diode array detection-mass spectrometry (LC-DAD-MS) and we identified several UV peaks that were possibly related to acetylaranotin (**1**) (Figure 3). For further verification, the strain was subjected to large scale cultivation, and each peak was initially purified by flash chromatography and subsequently by preparative HPLC. Using NMR spectroscopy, we determined the structures of five known compounds including acetylaranotin (**1**),[21] bisdethiobis (methylthio)-acetylaranotin (**2**),[27] acetylapoaranotin (**3**),[7] bisdethiobis (methylthio)-acetylapoaranotin (**4**),[27] and cyclo-(L-phe-L-phe) (**5**)[28], along with a previously not reported natural product, **16** [Figures 2 and 3; spectral data available in Supporting Information (SI)]. We also identified several *A. terreus* secondary metabolites that are not from the acetylaranotin (**1**) pathway. Several of these compounds have been characterized and reported earlier (Figure S2).[29] The other, unidentified, metabolites possess different chromophores from that of acetylaranotin (**1**) (Figures S1 and S2).

Identification of the gene cluster responsible for acetylaranotin (1) biosynthesis

Previous bioinformatic analysis has uncovered two putative ETP biosynthetic gene clusters in *A. terreus* NIH 2624.[26] Each of the two clusters contains an NRPS gene, ATEG_03470.1 and ATEG_08427.1, one of which was likely to be responsible for the biosynthesis of the DKP scaffold of acetylaranotin (**1**). A previous study using isotopic precursors showed that compound **5** could be incorporated into compound **2**.[30] Domain analysis by NCBI Conserved Domain Database (CDD) search revealed the putative protein coded by ATEG_03470.1 contains one adenylation (A) domain whereas two such domains were identified in the amino acid sequence of ATEG_08427.1. The A domain recognizes and activates a specific amino acid or aryl acid substrate.[31] We hypothesized that ATEG_03470.1 is the NRPS responsible for acetylaranotin (**1**) biosynthesis based on the observation that only one amino acid substrate, L-Phe, is activated and condensed to form **5**. To confirm our hypothesis, we deleted the gene ATEG_03470.1 in an *A. terreus* NIH 2624 strain with a *kusA*-, *pyrG*-background and cultivated the ATEG_03470.1Δ strain under the acetylaranotin (**1**) producing conditions. The *kusA* gene deletion improves the gene targeting efficiency due to high homologous recombination rates.[32] Analysis of the resultant secondary metabolites using LC-DAD-MS showed the complete elimination of compounds **1** to **5**, and **16** (Figure 3). Thus, we confirmed the involvement of this NRPS gene ATEG_03470.1 in acetylaranotin (**1**) production and designated it as *ataP*, using the letter designation developed for ETP biosynthesis genes.[26] Metabolites that are not related to acetylaranotin (**1**) biosynthesis are unaffected by the deletion of *ataP* (Figure 3).

Previous bioinformatic prediction of the putative acetylaranotin (**1**) cluster suggested that several genes downstream of *ataP* encode proteins that are homologous to the enzymes involved in gliotoxin biosynthesis. These enzymes include a cytochrome P450 monooxygenase (AtaF), glutathione S-transferase (AtaG), *O*-methyl transferase (AtaM), aminocyclopropane carboxylic acid synthase (ACCS) (AtaI), dipeptidase (AtaJ), thioredoxin reductase (AtaT), cytochrome P450 monooxygenase (AtaC), and major facilitator superfamily (MFS) transporter (AtaA).[26] We chose to name the genes involved in acetylaranotin (**1**) biosynthesis using the same letter code developed for ETP biosynthesis (Figure 3A and Table 1).[26]

However, in the updated *Aspergillus* Comparative Database, the *A. terreus* gene ATEG_03472.1 (*ataIMG*) is annotated as a trimodular gene with each module homologous to gliI, gliM, and gliG (Table 1). The gene ATEG_03474.1 (*ataTC*) is annotated as a bimodular gene with homology to gliT and gliC (Table 1). We investigated the mRNA sequences of these two genes products by using an RT-PCR approach. Using the cDNA of *ataIMG* as template and the genomic DNA as control, we were able to amplify two PCR fragments that flanked the sites where intron 1 and intron 8 were located, in the region encoding the AtaI and AtaG domains (Figure S4). This suggests that the mRNA of *ataIMG* includes the sequence that codes for a putative AtaI and AtaG domains. Another PCR fragment flanking the site where introns 4 and 5 were located was amplified. The gene sequence spanning introns 4 and 5 includes part of the coding sequence for both the putative AtaI and AtaM domains (Figure S4). The similar strategy was applied to analyze the mRNA sequence of *ataTC*. However, one PCR fragment that flanked intron T was amplified using *the* cDNA of *ataTC* as template. Intron T was not in the predicted sequence of *ataTC* but in the sequence of a putative transcript 3857_t (Figure S4A). The 3′-end of 3857_t overlaps with the 5′-end of *ataTC* by 579 bp. The predicted amino acid sequence of 3857_t is 48% similar to that of GliT (Table 1). The conserved motif (145-CXXC-148) in GliT, a hallmark for various sulfhydryl oxidases, can also be identified in the amino acid sequence of 3857_t (145-CXXC-148) but not in that of the annotated *ataTC*.[17,33] This indicates that the coding sequence of *ataTC* is likely to cover the sequence of one putative transcript 3857_t (Figure S4, see SI for detailed information of the transcript).

In addition, the two acetyl groups in the structure of acetylaranotin (**1**) suggest the involvement of an acetyl transferase gene. Analysis of the amino acid sequences of the genes in proximity to *ataP* using Basic Local Alignment Search Tool (BLAST) disclosed one gene ATEG_03466.1, which we named *ataH*, that encodes a putative acetyl transferase (Table 1). We then suspected that the three genes between *ataH* and *ataP* could also be involved. In all, an additional nine genes from *ataH* to *ataA* flanking the NRPS gene *ataP* were selected for deletion experiments. Four additional genes flanking the cluster, ATEG_03464.1, ATEG_03465.1, ATEG_03476.1, and ATEG_03477.1 were also selected for deletion analysis. Our bioinformatic analysis suggested that these four genes are beyond the acetylaranotin (**1**) biosynthetic gene cluster and examination of their secondary metabolite profiles would enable us to determine the two boundaries of the acetylaranotin (**1**) biosynthesis cluster.

The genes were individually deleted following the previously developed protocol.[29] Two verified deletants of each gene were cultivated under acetylaranotin (**1**) producing conditions. Examination of the LC/MS profiles of the mutant strains revealed that ATEG_03466.1Δ (*ataH*Δ), ATEG_03468.1Δ (*ataY*Δ), ATEG_03469.1Δ (*ataL*Δ), ATEG_03470.1Δ (*ataP*Δ), ATEG_03471.1Δ (*ataF*Δ), ATEG_03472.1Δ (*ataIMG*Δ), ATEG_03473.1Δ (*ataJ*Δ), ATEG_03474.1Δ (*ataTC*Δ), and ATEG_03475.1Δ (*ataA*Δ) impaired the production of compounds **1** to **4** (Figure 3), indicating that these nine genes are involved in the biosynthesis of acetylaranotin (**1**). The ATEG_03467.1Δ strain continued to

produce acetylaranotin (**1**) and its related natural products, suggesting that this gene, coding for a putative MFS transporter, is not involved in the biosynthesis pathway (Figure 3). As predicted, the secondary metabolite profiles remained unchanged after deletion of genes ATEG_03464.1, ATEG_03465.1, ATEG_03476.1, and ATEG_03477.1 indicating that these four genes are not involved in the biosynthesis of acetylaranotin (**1**) and that we have defined the two boundaries of the cluster (Figure 3).

Identification, purification, and structural determination of the intermediates and shunt products from mutant strains

Compound **5** is still identified in the metabolite profiles of all the deletion strains except the *ataP*Δ strain, suggesting that **5** may be the first intermediate synthesized by AtaP (Figure 3). However, we cannot exclude the alternative that **5** stands as a shunt product that could be derived from the non-enzymatic cyclization of tethered di-tyrosine.

Mutant strains containing gene replacements of four genes, *ataH*, *ataY*, *ataF* and *ataIMG*, accumulated sufficient amounts of chemically stable intermediates or shunt products for structure determination (Figure 3). Two natural products **6** and **7** were purified from scaled-up cultures of the *ataH*Δ strain (Figures 2 and 3). Compound **6** has been previously isolated from *Menisporopsis theobromae*.[34] We confirmed the structure of **7** using 1D and 2D NMR spectroscopy (see SI for detailed information). Both compounds have similar carbon skeletons as acetylaranotin (**1**), suggesting that the acetylation of **1** occurs in the later stage of its biosynthesis. From the culture of the *ataY*Δ strain we purified and characterized four secondary metabolites, **9** to **12 (**Figures 2 and 3). Compound **10** was previously reported as heamatocin, an antifungal DKP isolated from *Nectria haematococca*, a fungus causing nectria blight disease on ornamental plants.[10] Compounds **9**, **11**, and **12** are previously not reported compounds and their structures were elucidated via a complete 1D and 2D NMR and HRESIMS analysis (see SI for detailed information). The epitetrasulfide bridge, as shown in compound **9**, also presents in gliotoxin G which is a derivative of the disulfide gliotoxin.[35] Compounds **8**, **13**, **14**, and **15**, isolated from large-scale cultures of the *ataF*Δ strain are all derivatives of the dipeptide compound **5** (Figures 2 and 3). Reexamination of the LC/MS profiles revealed that compound **13** was also produced in the wild type and other mutant strains except the *ataP*Δ, *ataIMG*Δ, and *ataTC*Δ strains (Figure 3). Removal of gene *ataF* led to the accumulation of this minor metabolite and made it feasible to be analyzed by NMR. We were also able to purify and resolve the structure of **17** from the *ataIMG*Δ strain (Figures 2 and 3). Surprisingly, the metabolite **17** no longer maintains the DKP scaffold, indicating it might be a shunt product involved in acetylaranotin (**1**) biosynthesis.

To better differentiate between the intermediates and shunt products isolated in our study, compounds **4** to **7** and **10** were fed into the *ataP*Δ strain under acetylaranotin (**1**) producing conditions. We hypothesized that if one compound is an intermediate, feeding it into the *ataP*Δ strain might recover some of the strain's ability to reproduce its downstream natural products in the pathway. We could detect the accumulation of compound **2** in the extract of the *ataP*Δ strain after feeding compound **4** (Figure S6A). Compounds **1**, **2**, and **4** could be detected after feeding compound **5** (Figure S6B). We can identify compound **4** after feeding substrate **6** (Figure S6C) and compound **2** after feeding substrate **7** (Figure S6D). Our data suggest that compounds **4** to **7** are likely the intermediates involved in the biosynthesis of acetylaranotin (**1**) or its methylated form **2** (Figures 4 and S6). No downstream metabolites can be identified after feeding compound **10** (Figure S6E) indicating that it is likely a shunt product in the biosynthesis of **2** or alternatively it might not be able to cross the cell wall (Figure 4).

Discussion

We used a combination of genomics, efficient gene targeting and natural product chemistry to identify the genes encoding the biosynthesis pathway for the ETP toxin acetylaranotin (**1**) in *A. terreus*. After an optimal culture condition was found for acetylaranotin (**1**) production, we embarked on a targeted gene knock out approach to identify the genes involved in the biosynthesis pathway. We showed that the acetylaranotin (**1**) biosynthesis pathway is complex and involves, at least, the products of nine clustered genes. In addition to acetylaranotin (**1**), sixteen additional compounds produced from the pathway were isolated. Nine of these compounds are previously not reported. Bioinformatic analysis of the responsible genes, along with comparison to the similar gliotoxin biosynthesis pathway allowed us to propose a biosynthetic pathway for acetylaranotin (**1**).

In our proposed pathway, the first gene, *ataP*, encodes an NRPS for the biosynthesis of diketopiperazine **5**. AtaP has the domain architecture (T_1-C_1-A_1-T_2-C_2). AtaP shares 48% similarity to GliP, the NRPS in the gliotoxin pathway (Table 1).[14,15] The single A domain in AtaP suggests that one specific substrate, L-Phe, is loaded onto the T domains of AtaP. The sequence of AtaP A domain was aligned with several A domains of fungal NRPSs that have been shown to accept phenylalanine as their substrates (Figure S7). The alignment shows several conserved motifs that might be involved in the phenylalanine selection and activation by these A domains, but further investigation is necessary to define the A domain signature motifs for amino acid recognition in fungal NRPSs. A previous study using isotopic precursors has shown the intact incorporation of compound **5** into compound **2** in *A. terreus*.[30] We were able to detect accumulation of **5** in all mutant strains except the *ataP*Δ strain (Figure 3), indicating **5** is an early precursor involved in acetylaranotin (**1**) biosynthesis.

Five genes (*gliC*, *gliG*, *gliJ*, *gliI*, and *gliT*) are proposed to be involved in the sulfurization of the cylco-L-Phe-L-Ser DKP skeleton in gliotoxin biosynthesis.[16–20] In addition a sixth gene encoding the *O*-methyltransferase GliM may catalyze the methylation of the free thiols.[18,36] Homologs of these six gliotoxin genes can be identified in the acetylaranotin (**1**) gene cluster in *A. terreus* (Table 1). We propose that the AtaC domain of AtaTC catalyzes the formation of bishydroxylation of compound **5** to yield intermediate **19** (Figure 4). The *ataTC*Δ strain failed to produce any detectable intermediates or shunt products from the pathway except compound **5**, providing evidence that the dual hydroxylation is an early step in the biosynthesis pathway (Figures 3 and 4). The glutathione S-transferase domain AtaG in AtaIMG catalyzes the conjugation of two glutathiones to the bishydroxylated **19** to give precursor **20**, similar to the role of GliG in gliotoxin biosynthesis (Figure 4).[18,19] Next, we propose that the dipeptidase AtaJ removes the Glu resides of intermediate **20** to give **21** based on the proposed function of its homolog GliJ (Figure 4 and Table 1).[18] From the *ataJ*Δ strain we were able to identify intermediate **5** and two shunt metabolites **13** and **16**. These compounds do not contain a pyrrolidine ring moiety as shown in **1** suggesting that AtaJ is involved in the early steps of acetylaranotin (**1**) biosynthesis. Next, the carbon sulfur lyase domain AtaI of AtaIMG may catalyze a similar reaction to GliI converting the bis-cysteinyl adduct **21** to yield the epidithiol intermediate **22** (Figure 4). Purification of shunt product **8** projected the existence of the intermediate **23** (Figure 4). On the basis of homology and the conserved CXXC motif, we propose that the AtaT domain identified from the adjusted AtaTC protein sequence catalyzes the oxidation of the free dithiols in precursor **22** to yield **23**.

Next, the cyclization of **23** to give the intermediate **25** may be catalyzed by the cytochrome P450 AtaF. From the *ataF*Δ strain we identified compounds **5**, **8**, and **13** to **16** (Figure 3). The pyrrolidine structure in acetylaranotin (**1**) is noticeably missing in these compounds

(Figure 2). AtaF shares 70% similarity to GliF, which has not been characterized genetically or biochemically. AtaF probably acts as an epoxidase to promote a dual epoxidation formation at C8 and C9 along with C8′ and C9′ of **24**, followed by the spontaneous nucleophilic attack of the amide nitrogens N10 and N10′ of **24** to yield **25** with the pyrrolidine partial structure (Figure 4).

The final steps of acetylaranotin (**1**) biosynthesis involve the acetylation and ring rearrangement of intermediate **25** to produce acetylaranotin (**1**). These transformations are not found in the gliotoxin biosynthesis pathway. Therefore the genes involved in the transformation from **25** to **1** may not have homologs in the gliotoxin gene cluster. AtaH has 57% amino acid similarity to Tri7 of *Gibberella zeae* (Table 1). Tri7 has been shown to be involved in the acetylation of the oxygen at C4 of nivalenol.[37] AtaY has 51% similarity with a putative benzoate para-hydroxylase found in *Arthroderma otae* CBS 113480. The benzoate para-hydroxylase homolog (CYP51A) in *A. niger* has been previously purified and characterized.[38] We propose that AtaH catalyzes the acetylation of **25** to **26** and AtaY is responsible for the formation of the dihydrooxepin moiety that converts **26** to acetylaranotin (**1**) via compound **3** (Figure 4 and Table 1). Our data suggest that both enzymes could function independently in the absence of the other. From the *ataH*Δ strain we were able to isolate putative intermediates **6** and **7** that are not acetylated but both contain the dihydrooxepine partial structure. From the *ataY*Δ strain we were able to identify compounds **5**, **9** to **13**, and **16**. None of these compounds have the dihydrooxepine partial structure. Isolation of tetrasulfide **9** projected the existence of intermediate **26** in the biosynthesis of acetylaranotin (**1**). The acetyl groups on compound **9** suggest that the acetylation step can still occur in the absence of AtaY.

We identified an MFS transporter AtaA encoded in the acetylaranotin (**1**) biosynthesis cluster. The MFS genes found in secondary metabolite gene clusters sometimes code for transporters that are involved in the specific efflux of a secondary metabolite. Tri12, for instance, plays a role in the transport of trichothecene in *Fusarium* species.[39] In the *ataA*Δ strain, we observed that the production of acetylaranotin (**1**) was greatly reduced. This would suggest that AtaA could be involved in the efflux of acetylaranotin (**1**), or the *ataA* deletion could lead to a degradation product of **1** that cannot be secreted. It is also possible that *ataA* deletion could down regulate the biosynthesis of **1**, but it should be noted that other compounds (**2** to **5**, **13**, and **16**) are produced in the *ataA*Δ strain at a comparable level as in the wild type strain (Figure 3). Thus, the function of AtaA is still not clear and awaits further investigation.

There is another gene ATEG_03469.1 (*ataL*), for which we were not able to definitively assign the function of its putative product. BLAST analysis of AtaL reveals it has an uncharacterized ligand-binding domain and is 52% similar to GliH (Table 1). A previous study reported that deletion of *gliH* abolished gliotoxin production, suggesting that this gene is involved in gliotoxin biosynthesis or secretion.[16] Only compounds **5**, **13**, and **16** can be identified in the secondary metabolite profiles of the *ataL*Δ strain, indicating it may be involved in the substitution of the C-O bond in **19** to the C-S bond in **22**, or in the secretion of the acetylaranotin (**1**) and its related natural products. However, additional experiments will be necessary to further clarify the specific function of AtaL.

Methylation of the free dithiols of precursor **22** seems to be a branching point for the biosynthesis of acetylaranotin (**1**) and its sulfur-methylated form **2** (Figure 4). We propose that the AtaM domain of AtaIMG, a homolog of a putative *S*-methyl transferase GliM (Table 1), is responsible for this methylation step. Methylation of **22** by the AtaM domain of AtaIMG gives **28**, followed by AtaF-catalyzed cyclization that converts **28** to **29**. Ring expansion by AtaY converts **29** to **7** via intermediate **6**, and acetylation by AtaH produces

compound **2**. Another pathway that converts **6** to **2** involves the acetylation of **6** by AtaH to produce intermediate **4** followed by the ring expansion by AtaY that converts **4** to **2**.

Other shunt products (**11** to **17**) in acetylaranotin (**1**) biosynthesis were also isolated from gene deletion strains (Figure 5). Two molecules of water are probably eliminated from **19** to yield an unstable imine **30** followed by N-hydroxylation to produce **16** (Figure 5). Compound **17**, isolated from the *ataIMG*Δ strain, no longer maintains the DKP framework, suggesting that other genes beyond the *ata* cluster could be involved in its formation. The 2′ olefinic moiety found in **13** to **15** is probably formed via the elimination of one water molecule at C2′ of **19** (Figure 5). Methylation of the C2 hydroxyl in **14** and C2 thiol in **15** could be catalyzed by the AtaM domain of AtaIMG.

We were able to isolate two shunt products, **11** and **12**, from the *ataY*Δ strain. We hypothesized that AtaF is capable of catalyzing the dual epoxidation at C5 and C6 along with C5′ and C6′ of **19** to produce intermediate **31**, followed by spontaneous nucleophilic attack initiated by the free hydroxyl at C2 and C2′ that leads to the formation of the spiro partial structure in **11** (Figure 5). This mechanism has been mentioned in a previous study to explain the formation of two gliotoxin shunt products with a spiro partial structure.[36] The formation of compound **12** seems more complex. Sulfurization at C2′ of intermediate **19**, yielding a C2′ thiol intermediate **32**, likely involves AtaJ and AtaIMG. AtaF then catalyzes the epoxidation of **32** to **33** which spontaneously converts to **12** (Figure 5). The formation of the spiro shunt products **11** and **12** is probably due to the substitution of their benzene oxide intermediates **31** and **33** by the more nucleophilic hydroxyl or thiol groups. The epidisulfide bridge formation or the methylation of the free thiols or hydroxyls might reduce their nucleophilicity and enables the oxide intermediate **24** to be attacked by the amide nitrogen to give the pyrrolidine partial structure as shown in acetylaranotin (**1**).

Previous studies proposed that the reduced oxepine moiety as shown in compounds **1** to **4** is formed via equilibrium with its arene oxide tautomeric components prior to the pyrrolidine ring cyclization.[6,7] Then the epoxidation of the oxepine ring generates a *sym*-oxepine oxide moiety that is attacked by the nucleophilic amide nitrogen to give the pyrrolidine moiety.[6,7] Indeed, it is known that benzene oxide and oxepine exist in valence-tautomeric equilibrium with each other.[40] Though the calculated energy of benzene oxide is 1.7 kcal mole^{-1} lower than that of oxepine, the equilibrium is in favor of the oxepine form at room temperature due to the associated entropy gain ($\Delta G \approx -1.3$ kcal mole^{-1}).[40] The epoxide intermediates **24** (involved in pyrrolidine ring cyclization), **31**, and **33** (involved in spiro shunt product biosynthesis) are therefore likely to be in equilibrium with their more favorable oxepine forms (Figures 4 and 5). However, the methylene substitution at the 4 and 4′ position of these intermediates (Figures 4 and 5) may shift the equilibrium toward their oxide forms which may undergo nucleophilic substitution.[40,41] In our study, we can identify compound **2** in the total extract of the *ataP*Δ strain after feeding the strain with compound **4** under acetylaranotin (**1**) producing conditions (Figures 4 and S6A). This suggests that the oxepine moiety formation can occur after the pyrrolidine ring cyclization. Further investigation focusing on the catalytic function of AtaY will help to elucidate the mechanism involved in the dihydrooxepine rings formation in acetylaranotin (**1**).

Another interesting aspect is the self protection strategy adapted by *A. terreus* against the toxin acetylaranotin (**1**). GliT mediates self-resistance of *A. fumigatus* against gliotoxin. Removal of *gliT* accumulates the metabolite with free dithiols which may generate ROS or protein conjugates.[16,17] We did not observe any growth inhibition of *A. terreus* wild type under acetylaranotin (**1**) producing conditions suggesting that the AtaT domain functions normally. A recent study suggests that S-methylation of holomycin intermediates serves as another self-protection strategy taken by *Streptomyces clavuligerus*.[42] Compounds **1** and **3**

with an epidisulfide bridge and their S-methylated forms (**2** and **4**) were identified suggesting that *A. terreus* deploys both auto-protection strategies. The methylated sulfhydryl groups are less readily converted to dithiols compared to the disulfide moiety. Therefore, cellular toxicity of the methylated forms may be reduced.[42]

Conclusion

In summary using a targeted gene deletion approach we have shown that a total of nine clustered genes are involved in acetylaranotin (**1**) biosynthesis in *A. terreus*. By combining bioinformatic information and intermediates or shunt products isolated from the individual gene deletion mutants, we have proposed a biosynthetic pathway for the ETP toxin acetylaranotin (**1**).

Materials and Methods

Strains and molecular manipulations

Primers used in this study are listed in Table S1. *A. terreus* wild type and mutant strains used in this project are listed in Table S2. Each targeted gene was individually deleted by the *A. fumigatus pyrG* gene (*AfpyrG*) in the *kusA-, pyrG-* background of *A. terreus*. The construction of double joint fusion PCR products, protoplast generation, and transformation were carried out according to previous procedures.[29] Three putative gene replacements for each gene were randomly picked and verified by diagnostic PCR. At least two of the three transformants were determined to be correct for each gene (see SI for more information). The scheme of diagnostic PCR of the mutant strains is presented in Figure S3. The difference in size between the gene replaced by the resistant marker and the native gene allowed us to determine if the transformants carried correct gene replacement. If the targeted gene size is similar to the marker *AfpyrG*, an internal primer that specifically binds to AfpyrG was used with an outside primer. If the targeted gene has been successfully replaced by *AfpyrG*, a PCR fragment of around 4kb can be amplified from the mutant strains (Figure S5).

RT-PCR analyses

The *A. terreus* wild type strain was cultivated under acetylaranotin (**1**) producing conditions for 72 hours. Total RNA of *A. terreus* wild type strain was extracted using the Qiagen RNeasy Plant Mini Kit according to the manufacturer's instructions. Contaminating DNA was removed by digestion with DNase I (Fermentas). The cDNA was synthesized with the specific primers ATEG_03472.1RT and ATEG_03474.1RT (Table S1) using TaqMan Reverse Transcription Reagents (Applied Biosystems) following the supplied protocols. The cDNA was then used as template for PCR amplification with following specific primer sets of ATEG_03472.1 and ATEG_03474.1 flanking the putative introns (Figure S4 and Table S1). Amplification products were analyzed by electrophoresis in 1% agarose gels stained with ethidium bromide.

Fermentation and LC-MS analysis

The wild type and mutant *A. terreus* strains were cultivated at 30 °C on Czapek's agar plates (3 g $NaNO_3$/liter, 0.5 g KCl/liter, 0.50 g $MgSO_4{\cdot}7H_2O$/liter, 1.0 g K_2HPO_4/liter, 0.01 g $FeSO_4{\cdot}7H_2O$/liter, 30 g sucrose/liter, 15 g agar/liter) at 20 $\times 10^6$ spores per 10 cm plate. After 5 days, agar was chopped into small pieces and extracted with 50 ml of 1:1 CH_2Cl_2/MeOH with 1.0 hour sonication. The extract was evaporated in vacuo to yield a water residue, which was suspended in 25 ml water and partitioned with 25 ml ethyl acetate (EA). The EA layer was evaporated *in vacuo*, re-dissolved in 1 ml of 20% DMSO/MeOH and a 10μL portion was examined by HPLC–DAD–MS analysis. HPLC–MS was carried out

using a ThermoFinnigan LCQ Advantage ion trap mass spectrometer with a reverse-phase C18 column (Alltech Prevail C18 2.1 mm by 100 mm with a 3μm particle size) at a flow rate of 125 μl/min. The solvent gradient system for HPLC and the conditions for MS analysis were carried out as described previously.[29]

Isolation and characterization of secondary metabolites

For structure elucidation, the *A. terreus* wild type and mutant strains was cultivated on 40 × 150 mm Czapek's agar plates (~50 ml of medium per plate) at 30°C for five days. The UV-active secondary metabolites were isolated via flash chromatography and reverse-phase HPLC. A portion of 10 μL of each purified compound solution from the reverse-phase HPLC was then examined by an Agilent Technologies 1200 series high-resolution mass spectrometer. Melting points were measured with a Yanagimoto micromelting point apparatus that are uncorrected. IR spectra were recorded on a GlobalWorks Cary 14 Spectrophotometer. Optical rotations were measured on a JASCO P-200 digital polarimeter. NMR spectra were collected on a Varian Mercury Plus 400 spectrometer (see SI for more details).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The project described was supported in part by PO1GM084077 from the National Institute of General Medical Sciences to C.C.C.W. Research conducted at the Pacific Northwest National Laboratory was supported by the Department of Energy, Office of the Biomass Program. The Wang group is additionally supported by the Snyder Foundation.

References

1. Gardiner DM, Waring P, Howlett BJ. Microbiology. 2005; 151:1021–1032. [PubMed: 15817772]
2. Fox EM, Howlett BJ. Mycol Res. 2008; 112:162–169. [PubMed: 18272357]
3. Pardo J, Urban C, Galvez EM, Ekert PG, Muller U, Kwon-Chung J, Lobigs M, Mullbacher A, Wallich R, Borner C, Simon MM. J Cell Biol. 2006; 174:509–519. [PubMed: 16893972]
4. Pahl HL, Krauss B, Schulze-Osthoff K, Decker T, Traenckner EB, Vogt M, Myers C, Parks T, Warring P, Muhlbacher A, Czernilofsky AP, Baeuerle PA. J Exp Med. 1996; 183:1829–1840. [PubMed: 8666939]
5. Nagarajan R, Huckstep LL, Lively DH, DeLong DC, Marsh MM, Neuss N. J Am Chem Soc. 1968; 90:2980–2982.
6. Neuss N, Boeck LD, Brannon DR, Cline JC, DeLong DC, Gorman M, Huckstep LL, Lively DH, Mabe J, Marsh MM, Molloy BB, Nagarajan R, Nelson JD, Stark WM. Antimicrob Agents Ch. 1968; 8:213–219.
7. Neuss N, Nagarajan R, Molloy BB, Huckstep LL. Tetrahedron Lett. 1968; 9:4467–4471.
8. Miller PA, Trown PW, Fulmor W, Morton GO, Karliner J. Biochem Biophys Res Commun. 1968; 33:219–221. [PubMed: 4301873]
9. Choi EJ, Park JS, Kim YJ, Jung JH, Lee JK, Kwon HC, Yang HO. J Appl Microbiol. 2011; 110:304–313. [PubMed: 21122037]
10. Suzuki Y, Takahashi H, Esumi Y, Arie T, Morita T, Koshino H, Uzawa J, Uramoto M, Yamaguchi I. J Antibiot (Tokyo). 2000; 53:45–49. [PubMed: 10724007]
11. Vigushin D, Mirsaidi N, Brooke G, Sun C, Pace P, Inman L, Moody C, Coombes R. Med Oncol. 2004; 21:21–30. [PubMed: 15034210]
12. Gardiner DM, Cozijnsen AJ, Wilson LM, Pedras MS, Howlett B. J Mol Microbiol. 2004; 53:1307–1318.

13. Gardiner DM, Howlett BJ. FEMS Microbiol Lett. 2005; 248:241–248. [PubMed: 15979823]
14. Cramer RA Jr, Gamcsik MP, Brooking RM, Najvar LK, Kirkpatrick WR, Patterson TF, Balibar CJ, Graybill JR, Perfect JR, Abraham SN, Steinbach WJ. Eukaryot Cell. 2006; 5:972–980. [PubMed: 16757745]
15. Balibar CJ, Walsh CT. Biochemistry. 2006; 45:15029–15038. [PubMed: 17154540]
16. Schrettl M, Carberry S, Kavanagh K, Haas H, Jones GW, O'Brien J, Nolan A, Stephens J, Fenelon O, Doyle S. PLoS Pathog. 2010; 6:e1000952. [PubMed: 20548963]
17. Scharf DH, Remme N, Heinekamp T, Hortschansky P, Brakhage AA, Hertweck C. J Am Chem Soc. 2010; 132:10136–10141. [PubMed: 20593880]
18. Davis C, Carberry S, Schrettl M, Singh I, Stephens JC, Barry SM, Kavanagh K, Challis GL, Brougham D, Doyle S. Chem Biol. 2011; 18:542–552. [PubMed: 21513890]
19. Scharf DH, Remme N, Habel A, Chankhamjon P, Scherlach K, Heinekamp T, Hortschansky P, Brakhage AA, Hertweck C. J Am Chem Soc. 2011; 133:12322–12325. [PubMed: 21749092]
20. Scharf DH, Chankhamjon P, Scherlach K, Heinekamp T, Roth M, Brakhage AA, Hertweck C. Angew Chem Int Ed. 2012; 51:10064–10068.
21. Codelli JA, Puchlopek AL, Reisman SE. J Am Chem Soc. 2012; 134:1930–1933. [PubMed: 22023250]
22. Fujiwara H, Kurogi T, Okaya S, Okano K, Tokuyama H. Angew Chem Int Ed. 2012; 51:13062–13065.
23. Haritakun R, Rachtawee P, Chanthaket R, Boonyuen N, Isaka M. Chem Pharm Bull. 2010; 58:1545–1548. [PubMed: 21048353]
24. Kamata S, Sakai H, Hirota A. Agric Biol Chem. 1983; 47:2637–2638.
25. Haritakun R, Rachtawee P, Komwijit S, Nithithanasilp S, Isaka M. Helv Chim Acta. 2012; 95:308–313.
26. Patron NJ, Waller RF, Cozijnsen AJ, Straney DC, Gardiner DM, Nierman WC, Howlett BJ. BMC Evol Biol. 2007; 7:174. [PubMed: 17897469]
27. Lin Z, Zhu T, Fang Y, Gu Q. Magn Reson Chem. 2008; 46:1212–1216. [PubMed: 18846581]
28. Wang JM, Ding GZ, Fang L, Dai JG, Yu SS, Wang YH, Chen XG, Ma SG, Qu J, Xu S, Du D. J Nat Prod. 2010; 73:1240–1249. [PubMed: 20550196]
29. Guo CJ, Knox BP, Chiang YM, Lo HC, Sanchez JF, Lee KH, Oakley BR, Bruno KS, Wang CCC. Org Lett. 2012; 14:5684–5687. [PubMed: 23116177]
30. Boente MIP, Kirby GW, Robins DJ. J Chem Soc, Chem Commun. 1981:619–621.
31. Fischbach MA, Walsh CT. Chem Rev. 2006; 106:3468–3496. [PubMed: 16895337]
32. Ninomiya Y, Suzuki K, Ishii C, Inoue H. Proc Natl Acad Sci USA. 2004; 101:12248–12253. [PubMed: 15299145]
33. Heckler EJ, Rancy PC, Kodali VK, Thorpe C. BBA-Mol Cell Res. 2008; 1783:567–577.
34. Chinworrungsee M, Kittakoop P, Saenboonrueng J, Kongsaeree P, Thebtaranonth Y. J Nat Prod. 2006; 69:1404–1410. [PubMed: 17067151]
35. Kirby GW, Rao GV, Robins DJ. J Chem Soc, Perkin Trans. 1988; 1:301–304.
36. Forseth RR, Fox EM, Chung D, Howlett BJ, Keller NP, Schroeder FC. J Am Chem Soc. 2011; 133:9678–9681. [PubMed: 21612254]
37. Lee T, Han YK, Kim KH, Yun SH, Lee YW. Appl Environ Microbiol. 2002; 68:2148–2154. [PubMed: 11976083]
38. Faber BW, van Gorcom RF, Duine JA. Arch Biochem Biophys. 2001; 394:245–254. [PubMed: 11594739]
39. Alexander NJ, McCormick SP, Hohn TM. Mol Gen Genet. 1999; 261:977–984. [PubMed: 10485289]
40. Vogel E, Güther H. Angew Chem Int Ed. 1967; 6:385–401.
41. Hayes DM, Nelson SD, Garland WA, Kollman PA. J Am Chem Soc. 1980; 102:1255–1262.
42. Li B, Forseth RR, Bowers AA, Schroeder FC, Walsh CT. Chembiochem. 2012; 13:2521–2526. [PubMed: 23097183]

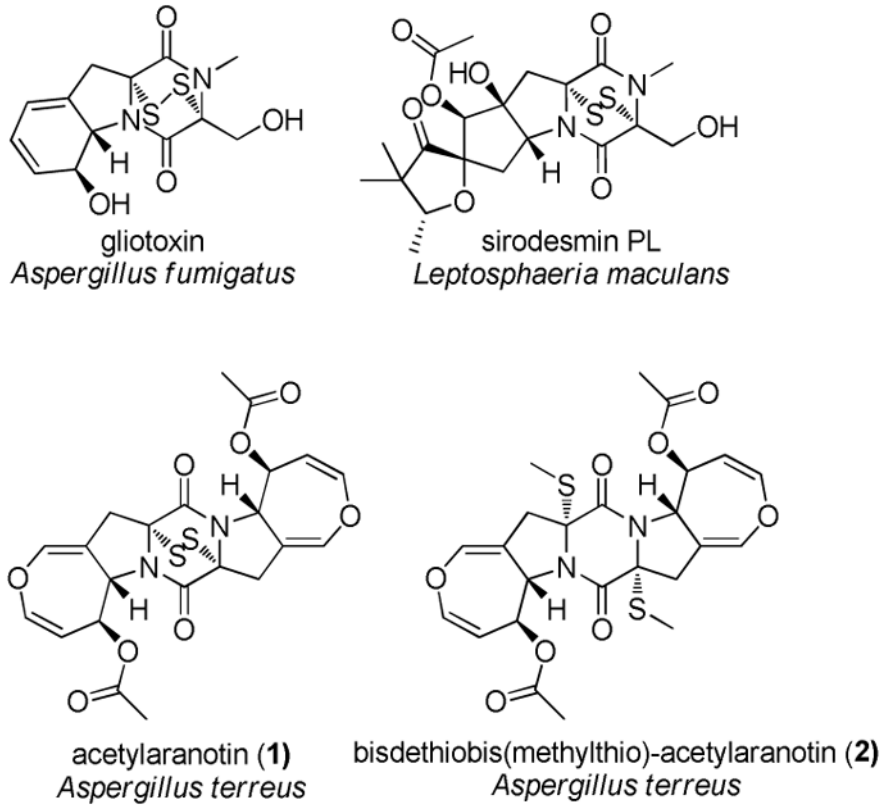

Figure 1.
ETPs and their derivatives identified from different fungal species.

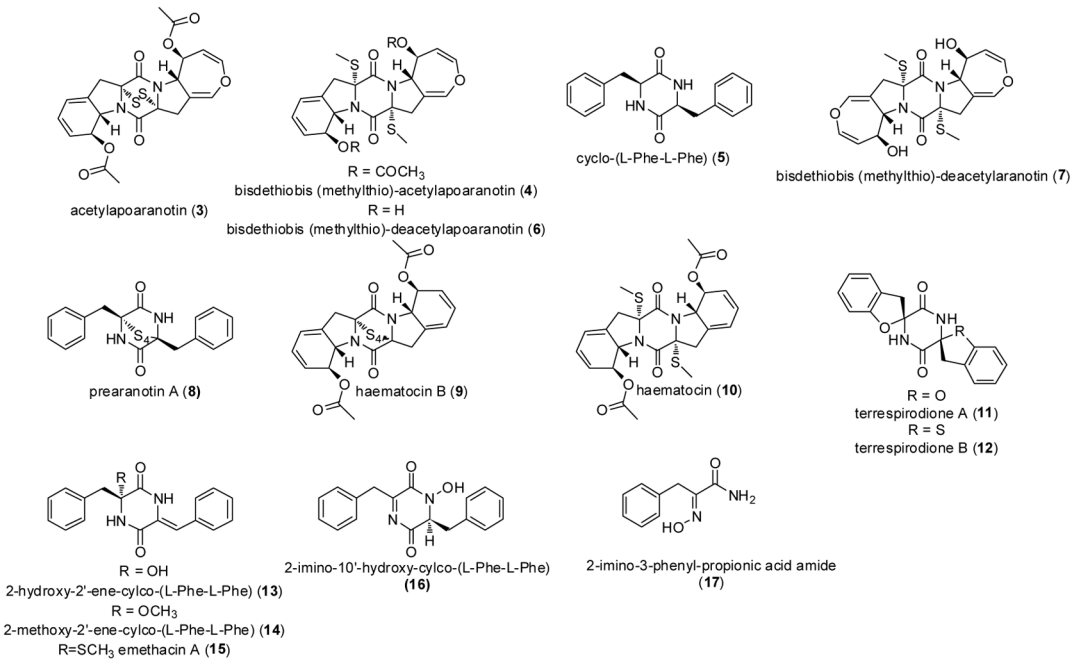

Figure 2.
Natural products from the acetylaranotin (**1**) pathway isolated from this study.

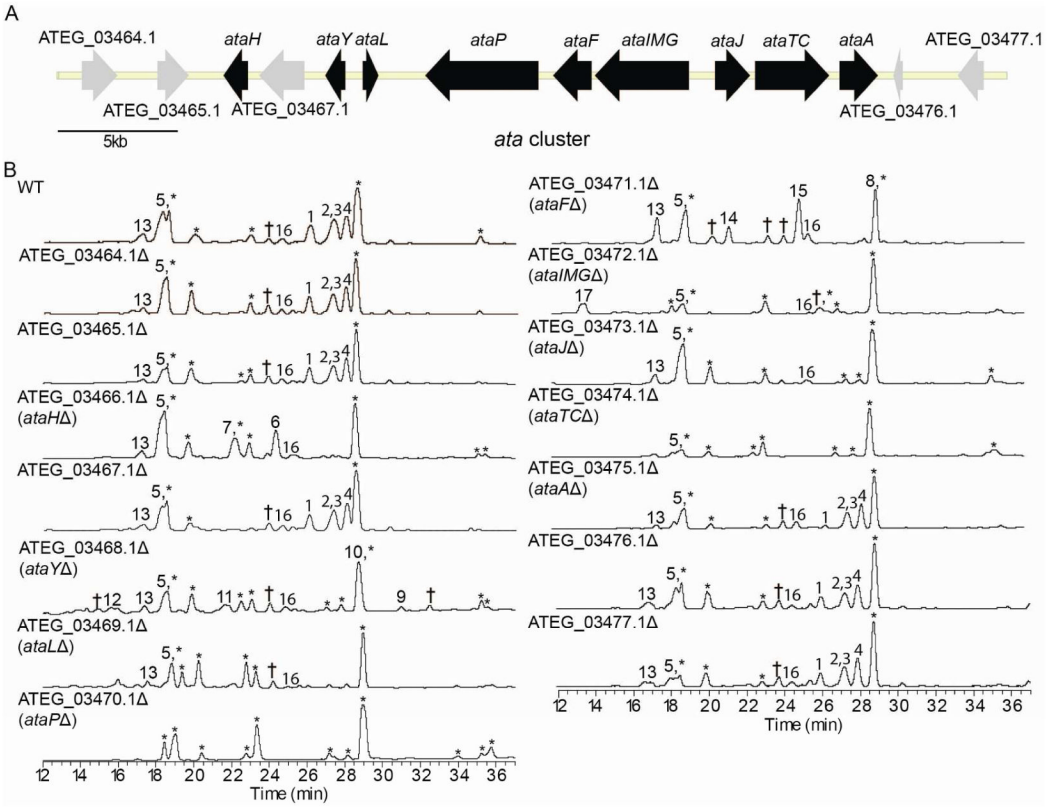

Figure 3.

(A) Organization of the acetylaranotin (**1**) biosynthesis gene cluster in *A. terreus*. Each arrow indicates relative size and the direction of transcription of the open reading frames deduced from analysis of nucleotide sequences. Genes in black are involved in acetylaranotin (**1**) biosynthesis while those in grey are not.

(B) HPLC profile of extracts of strains in the cluster as detected by UV at total scan. Numbering on peaks corresponds to the natural products shown in Figure 2.

* These compounds can be identified in ATEG_03470.1Δ and are not involved in acetylaranotin (**1**) biosynthesis.

† These compounds cannot be characterized due to poor yield produced or instability.

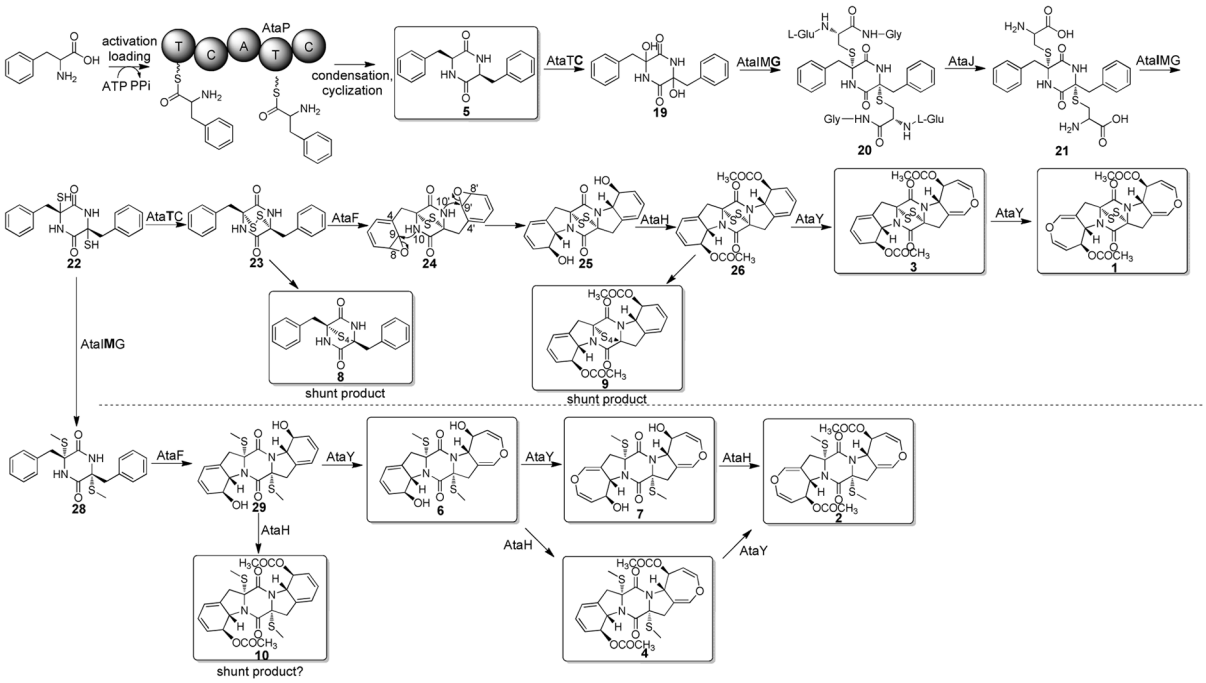

Figure 4.

Proposed biosynthetic pathway for acetylaranotin (**1**) and its methylated form (**2**). The pathway bifurcates to **1** and **2** and is separated by a dashed line at this point. For proteins with fused catalytic domains, the domain that is active at a particular step is in bold. All the natural products isolated from this study are boxed.

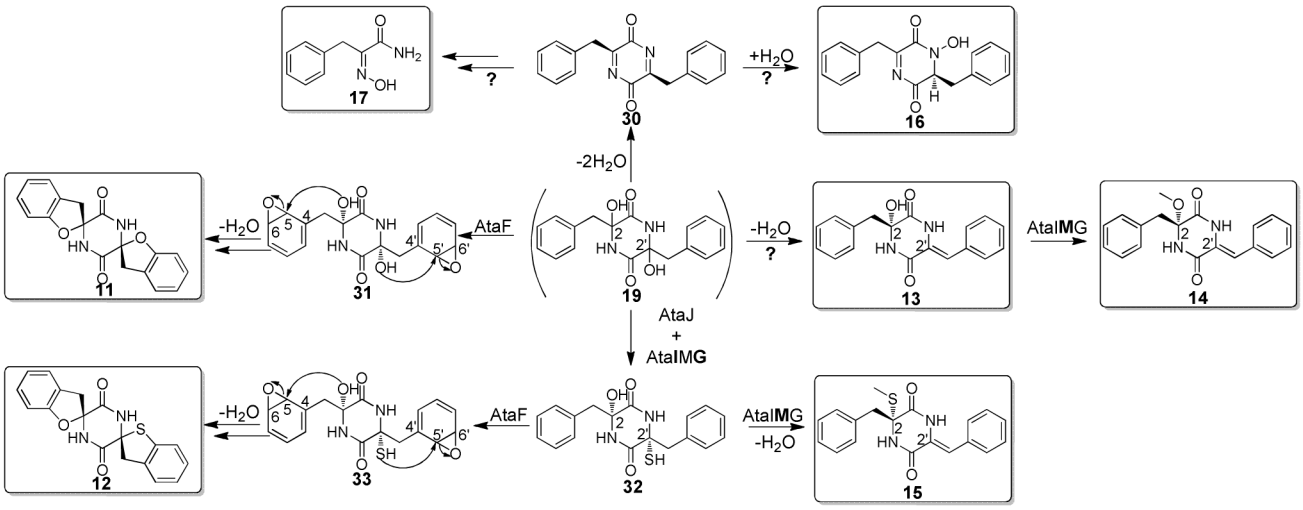

Figure 5.

Additional shunt pathways in acetylaranotin (**1**) biosynthesis. For proteins with fused catalytic domains, the domain that is active at a particular step is in bold. The starting molecule **19** is shown in round brackets. All compounds isolated from this study are boxed.

Table 1

Putative functions and homologs of the genes in the *ata* cluster.

Gene	Size (gene/aa)	Characterized homolog	Identity/similarity (%)	Putative function
ATEG_03466.1 (*ataH*)	979/303	Tri7, *Gibberella zeae*[37]	38/57	acetyltransferase
ATEG_03468.1 (*ataY*)	795/228	MCYG_07058[a], *Arthroderma otae* CBS 113480	33/51	benzoate para-hydroxylase
ATEG_03469.1 (*ataL*)	629/155	GliH, *A. fumigatus* Af293[16]	36/52	putative protein
ATEG_03470.1 (*ataP*)	4724/1507	GliP, *A. fumigatus* Af293[14,15]	30/48	NRPS
ATEG_03471.1 (*ataF*)	1572/489	GliF[a], *A. fumigatus* Af293	50/70	epoxidase
ATEG_03472.1 (*ataIMG*)	3923/874	GliI, *A. fumigatus* Af293[20] GliM[a], *A. fumigatus* Af293 GliG, *A. fumigatus* A1163[18,19]	27/44 41/57 39/56	carbon-sulfur lyase *O*-methyltransferase glutathione S-transferase
ATEG_03473.1 (*ataJ*)	1437/410	GliJ[a], *A. fumigatus* Af293	51/66	membrane dipeptidase
ATEG_03474.1 (*ataTC*)	2505/718	GliT, *A. fumigatus* Af293[16,17] GliC[a], *A. fumigatus* Af293	35/48 40/59	sulfhydryl oxidase hydroxylase
ATEG_03475.1 (*ataA*)	1581/449	GliA[a], *A. fumigatus* Af293	35/52	MFS

[a]The functions of these enzymes have been proposed but needs further verification.