

Published in final edited form as:

Chem Biol. 2014 February 20; 21(2): 257–263. doi:10.1016/j.chembiol.2013.12.005.

Aryl-aldehyde formation in fungal polyketides: Discovery and characterization of a distinct biosynthetic mechanism

Meng Wang¹, Mirko Beissner¹, and Huimin Zhao^{1,2,*}

¹Department of Chemical and Biomolecular Engineering, Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

²Departments of Chemistry, Biochemistry, and Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

SUMMARY

Aryl-aldehydes are a common feature in fungal polyketides, which are considered to be exclusively generated by the R domain of non-reducing polyketide synthases (NR-PKSs). However, by cloning and heterologous expression of two cryptic NR-PKS and non-ribosomal peptide synthase (NRPS)-like genes from *Aspergillus terreus* in *Saccharomyces cerevisiae*, we discovered a distinct mechanism for aryl-aldehyde formation in which a NRPS-like protein activates and reduces an aryl-acid produced by the accompanying NR-PKS to an aryl-aldehyde. Bioinformatics study indicates such a mechanism may be widely used throughout the fungi kingdom.

INTRODUCTION

Fungal polyketides are an important class of natural products, which have led to the development of numerous pharmaceuticals such as the blockbuster drug lovastatin. Genome sequencing of various fungal strains has revealed that each strain contains many more polyketide (PK) and/or non-ribosomal peptide (NRP) biosynthetic pathways than what was known (Galagan, et al., 2005). These unknown or cryptic pathways present both great opportunities and challenges for natural product discovery.

Aryl-aldehydes are a common feature in fungal polyketides. They can be retained in the final products, such as 3-(2,4-dihydroxy-6-methylbenzyl) orsellinaldehyde (Ahuja, et al., 2012) **1**, or serve as an active intermediate such as **2** and **3** for further modifications, like in asperfuranone (Chiang, et al., 2009) and tropolone (Davison, et al., 2012) biosynthesis (Figure 1). On the genetic and biochemical level, all previously characterized aryl-aldehydes in fungal polyketides are generated via a reducing domain (R domain), which is covalently attached to the end of a non-reducing polyketide synthase (NR-PKS). During our efforts to discover compounds from cryptic fungal pathways, we noticed that a large number of uncharacterized NRPS-like genes are present in fungal genomes (Sanchez, et al., 2012). Domain structure analysis of NRPS-like gene reveals that there are two sub-types of NRPS-

© 2013 Elsevier Ltd. All rights reserved.

*Correspondence: zhao5@illinois.edu.

SUPPLEMENTAL INFORMATION

Supplemental Information includes two tables, two figures and Supplementary experimental procedures

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

like proteins. Sub-type I NRPS-like proteins have an adenylation (A) domain, an acyl-carrier protein (ACP) domain, and a thioesterase (TE) domain. Sub-type I proteins have been characterized to be involved in the generation of various fungal secondary metabolites, such as microperfurane (Yeh, et al., 2012) and atromentin (Wackler, et al., 2012). However, sub-type II NRPS-like proteins, which have an A domain, an ACP domain, and an R domain, have never been characterized until very recently by Forseth *et al.* In their study, the NRPS-like gene *InaA* is proposed to reduce a L-tyrosine to an aldehyde intermediate, which then is converted to various metabolites (Forseth, et al., 2013). It is interesting that some of those NRPS-like genes are accompanied by NR-PKS genes in the vicinity on their genomes. For example, ATEG_03630 from *A. terreus* is a sub-type II NRPS-like gene while ATEG_03629 is a NR-PKS gene, which consists of a starter unit: ACP transacylase (SAT) domain, a ketosynthase (KS) domain, an acyltransferase (AT) domain, a product template (PT) domain, two ACP domains, a methyltransferase (MT) domain, and a TE domain. It is well known that NR-PKSs can produce aryl-acids (Ahuja, et al., 2012; Sanchez, et al., 2012). In addition, it was documented that adenylation domain can accept an aryl-acid as substrate in NRP/PK biosynthesis such as enterobactin (Crosa and Walsh, 2002). Therefore, we hypothesized that in those cryptic pathways, a NR-PKS produces an aryl-acid, which is then converted to an aryl-aldehyde by a NRPS-like protein.

Several methods have been developed to discover unknown products from cryptic pathways (Brakhage and Schroeckh, 2011). For example, changing growth conditions of the native producer coupled with transcriptome analysis may link a certain compound to a particular pathway (Schroeckh, et al., 2009). In addition, adding epigenetic factors can also turn on cryptic pathways (Yakasai, et al., 2011). Recently, Ahuja *et al.* demonstrated that replacing the promoter of a PKS gene with the *alcA* promoter in *Aspergillus nidulans* can produce many new polyketides (Ahuja, et al., 2012). However, those methods are limited by the minimal, sometimes non-existing genetic manipulation methods of the native hosts. In addition, the complicated background metabolite profile of the native host can also hinder the discovery of new compounds. Therefore, in this study, we sought to heterologously express our genes of interest in *S. cerevisiae*. Heterologous expression in *S. cerevisiae* provides many advantages such as almost unlimited genetic manipulation capability and a much cleaner metabolite background. However, it also comes with its own limitations such as protein expression problems and unknown precursor requirements. Particularly, *S. cerevisiae* cannot process introns from other fungi correctly. Therefore, based on bioinformatics prediction, we used the DNA assembler method (Shao and Zhao, 2009) to construct our target genes in an “intron-less” fashion.

In this study, we discovered and partially characterized a cryptic pathway from *A. terreus*. Through cloning, heterologous expression and *in vitro* study, we discovered a distinct mechanism for aryl-aldehyde formation in which a NRPS-like protein activates and reduces an aryl-acid produced by the accompanying NR-PKS to an aryl-aldehyde. Bioinformatics study indicates such a mechanism may be widely used throughout the fungi kingdom.

RESULTS

Transcription Analysis of the Target Cryptic Pathway in Native Host *A. terreus*

To affirm our target gene cluster (Figure 2A) is indeed cryptic, we performed a quantitative-PCR analysis to determine the transcription level of each gene from this cluster under a particular growth condition (Figure S1E). As a result, ATEG_03630 and ATEG_03629 are expressed less than 10% of internal control, while other putative tailoring enzymes are expressed less than 20%, which indicates the pathway is likely to be cryptic under our growth condition (Gressler, et al., 2011).

Discovery of Compounds from the Target Cryptic Pathway by Heterologous Expression in *S. cerevisiae*

To confirm our hypothesis, we set out to investigate the function of ATEG_03630 and ATEG_03629 from *A. terreus* (Figure 2A). Using the DNA assembler method we recently developed (Shao and Zhao, 2009), each exon was first amplified and assembled into a full length gene under the control of a constitutive promoter. Then, a plasmid containing the two-gene cassette was constructed via the DNA assembler method (see Supporting Information for details on plasmid construction). It has been demonstrated that a fungal 4'-phosphopantetheinyl transferase (PPTase) is crucial for the posttranslational modification of PKSs in *S. cerevisiae* (Wattanachaisaareekul, et al., 2007). Therefore, we also cloned the *npgA* gene encoding a PPTase from *A. nidulans* under the GPM1p promoter into plasmid pRS416-GPM1p-npgA-GPM1t.

Next, we tested the function of ATEG_03629 and ATEG_03630 by co-transforming pRS414-ATEG_03629 or pRS414-ATEG_03629_03630 with pRS416-GPM1p-npgA-GPM1t to *S. cerevisiae* strain HZ848. The cell cultures of the corresponding strains were extracted with ethyl acetate and dissolved in methanol for HPLC-ESI-MS analysis. As shown in Figure 2B, one distinct peak with the molecular weight of 182 was found in the strain expressing the ATEG_03629 gene and another peak with molecular weight of 166 was found in the strain expressing the ATEG_03629_03630 two-gene cassette. The decrease of molecular weight by 16 when the NRPS-like gene ATEG_03630 is included in the plasmid indicates that ATEG_03630 is performing a reduction reaction as we proposed. The two compounds were then purified from large-scale cultures for structure elucidation. Compound **4** was determined to be 5-methyl orsellinic acid (5-MOA) while compound **5** was 2,4-dihydroxy 5,6-dimethyl benzaldehyde (see Supporting Information for detailed characterization). Although compounds **4** and **5** were both reported in different literatures (Chen, et al., 2011; Ishiuchi, et al., 2012), their biosynthetic origin and connection have never been established.

To further unambiguously confirm that these two compounds were produced by heterologous expression of ATEG_03629 and ATEG_03630, we constructed loss of function mutants by mutating the posttranslationally modified serine of the ACPs to an alanine. Because ATEG_03629 has an unusual domain structure of two consecutive ACPs, we first constructed double mutant ATEG_03629m-S1651A-S1772A via the DNA assembler method. Single mutant ATEG_03630m-S611A was also constructed. The strain expressing mutated NR-PKS ATEG_03629m-S1651A-S1772A did not produce compound **4** or **5** (Figure 2B). In comparison, the strain harboring the plasmid containing the combination of a functional ATEG_03629 and the mutant ATEG_03630m-S611A protein only produced acid **4** (Figure 2B). These results confirm our hypothesis that the PKS ATEG_03629 produces 5-MOA which is then activated and reduced by the NRPS-like protein ATEG_03630 to 2,4-dihydroxy 5,6-dimethyl benzaldehyde (Figure 2C).

***In vitro* Characterization of the NRPS-like Protein**

To confirm that NRPS-like protein ATEG_03630 indeed performs the proposed chemical transformation, we over-expressed and purified ATEG_03630 from *S. cerevisiae* BJ5464-NpgA strain (see supporting information for plasmid construction). Using compound **4** as substrate, the purified ATEG_03630 protein can produce **5** *in vitro* with specific activity of 4.56 ± 0.21 $\mu\text{mol/h/mg}$ (Figure S1M).

Usually, the substrate of the A domain of NRPS can be predicted fairly well based on bioinformatics analysis (Rausch, et al., 2005). Therefore, we tried to predict the substrate specificity of ATEG_03630 using software NRPS predictor (<http://ab.inf.uni-tuebingen.de/>

[toolbox/index.php?view=domainpred](#)). However, no prediction could be obtained. With updated software NRPSpredictor2 (<http://nrps.informatik.uni-tuebingen.de>), 4-hydroxyphenyl-glycine (HPG) was predicted as a substrate. However, *in vitro* assay of HPG with purified ATEG_03630 protein indicates that no putative aldehyde product can be detected (Figure S1N). This indicates that ATEG_03630 may use an entirely different set of code for substrate recognition. Therefore, we performed sequence alignment of the A domain of ATEG_03630 with different A domains of NRPSs with known substrate specificity. Since our *in vivo* results indicate that 5-MOA is its native substrate, numerous aromatic acids such as 2,3-dihydroxy benzoic acid and anthranilic acid are included. The ten-letter nonribosomal code for ATEG_03630 was deduced from sequence alignment (Table S2). The nonribosomal code of ATEG_03630 is more similar to aromatic acids than aromatic amino acids. However, the code is still significantly different from that of known aromatic acid substrates, which supports our observation that, unlike LnaA, 5-MOA is the native substrate for ATEG_03630. To further differentiate ATEG_03630 from LnaA, three different aromatic amino acids (Tyr, Phe and Trp) were used as substrates to investigate the substrate specificity of ATEG_03630 *in vitro*. HPLC-MS analysis indicates that none of the three amino acids can be converted to their putative aldehyde products (Figures S1O, S1P and S1Q). The deduced ten-letter nonribosomal code of 5-MOA shall shed some light on *in silico* prediction of NRP.

To further demonstrate the function of the putative reducing domain and gain more insights on its structure and catalytic activity, we set out to characterize the R domain by mutagenesis. Multiple sequence alignment of the R domain with short-chain dehydrogenase/reductase family proteins indicates that it contains a conserved TGXXGXXG cofactor binding motif, and a YXXXXK active site motif (Persson, et al., 2003) (Figure S1F). Therefore, we generated three single mutants (T690A, Y863F and K867A) of ATEG_03630 to study its cofactor binding and catalytic activity. Compared to the wild-type protein, the mutant T690A only retains $0.128 \pm 0.001\%$ activity, likely because of the significantly reduced cofactor binding ability. The substitution of the catalytic residue lysine with alanine completely abolishes its activity, while mutant Y863F still retains $0.060 \pm 0.008\%$ activity. Some R domains from bacteria can perform an additional reduction step to generate alcohols from aldehyde intermediates (Li, et al., 2008). However, no unique peak with a molecular weight matching the putative alcohol product can be found in our *in vitro* and *in vivo* experiments. Based on our mutagenesis study, we propose a NADPH and Tyr-Lys dependent mechanism for the R domain catalyzed reduction of an aromatic thioester, in which the Tyr residue acts as a catalytic acid/base while the Lys residue can enhance its activity by decreasing the tyrosine hydroxyl pKa (Kavanagh, et al., 2008).

Bioinformatics Analysis of NRPS-like Gene Containing Clusters

To assess the prevalence of NRPS-like genes, we performed genome mining of all 85 available fungal genomes in the NCBI GenBank database. Using the ATEG_03630 gene as a template, over one hundred NRPS-like genes with sequence identity of between 27% and 40% were identified. To shed some light on the putative substrates of these NRPS-like proteins, phylogenetic analysis of the A domain was performed. Given our discovery that the NRPS-like protein can cooperate with the NR-PKS protein to produce an aryl-aldehyde, we examined each NRPS-like gene on its genome to determine if any PKS gene is in its vicinity. Interestingly, almost all NRPS-like genes with a NR-PKS gene in its vicinity were clustered together and form a well-defined clade, which indicates that they may accept similar substrates (Figure S1J). In addition, it is intriguing to find out that some of the NRPS-like genes are accompanied by highly-reducing PKS (HR-PKS) genes, which suggests that aldehydes of highly-reduced polyketides may also be generated in such an activation-reduction mechanism. The large number and high diversity of such PKSs and

NRPS-like gene-containing clusters indicate the high popularity of such aldehyde formation mechanism throughout the fungi kingdom.

Detailed bioinformatics analysis (Figure 2A and Table S2) of gene clusters containing the NR-PKS and NRPS-like genes reveals that they share a core set of genes besides the NR-PKS and NRPS-like genes, albeit that each cluster has its own unique feature. For example, the PKS gene ATEG_07379 from *A. terreus* does not contain the MT and TE domains, which indicates that another product releasing mechanism, such as an α/β hydrolase in the cluster, is required for the proper function of the PKS (Awakawa, et al., 2009; Li, et al., 2011). Apparently, the aryl-aldehyde **5** from the two-step biosynthesis we characterized so far is unlikely the final product of this pathway. There is no literature report on what kind of final product would be produced from this pathway in *A. terreus*. However, the phytotoxin cichorine is strongly linked to the PKS AN6448.2 (Figure 2A) from *A. nidulans* by genome-wide deletion study (Sanchez, et al., 2011). It is possible that our pathway from *A. terreus* may produce a compound that is similar to cichorine. However, there is no obvious candidate gene to incorporate the nitrogen atom in both gene clusters. Full elucidation of this biosynthetic pathway is in progress.

Native Host *A. terreus* does not Produce Compound **4** or **5**

To further support our previous result that the putative NRPS-like gene-containing cluster in *A. terreus* is cryptic, we tried to determine whether native host *A. terreus* can produce compound **4** and **5**. *A. terreus* was grown in YPAD for four days (the same conditions as the qPCR experiment). After extraction, HPLC-MS analysis was performed. As a result, no compound **4** or **5** was detected (Figure S1K). In addition, compound **4** was fed to *A. terreus* to determine whether the native host can convert **4** to **5**. HPLC-MS analysis indicates that no compounds **4** or **5** were detected after four days. It is likely that compound **4** was metabolized by *A. terreus* (Figure S1R) (Skellam, et al., 2010).

Characterization of the ACP Domains of PKSs

Tandem ACPs are rare but not an unprecedented feature in PKSs. The function of tandem ACPs from bacterial type I PKSs and polyunsaturated FASs have been investigated, which can be involved in recruiting trans-acting enzymes, β -branch formation and increasing metabolite production level (Gulder, et al., 2011). Only one fungal NR-PKS with tandem ACPs has been characterized (Fujii, et al., 2001). Each ACP was demonstrated to be self-sufficient to support the polyketide production. However, no quantitative study was performed to assess the functional equivalency of tandem ACPs in fungal PKSs. Therefore, we mutated the serine to alanine individually via the DNA assembler method to obtain single mutants ATEG_03629m-S1651A and ATEG_03629m-S1772A. Quantitative analysis of the 5-MOA production of wild type and both single mutants indicates each mutant can still produce 5-MOA but in a greatly reduced level (Supplementary Experimental Procedures). ATEG_03629m-S1651A and ATEG_03629m-S1772A only produce $8.3 \pm 0.3\%$ and $13.9 \pm 0.1\%$ 5-MOA, respectively, compared to WT. These results are different from the study of polyunsaturated FAS (Jiang, et al., 2008), in which mutating ACPs one by one only incrementally decreased the production level. The production level is still maintained at $\sim 37\%$ even when only one out of six ACPs remains. Although both polyunsaturated FAS and fungal PKS are iterative synthases which may share common structural features, the tandem ACPs of fungal NR-PKS seems to be more cooperative and only partially equivalent.

DISCUSSION

Structurally speaking, a NRPS-like gene containing A-ACP-R domains is not unique. It is known that the amino adipate reductase in the fungal lysine biosynthesis pathway (Ehmann, et al., 1999; Zabriskie and Jackson, 2000) has a similar domain structure, although their A domains have very low homology due to the difference in the substrate structure. On the other hand, the aryl-aldehyde oxidoreductase from bacteria *Nocardia* sp. also has a similar domain structure and can reduce an aryl-acid to an aldehyde (Venkitasubramanian, et al., 2007). It has broad substrate specificity and is believed to be involved in primary metabolism (Li and Rosazza, 1997). A terminal R domain involved reductive release mechanism is well-documented in polyketide and non-ribosomal peptide biosynthesis (Du and Lou, 2010; Qiao, et al., 2011). Only recently, a NRPS-like gene LnaA from *Aspergillus flavus* has been proposed to be involved in reducing tyrosine to its corresponding aldehyde (Forseth, et al., 2013). However, without isolating the aldehyde product, its exact mechanism remains unclear. To the best of our knowledge, such an aryl-aldehyde formation mechanism in the biosynthesis of polyketides has never been reported. All previously characterized aryl-aldehydes are generated via the R domain of NR-PKSs (Ahuja, et al., 2012; Sanchez, et al., 2012). At first glance, this activation-reduction mechanism seems redundant and energy-wasting compared to the typical direct reduction mechanism. However, given the relatively reactive nature of aldehyde compounds, the additional step of activation-reduction may be a good regulation node for the proper timing of aldehyde production in response to environmental stimuli.

In conclusion, we have successfully cloned two “intron-less” cryptic PKS and NRPS-like genes directly from fungal genomic DNA of *A. terreus*. By heterologously expressing these genes individually or in combination in *S. cerevisiae*, we discovered a distinct mechanism for aryl-aldehyde formation in polyketide biosynthesis. Bioinformatics study indicates that such a mechanism may be widely used in fungal secondary metabolite biosynthetic pathways. Our mutagenesis and *in vitro* study provide the first direct investigation of the structure and function of sub-type II NRPS-like gene. In addition, we demonstrated that the combination of synthetic biology tools such as the DNA assembler method and heterologous expression resulted in the identification of compounds from a cryptic pathway. This strategy may enable the discovery of new secondary metabolites from cryptic biosynthetic pathways identified in sequenced genomes and metagenomes (Cobb and Zhao, 2012; Shao, et al., 2011; Shao, et al., 2013; Luo, et al. 2013).

SIGNIFICANCE

By cloning and heterologous expression of two cryptic NR-PKS and NRPS-like genes from *A. terreus* in *S. cerevisiae*, we discovered a distinct mechanism for aryl-aldehyde formation in which a NRPS-like protein activates and reduces an aryl-acid produced by the accompanying NR-PKS to an aryl-aldehyde. Bioinformatics study indicates such a mechanism may be widely used throughout the fungi kingdom. In addition, we provide the first detailed characterization of the structure and function of the NRPS-like protein. Finally, we demonstrated that the combination of synthetic biology tools with heterologous expression may enable the discovery of novel secondary metabolites from cryptic biosynthetic pathways identified in sequenced genomes and metagenomes.

EXPERIMENTAL PROCEDURES

Materials and Reagents

Plasmids pRS414 and pRS416 were purchased from New England Biolabs (Beverly, MA). Failsafe™ 2× premix buffer G was purchased from EPICENTRE Biotechnologies (Madison,

WI). Synthetic complete drop-out medium lacking uracil (SC-Ura), tryptophan (SC-Trp) or both (SC-Trp-Ura) from MP Biomedicals (Solon, OH) were used to select *S. cerevisiae* transformants containing the assembled plasmids. *S. cerevisiae* HZ848 (*MAT α* , *ade2-1*, *Δura3*, *his3-11, 15*, *trp1-1*, *leu2-3, 112*, and *can1-100*) was used as the host for DNA assembly and heterologous expression. *S. cerevisiae* strain BJ5464-NpgA (*MAT α* *ura3-52* *his3-Δ200* *leu2- Δ1* *trp1* *pep4::HIS3* *prb1 Δ1.6R* *can1 GAL*) was used as the host for protein over-expression.

Plasmid Construction

Cloning of ATEG_03629 is described here as an example for all our plasmid construction using the DNA assembler method. See supporting information for details of other plasmid construction. PCR was used to amplify each of the three exons that in combination encode the ATEG_03629 gene from the genomic DNA of *A. terreus* using the corresponding primers listed in Table S1. Helper plasmid pRS414-TEF1p-HXT7t was linearized by *XhoI* digestion. The full ATEG_03629 gene under the control of the TEF1p promoter was assembled via the DNA assembler method (Shao and Zhao, 2009) to obtain the plasmid pRS414-ATEG_03629.

Quantitative PCR Analysis

A. terreus was inoculated to 4 ml of YPAD medium and grow at 30°C and 250 rpm for 96h. The total RNA was isolated using the RNeasy plant mini kit (QIAGEN, Valencia, CA, USA) following the manufacturer's instructions. Genomic DNA was removed using Ambion® TURBO DNA-free kit (Life Technologies, Grand Island, NY, USA) following the manufacturer's instructions. The RNA samples were then reverse transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis kit using oligo-dT primer following the manufacturer's instructions (Roche, Indianapolis, IN, USA). The qPCR experiments were carried out using the LightCycler® 480 system (Roche, Indianapolis, IN, USA) using the SYBR Green-based method following the manufacturer's instructions. Primers used in qPCR experiments were listed in Table S1.

Heterologous Expression in *S. cerevisiae*

Each expression vector was co-transformed with pRS414-GPM1p-npgA-GPM1t to *S. cerevisiae* strain HZ848 using the lithium acetate/single stranded carrier DNA/polyethylene glycol (PEG) method (Gietz and Woods, 2002). Transformants were grown in SC-Trp-Ura dropout media for either HPLC-ESI-MS analysis or product purification.

HPLC-ESI-MS Analysis

Yeast transformants were grown in 3 ml SC-Trp-Ura dropout media for two days. Culture broth was used directly for HPLC-ESI-MS analysis. HPLC-ESI-MS was performed on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer (Agilent, Palo Alto, CA) with a reverse-phase kinetex C18 column (Phenomenex, Torrance, CA). HPLC parameters for detection of compounds **4** and **5** were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 100% B in 20 min, maintain at 100% B for 10 min, return and maintain at 10% B for 7 min; flow rate 0.3 ml/min; detection by UV spectroscopy at 300 nm or 330nm. Under such conditions, compounds **4** and **5** are eluted at 13.1 min and 15.6 min, respectively. Mass spectra were acquired in ultra scan mode using electrospray ionization (ESI) with positive/negative polarity. The MS system was operated using a drying temperature of 350 °C, a nebulizer pressure of 35 psi, a drying gas flow of 9 l/min, and a capillary voltage of 4500 V.

Protein Expression and Purification from *S. cerevisiae*

Each expression vector was transformed to *S. cerevisiae* strain BJ5464-NpgA using the lithium acetate/single stranded carrier DNA/polyethylene glycol (PEG) method (Gietz and Woods, 2002). Transformants were grown in 20 ml of SC-Ura dropout media for two days and inoculate to 1L of YPAD medium. Cells were grown at 30 °C and 250 rpm for 72 hours. The cells were harvested by centrifugation (4000 g, 15 minutes, 4 °C), resuspended in 30 ml lysis buffer (50 mM NaH₂PO₄ pH 8.0, 0.15 M NaCl, 10 mM imidazole) and lysed through sonication on ice. His-tagged proteins were purified by using Ni-NTA Agarose (QIAGEN, Valencia, CA), following the supplied protocols. The cleared cell lysate following sonication and centrifugation was directly applied onto a column which was packed with Ni-NTA Agarose. After instructed washing steps, the protein was eluted with the elution buffer containing 250 mM imidazole. Purified proteins were concentrated and buffer exchanged into storage buffer (50 mM Tris-HCl, 100 mM NaCl, pH=7.9) containing 10% glycerol. The concentrated enzyme solutions were aliquoted and flash frozen. Protein concentrations were determined with the Bradford assay (BioRad) using BSA as a standard.

In vitro Characterization of ATEG_03630 and its Mutants

For *in vitro* enzymatic assays, the final concentrations of enzymes were 1 μM, 1 mM 5-MOA as substrate, with cofactor concentrations as 2 mM NADPH, 10 mM ATP and 10 mM MgCl₂. The assays were carried out in 50 mM Tris-HCl pH 8.5 buffer. The reaction mixtures were incubated at room temperature.

A typical volume of the reaction is 200 μl. 20 μl of reaction mixture was taken out at various time points and quenched with HCl. Each reaction mixture was used for HPLC-ESI-MS analysis. HPLC-ESI-MS was performed on an Agilent 1100 series LC/MSD XCT plus ion trap mass spectrometer (Agilent, Palo Alto, CA) with a reverse-phase kinetex C18 column (Phenomenex, Torrance, CA). HPLC parameters for detection of compound **5** were as follows: solvent A, 0.1% formic acid in water; solvent B, 0.1% formic acid in acetonitrile; gradient, 0% B to 100% B in 20 min, maintain at 100% B for 10 min, return and maintain at 10% B for 7 min; flow rate 0.3 ml/min; detection by UV spectroscopy at 330 nm. The amount of product **5** was quantified by area integration of the UV peak at 330 nm. A standard curve was generated using isolated compound **5** with the same HPLC condition.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Academies Keck Futures Initiative on Synthetic Biology and National Institutes of Health (GM077596). We thank Professor Y. Tang for providing *S. cerevisiae* strain BJ5464-NpgA.

References

- Ahuja M, Chiang YM, Chang SL, Praseuth MB, Entwistle R, Sanchez JF, Lo HC, Yeh HH, Oakley BR, Wang CC. Illuminating the Diversity of Aromatic Polyketide Synthases in *Aspergillus nidulans*. *J Am Chem Soc.* 2012; 134:8212–8221. [PubMed: 22510154]
- Awakawa T, Yokota K, Funa N, Doi F, Mori N, Watanabe H, Horinouchi S. Physically discrete beta-lactamase-type thioesterase catalyzes product release in atrochryson synthesis by iterative type I polyketide synthase. *Chem Biol.* 2009; 16:613–623. [PubMed: 19549600]
- Brakhage AA, Schroeckh V. Fungal secondary metabolites - strategies to activate silent gene clusters. *Fungal Genet Biol.* 2011; 48:15–22. [PubMed: 20433937]

- Chen L, Liu W, Huang K, Hu X, Fang ZX, Wu JL, Zhang QQ. Penicitrinols F-I, New Citrinin Derivatives from the Marine-Derived Fungus *Penicillium Citrinum*. *Heterocycles*. 2011; 83:1853–1858.
- Chiang YM, Szewczyk E, Davidson AD, Keller N, Oakley BR, Wang CC. A gene cluster containing two fungal polyketide synthases encodes the biosynthetic pathway for a polyketide, asperfuranone, in *Aspergillus nidulans*. *J Am Chem Soc*. 2009; 131:2965–2970. [PubMed: 19199437]
- Cobb RE, Zhao H. Direct cloning of large genomic sequences. *Nat Biotechnol*. 2012; 30:405–406. [PubMed: 22565964]
- Crosa JH, Walsh CT. Genetics and assembly line enzymology of siderophore biosynthesis in bacteria. *Microbiol Mol Biol Rev*. 2002; 66:223–249. [PubMed: 12040125]
- Davison J, Al Fahad A, Cai M, Song Z, Yehia SY, Lazarus CM, Bailey AM, Simpson TJ, Cox RJ. Genetic, molecular, and biochemical basis of fungal tropolone biosynthesis. *Proc Natl Acad Sci U S A*. 2012; 109:7642–7647. [PubMed: 22508998]
- Du L, Lou L. PKS and NRPS release mechanisms. *Nat Prod Rep*. 2010; 27:255–278. [PubMed: 20111804]
- Ehmann DE, Gehring AM, Walsh CT. Lysine biosynthesis in *Saccharomyces cerevisiae*: mechanism of alpha-aminoadipate reductase (Lys2) involves posttranslational phosphopantetheinylation by Lys5. *Biochemistry*. 1999; 38:6171–6177. [PubMed: 10320345]
- Forseth RR, Amaike S, Schwenk D, Affeldt KJ, Hoffmeister D, Schroeder FC, Keller NP. Homologous NRPS-like gene clusters mediate redundant small-molecule biosynthesis in *Aspergillus flavus*. *Angew Chem Int Ed Engl*. 2013; 52:1590–1594. [PubMed: 23281040]
- Fujii I, Watanabe A, Sankawa U, Ebizuka Y. Identification of Claisen cyclase domain in fungal polyketide synthase WA, a naphthopyrone synthase of *Aspergillus nidulans*. *Chem Biol*. 2001; 8:189–197. [PubMed: 11251292]
- Galagan JE, Calvo SE, Cuomo C, Ma LJ, Wortman JR, Batzoglou S, Lee SI, Basturkmen M, Spevak CC, Clutterbuck J, et al. Sequencing of *Aspergillus nidulans* and comparative analysis with *A. fumigatus* and *A. oryzae*. *Nature*. 2005; 438:1105–1115. [PubMed: 16372000]
- Gietz RD, Woods RA. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol*. 2002; 350:87–96. [PubMed: 12073338]
- Gressler M, Zaehle C, Scherlach K, Hertweck C, Brock M. Multifactorial induction of an orphan PKS-NRPS gene cluster in *Aspergillus terreus*. *Chem Biol*. 2011; 18:198–209. [PubMed: 21236704]
- Gulder TA, Freeman MF, Piel J. The Catalytic Diversity of Multimodular Polyketide Synthases: Natural Product Biosynthesis Beyond Textbook Assembly Rules. *Top Curr Chem*. 2011:1–53.
- Ishichi K, Nakazawa T, Ookuma T, Sugimoto S, Sato M, Tsunematsu Y, Ishikawa N, Noguchi H, Hotta K, Moriya H, et al. Establishing a new methodology for genome mining and biosynthesis of polyketides and peptides through yeast molecular genetics. *Chembiochem*. 2012; 13:846–854. [PubMed: 22447505]
- Jiang H, Zirkle R, Metz JG, Braun L, Richter L, Van Lanen SG, Shen B. The role of tandem acyl carrier protein domains in polyunsaturated fatty acid biosynthesis. *J Am Chem Soc*. 2008; 130:6336–6337. [PubMed: 18444614]
- Kavanagh KL, Jornvall H, Persson B, Oppermann U. The SDR superfamily: functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol Life Sci*. 2008; 65:3895–3906. [PubMed: 19011750]
- Li T, Rosazza JP. Purification, characterization, and properties of an aryl aldehyde oxidoreductase from *Nocardia* sp strain NRRL 5646. *J Bacteriol*. 1997; 179:3482–3487. [PubMed: 9171390]
- Li Y, Chooi YH, Sheng Y, Valentine JS, Tang Y. Comparative characterization of fungal anthracenone and naphthacenedione biosynthetic pathways reveals an alpha-hydroxylation-dependent Claisen-like cyclization catalyzed by a dimanganese thioesterase. *J Am Chem Soc*. 2011; 133:15773–15785. [PubMed: 21866960]
- Li Y, Weissman KJ, Muller R. Myxochelin biosynthesis: direct evidence for two- and four-electron reduction of a carrier protein-bound thioester. *J Am Chem Soc*. 2008; 130:7554–7555. [PubMed: 18498160]
- Luo Y, Huang H, Liang J, Wang M, Lu L, Shao Z, Cobb RE, Zhao H. Activation and characterization of a cryptic polycyclic tetramate macrolactam biosynthetic gene cluster. *Nat Commun*. accepted.

- Persson B, Kallberg Y, Oppermann U, Jornvall H. Coenzyme-based functional assignments of short-chain dehydrogenases/reductases (SDRs). *Chem Biol Interact.* 2003; 143–144:271–278.
- Qiao K, Zhou H, Xu W, Zhang W, Garg N, Tang Y. A fungal nonribosomal peptide synthetase module that can synthesize thiopyrazines. *Org Lett.* 2011; 13:1758–1761. [PubMed: 21384820]
- Rausch C, Weber T, Kohlbacher O, Wohlleben W, Huson DH. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). *Nucleic Acids Res.* 2005; 33:5799–5808. [PubMed: 16221976]
- Sanchez JF, Entwistle R, Hung JH, Yaegashi J, Jain S, Chiang YM, Wang CC, Oakley BR. Genome-based deletion analysis reveals the prenyl xanthone biosynthesis pathway in *Aspergillus nidulans*. *J Am Chem Soc.* 2011; 133:4010–4017. [PubMed: 21351751]
- Sanchez JF, Somoza AD, Keller NP, Wang CC. Advances in *Aspergillus* secondary metabolite research in the post-genomic era. *Nat Prod Rep.* 2012; 29:351–371. [PubMed: 22228366]
- Schroeckh V, Scherlach K, Nutzmann HW, Shelest E, Schmidt-Heck W, Schuemann J, Martin K, Hertweck C, Brakhage AA. Intimate bacterial-fungal interaction triggers biosynthesis of archetypal polyketides in *Aspergillus nidulans*. *Proc Natl Acad Sci U S A.* 2009; 106:14558–14563. [PubMed: 19666480]
- Shao Z, Rao G, Li C, Abil Z, Luo Y, Zhao H. Refactoring the silent spectinabilin gene cluster using a plug-and-play scaffold. *ACS Syn Bio.* 2013 in press.
- Shao Z, Luo Y, Zhao H. Rapid characterization and engineering of natural product biosynthetic pathways via DNA assembler. *Mol Biosyst.* 2011; 7:1056–1059. [PubMed: 21327279]
- Shao Z, Zhao H. DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways. *Nucleic Acids Res.* 2009; 37:e16. [PubMed: 19074487]
- Skellam EJ, Hurley D, Davison J, Lazarus CM, Simpson TJ, Cox RJ. Mutation of key residues in the C-methyltransferase domain of a fungal highly reducing polyketide synthase. *Mol Biosyst.* 2010; 6:680–682. [PubMed: 20237645]
- Venkatasubramanian P, Daniels L, Rosazza JP. Reduction of carboxylic acids by *Nocardia* aldehyde oxidoreductase requires a phosphopantetheinylated enzyme. *J Biol Chem.* 2007; 282:478–485. [PubMed: 17102130]
- Wackler B, Lackner G, Chooi YH, Hoffmeister D. Characterization of the *Suillus grevillei* quinone synthetase GreA supports a nonribosomal code for aromatic alpha-keto acids. *Chembiochem.* 2012; 13:1798–1804. [PubMed: 22730234]
- Wattanachaisaareekul S, Lantz AE, Nielsen ML, Andresson OS, Nielsen J. Optimization of heterologous production of the polyketide 6-MSA in *Saccharomyces cerevisiae*. *Biotechnol Bioeng.* 2007; 97:893–900. [PubMed: 17171715]
- Yakasai AA, Davison J, Wasil Z, Halo LM, Butts CP, Lazarus CM, Bailey AM, Simpson TJ, Cox RJ. Nongenetic reprogramming of a fungal highly reducing polyketide synthase. *J Am Chem Soc.* 2011; 133:10990–10998. [PubMed: 21675761]
- Yeh HH, Chiang YM, Entwistle R, Ahuja M, Lee KH, Bruno KS, Wu TK, Oakley BR, Wang CC. Molecular genetic analysis reveals that a nonribosomal peptide synthetase-like (NRPS-like) gene in *Aspergillus nidulans* is responsible for microperforanone biosynthesis. *Appl Microbiol Biotechnol.* 2012; 96:739–748. [PubMed: 22627757]
- Zabriskie TM, Jackson MD. Lysine biosynthesis and metabolism in fungi. *Nat Prod Rep.* 2000; 17:85–97. [PubMed: 10714900]

Highlights

- Discovery of a distinct mechanism for aryl-aldehyde generation in polyketides
- An NRPS-like protein can activate and reduce an aryl-acid to an aryl-aldehyde
- Such a distinct mechanism may be widely used throughout the fungi kingdom.

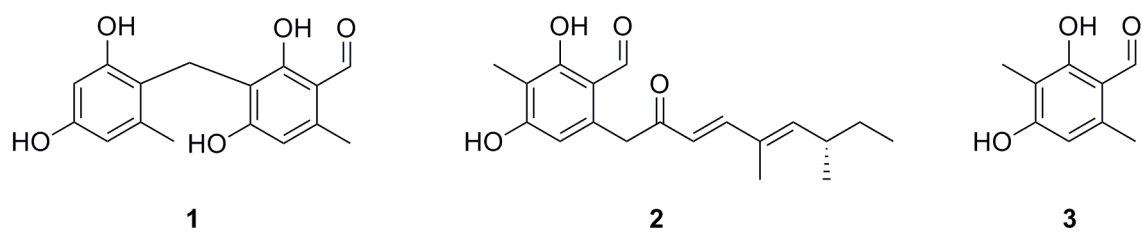


Figure 1. Structures of Selected Aryl-aldehyde Intermediates or Final Products from Fungal Polyketide Biosynthetic Pathways.

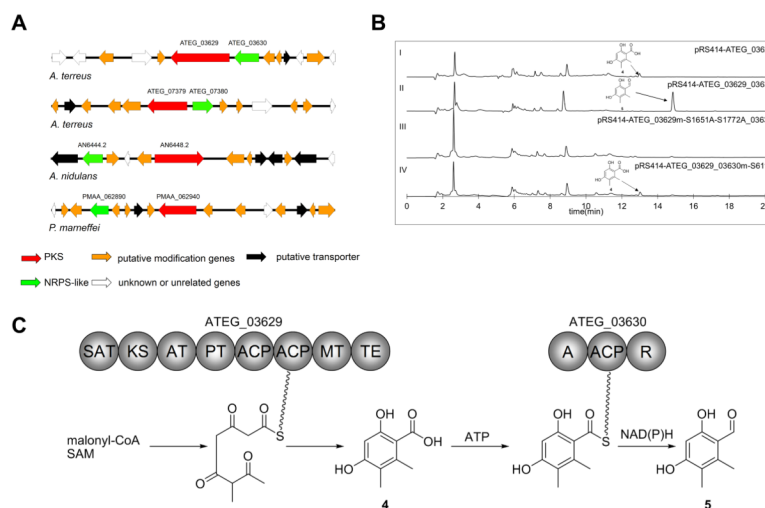


Figure 2. Characterization of NRPS-like Gene Containing Cluster via Heterologous Expression in *S.cerevisiae*

(A) Gene Organization of Similar Clusters that Contain NR-PKS and NRPS-like Genes (Detailed function prediction of each gene is provided in the Supporting Information).

(B) HPLC Analysis (UV = 300 nm) of the Supernatants of *S. cerevisiae* Strains Harboring Plasmid pRS414-ATEG_03629 (I), pRS414-ATEG_03629_03630 (II), pRS414-ATEG_03629m-S1651A-S1772A_03630 (III), and pRS414-ATEG_03629_03630m-S611A (IV).

(C) Proposed Mechanism for Aryl-aldehyde Biosynthesis. SAM=S-Adenosyl methionine.