Biochemistry Supporting Information

Anthranilate-Activating Modules from Fungal Nonribosomal Peptide Assembly Lines

Brian D. Ames and Christopher T. Walsh

Department of Biological Chemistry and Molecular Pharmacology

Harvard Medical School, Boston, MA 02115

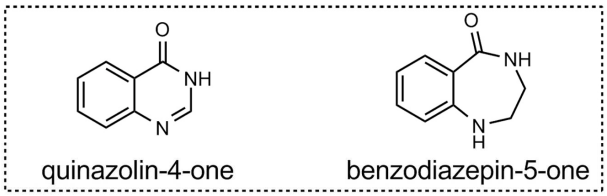
Supporting Discussion

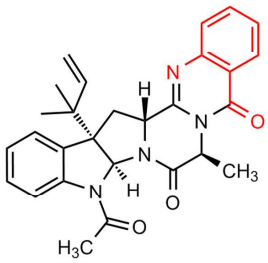
Bioinformatics analysis suggests that Orf12080 may be involved in fumiquinazoline biosynthesis. The fungus *A. fumigatus* Af293 is a known producer of the Ant-containing alkaloid fumiquinazoline A (*1*), a member of a group of fumiquinazoline compounds originally isolated from a strain of *A. fumigatus* separated from the gastrointestinal tract of the marine fish *Pseudolabrus japonicas* (*2*). The *A. fumigatus* Af293 genome contains fourteen genes identified as putative nonribosomal peptide synthetases (*3*). Of these fourteen, Orf12080 (also designated *pesM* (*3, 4*)) is part of an eight-gene cluster that we predict to contain the necessary components for fumiquinazoline biosynthesis including a single-module NRPS (Orf12050), two putative FAD-dependent oxygenases (Orfs 12060 and 12070), a putative N-acyltransferase (Orf12100), and a gene encoding a predicted anthranilate synthase (Orf12110) (Figure S6).

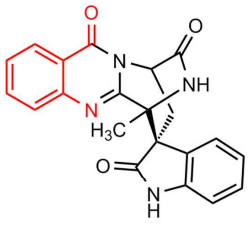
In this work we demonstrate that module 1 of Orf12080 selectively activates anthranilate as Ant-AMP and loads Ant onto the T-domain of module 1 as a phosphopantetheine-tethered product. Reconstitution of fumiquinazoline F biosynthesis by Orf12080 would required that module 2 selectively load and epimerize Trp, while module 3 select for Ala. The predicted specificity sequence of module 2 is 70% identical to the Trp-selective A-domain of TxtB (thaxtomin synthetase B) (*5*); and the module 3 specificity sequence is homologous to the Ala-specific module 3 of Tex1 (peptaibol synthetase from *Trichoderma virens*) (*6*). These findings are fully consistent with the three module Orf12080 as the Ant-Trp-Ala assembly line for production of fumiquinazoline F. The presence of an E-domain in module 2 fits with the (*R*)-stereogenic center at the C_α of the Trp residue which would likely arise from epimerization of the tryptophanyl moiety at the dipeptidyl stage. The terminal C-domain may have cyclase activity and/or diketopiperazine formation may drive spontaneous cyclization and chain release (*7*): indeed, an Ant_1-Trp_2-Ala_3-tethered intermediate would have to undergo double cyclization to morph the linear tripeptidyl intermediate to the tricyclic framework of fumiquinazoline F.

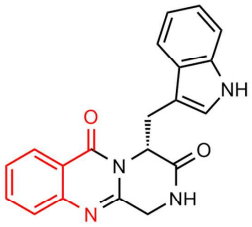
The Orfs neighboring Orf12080 are proposed to catalyze the tailoring reactions for the conversion of fumiquinazoline F into fumiquinazoline A. Orf12050 is a free standing A-T-C three-domain NRPS module; the A-domain of Orf12050 possesses a 10-residue specificity sequence that shares 80% similarity to the alanine specific A-domain from module 11 of cyclosporin A synthethase (*8*). This A-domain could be the Ala-activating enzyme for acylation of the indole nitrogen of fumiquinazoline F. Orf12060 and 12070 are predicted FAD-dependent oxygenases, one of which may act as the 3-indole monooxygenase involved in epoxidation of the pyrrole for subsequent intramolecular cyclization of the *N*-aminoacyl-indolyl. Orf12100 is a predicted N-acyltransferase which, along with Orf12050, should couple alanine to the indole N1. Finally, Orf12110 is a predicted anthranilate synthase which is just the enzyme needed to ensure an adequate supply of the Ant building block when this conditional natural product pathway is turned on.

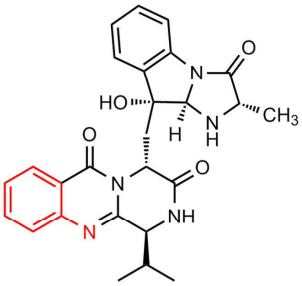
Figure S1. Additional Ant-containing fungal alkaloids discussed in this work. The bicyclic quinazolinone and benzodiazepinone scaffold elements common to this group of natural products are boxed and provided for reference. The top five compounds (*N*-acetylardeemin to tryptoquivaline) are produced by strains of fungi that share genus and species classification with fungi whose genomes have been sequenced.

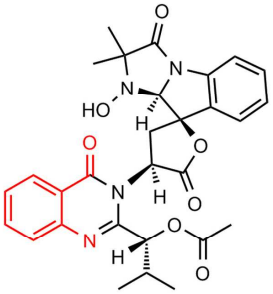


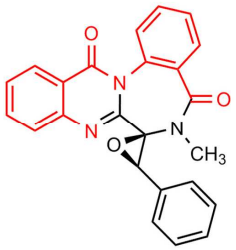

5-*N*-acetylardeemin

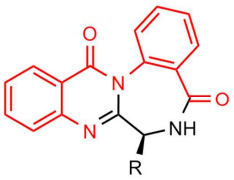

alantrypinone


glyantrypine

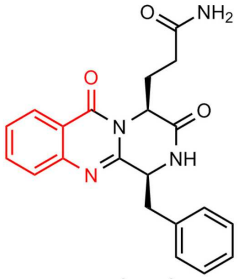

fiscalin A

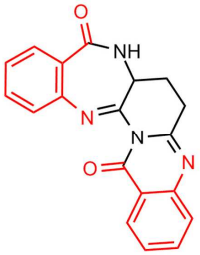

tryptoquivaline

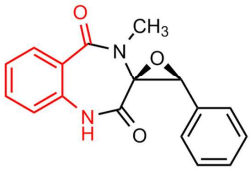

benzomalvin C


circumdatin F
(R = CH_3)

sclerotigen
(R = H)


verrucine A


auranthine


cyclopenin

(note: "cyclopeptine" is the same molecule without the expoxide)

Figure S2. Co-purification of *E. coli* chaperones with the three C*AT target proteins from Ni-affinity chromatography. The gel images represent pooled Ni-NTA elutions containing the indicated target protein (blue arrow or label); the level of purity illustrated was consistently observed from batch-to-batch for each target protein. Among the impurities present in the AnaPS Trx-C*AT preparation, peptide mass fingerprinting was used to identify two degradation fragments of full-length protein and three *E. coli* chaperones (red labels). A banding patterns consistent with these same three *E. coli* chaperones is also observed in the Ni-NTA elutions from cells overproducing AFUA_6g12080 C*AT and NFIA_057960 C*AT proteins.

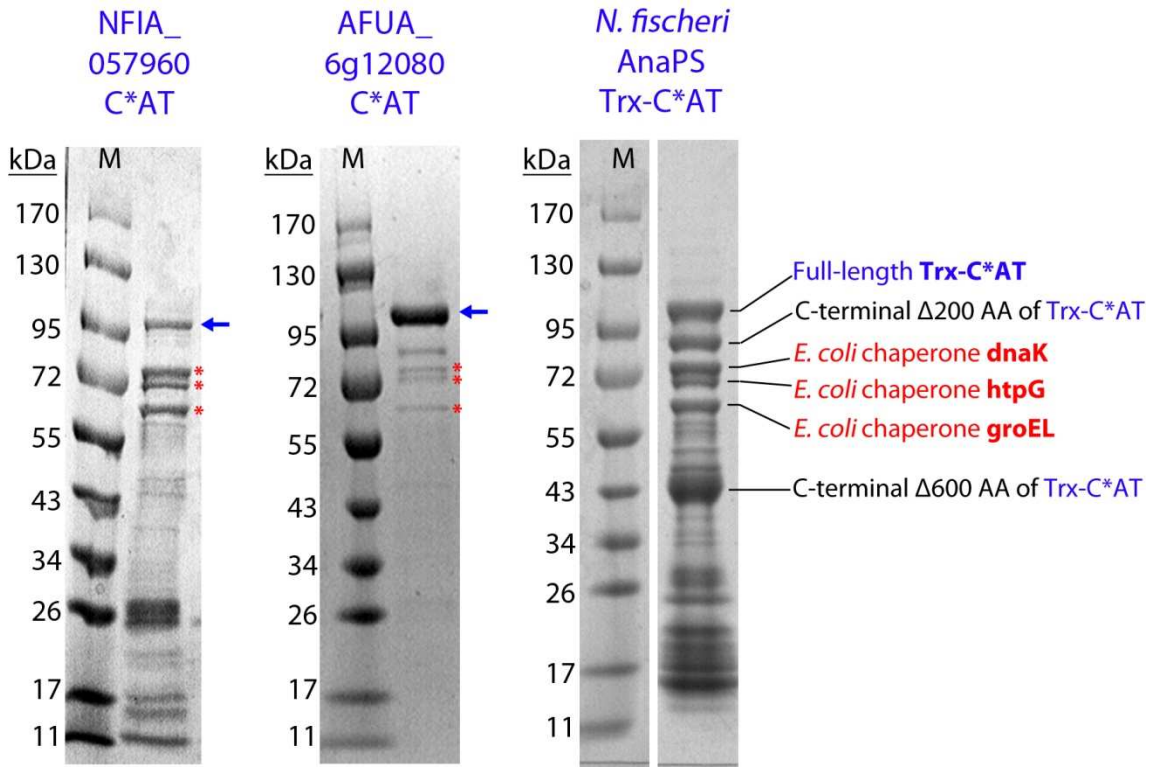


Figure S3. High-resolution LC-MS confirmation of Ant-AMP formation catalyzed by *N. fischeri* AnaPS Trx-C*AT (A), *A. fumigatus* Orf12080 C*AT (B), and *N. fischeri* NFIA_057960 C*AT (C). Left, LC chromatogram showing elution profile of Ant-AMP (V_t 5.5-5.6 min); right, corresponding MS spectrum. Panels (A) and (C) consist of data obtained for unlabeled anthranilate and ATP, whereas panel (B) contains all possible combinations of unlabeled or ^{15}N-labeled anthranilic acid (Ant) and ATP. Reactions were setup in a 50 µL reaction volume according to the HPLC-based assay to include: C*AT protein, 10 mM $MgCl_2$, 1 mM DTT, 50 mM Hepes (pH 7.4), and Ant + ATP (Rxn1), or ^{15}N-Ant + ATP (Rxn2), or Ant + $^{15}N_5$-ATP (Rxn3), or ^{15}N-Ant + $^{15}N_5$-ATP (Rxn4). An asterisk in the MS spectra denotes the AMP mass signal present due to partial product fragmentation during ionization.

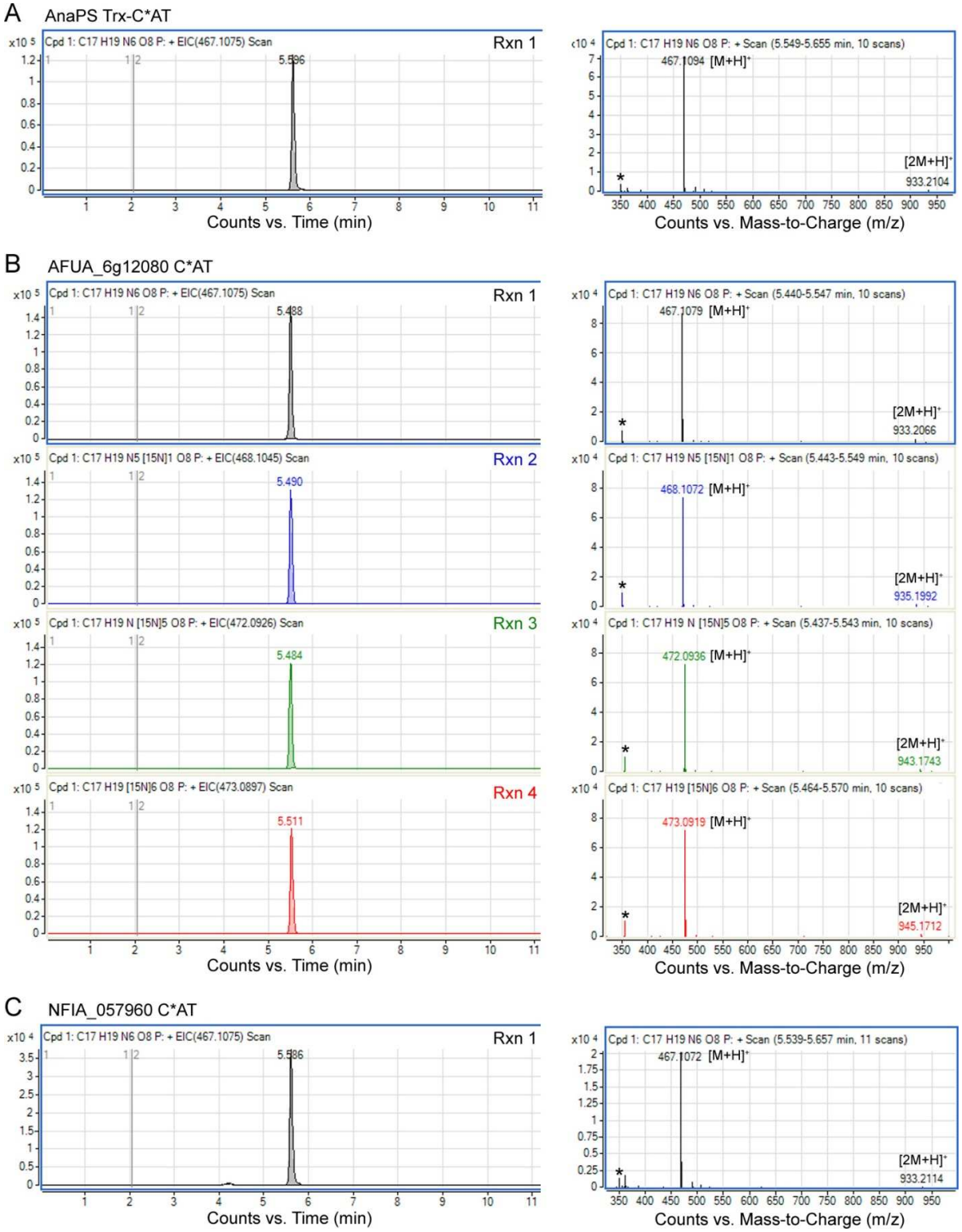


Figure S4. Partial sequence alignment of AnaPS A_1 to seventeen bacterial amino acid activating A-domains from the gramicidin, tyrocidin, syringomycin, and pristinamycin pathways. The aligned 10AA code for AnaPS A_1 (labeled as Pos1-10) precisely matches those residues extracted by the online prediction servers as GALFFAAGVK. Conserved regions neighboring the positions defining the 10AA code are boxed and annotated, and place the extracted specificity-residues of AnaPS A_1 in solid context in support of correct assignment. For example, the eight residues preceding Pos1 of AnaPS A_1 are identical to, or conserved in chemical property to, the majority of the aligned bacterial A-domain sequences. Other blocks of conservation are found as: seven residues following Pos3; five residues pre- and six residues post-Pos4; four residues post-Pos6; three pre- and five post-Pos7; and six post-Pos10. Sequence alignment was performed with ClustalW (*9*), and rendered using GeneDoc (*10*). Shading indicates the level of residue conservation as: red, 100%; green, 75%; and cyan 50% (with "Similarity Groups" enabled). Among strictly conserved residues, the Phe immediately preceding Pos1 and the Gly of block 7 are important for carboxylate interaction and positioning, the Pos10 Lys interacts with both the substrate carboxylate and ATP/AMP, and the Glu of block 7 coordinates the catalytically essential Mg^{2+} ion (*11*).

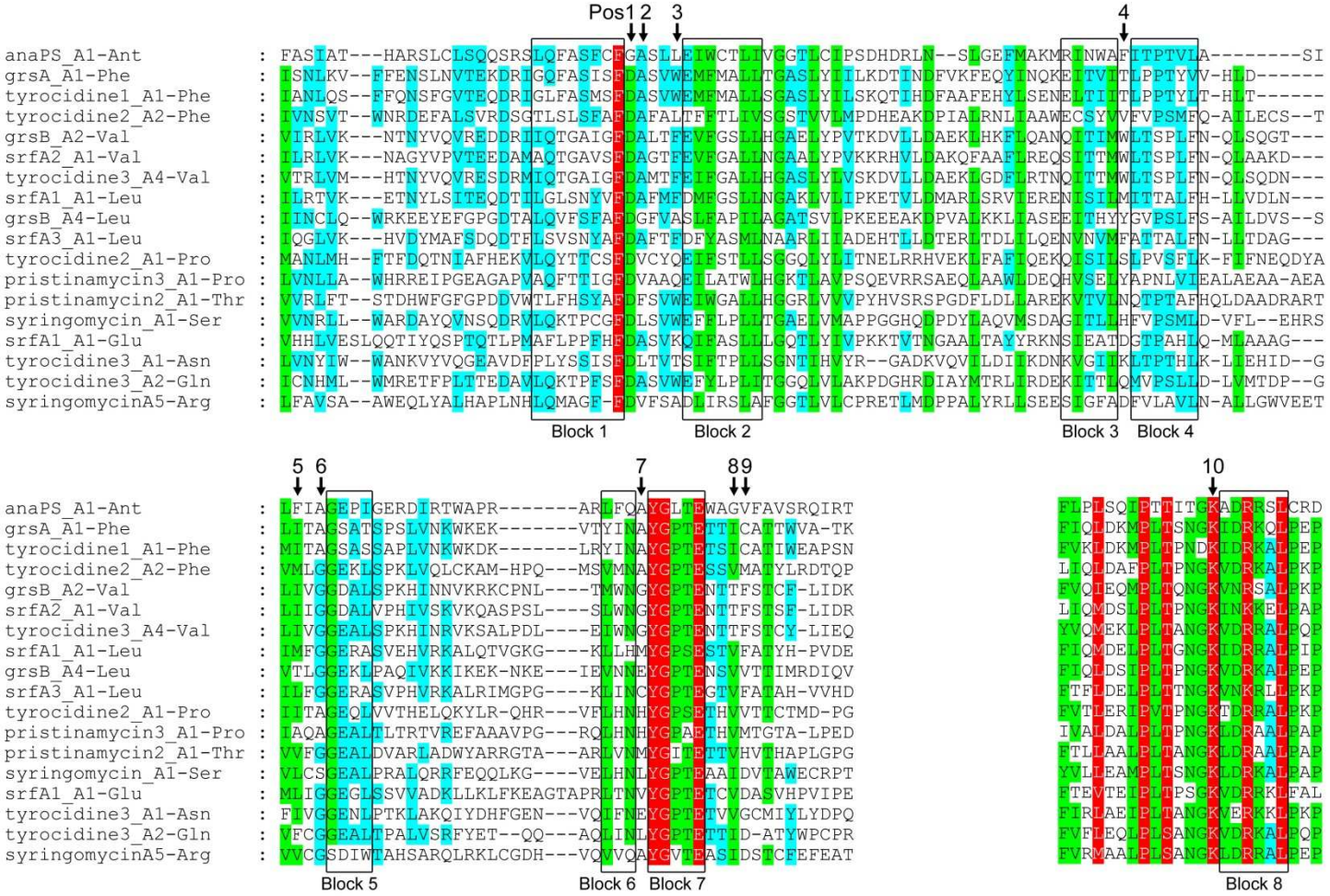


Figure S5. Sequence alignment comparing the substrate-binding/selectivity-determining region of the fungal Ant-activating A-domains identified in this work to select α-amino acid and aryl acid activating NRPS A-domains. GrsA-PheA and SrfAB_A5 activate the L-amino acids Phe and Asp respectively (labeled as "AA" on left side of alignment); DhbE and VibE activate 2,3-dihydroxybenzoic acid; YbtE and MbtA activate salicylic acid (both sets of enzymes grouped under the label "2-OH aryl"); anthramycin Orf21, sibiromycin SibE, and tomaymycin TomA are proposed to activate the substituted anthranilates ("S. Ant") 4-CH_3-3-OH-Ant, 4-CH_3-3,5-OH-Ant, and 4,5-OH-Ant respectively; AnaPS_A1, AFUA_6g12080_A1, and NFIA_057960_A1 are shown in this work to activate anthranilic acid ("Ant"), while the remaining seven sequences are predicted to activate anthranilic acid based on conservation of specificity sequence residues as detailed in the main text. The 10AA code residues proposed to be important for substrate selectivity are highlighted in yellow and numbered according to GrsA-PheA; three other conserved residues that contribute to carboxy-acid substrate binding are highlighted in grey. Secondary structural elements extracted from two NRPS adenylation-domain crystal structures are provided. Alignment was performed using ClustalW with secondary structure annotations generated using ESPript. Sequence alignment shows that the aryl acid-activating A-domains DhbE, VibE, YbtE, and MbtA, possess a core His followed by a Pos1 Asn (rather than the Phe-Asp pair of bacterial AA-activating enzymes or the Phe-Gly pair of AnaPS A_1), as well as conserved insertions/deletions around the specificity-determining residues at Pos3 and Pos8-9 of the 10AA code. These insertions/deletions were not appropriately gapped in the output from ClustalW (resulting in misalignment of the 10AA code); therefore, to accurately reflect these insertions/deletions manual adjustment of the output alignment was necessary and was performed based on superimposition of the DhbE and PheA crystal structures and structure-based sequence alignment.

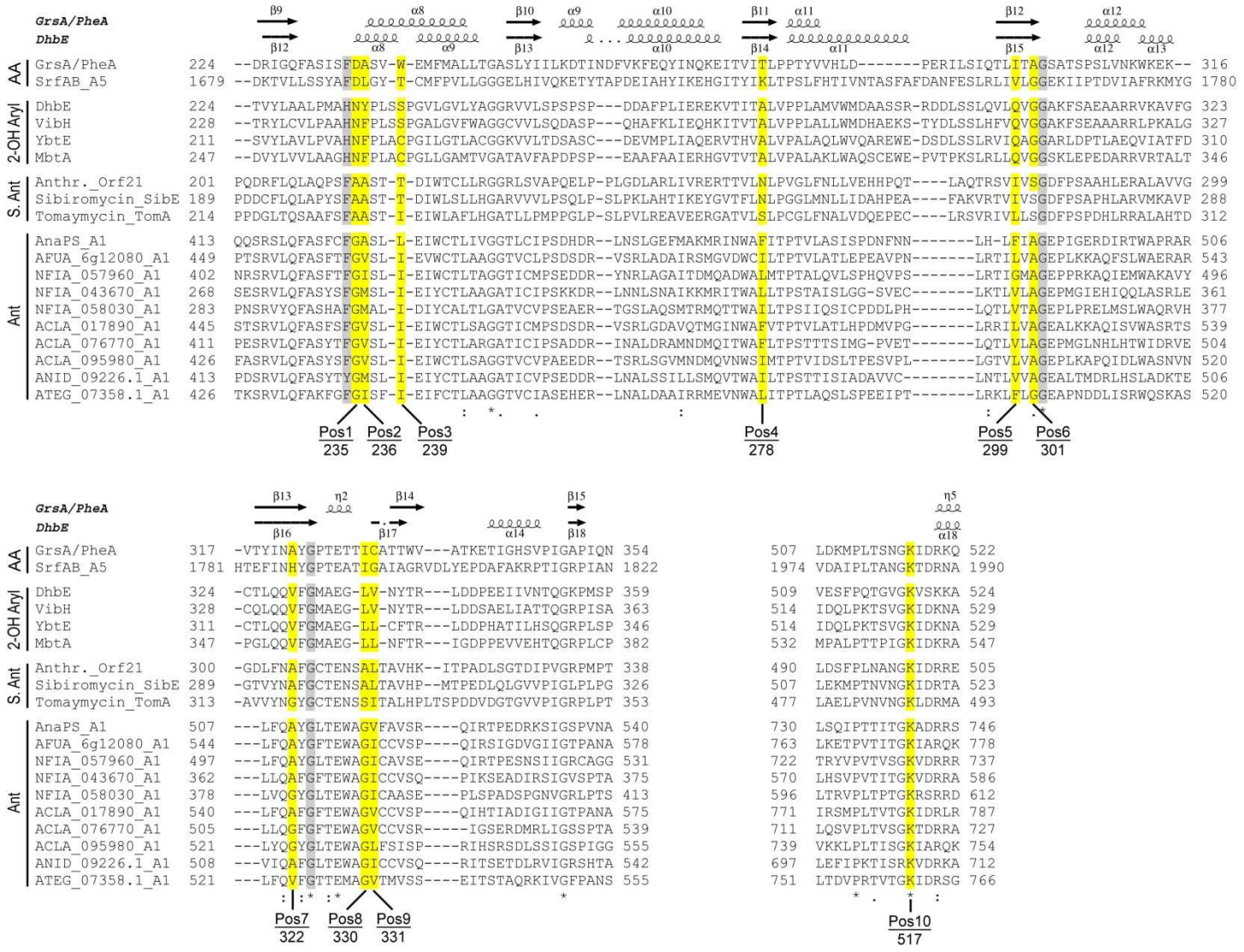


Figure S6. Putative gene cluster and hypothetical biosynthetic route for fumiquinazoline production by *A. fumigatus* Af293 Orfs 12040-12110. (A) Putative fumiquinazoline gene cluster identified by genome mining (introns are marked by vertical black lines). (B) Proposed route of fumiquinazoline biosynthesis illustrating the predicted role of the trimodular NRPS Orf12080 for the activation and loading of Ant, Trp (with subsequent epimerization), and Ala. Activation and loading of anthranilate was experimentally demonstrated in this work, while the selection of other amino acids as Trp and Ala are predicted from bioinformatics analysis (as described in SI Discussion). Transformation of the linear T-domain-tethered tripeptidyl intermediate into the tricyclic scaffold of fumiquinazoline is predicted to occur via two consecutive cyclization events: 1) chain-release via diketopiperazine formation, and 2) attack of the anthranilate 2-amino group on the Ala-derived carbonyl of the diketopiperazine ring. One or both of these cyclizations may be catalyzed by the C-terminal condensation domain. The hypothetical post-NRPS tailoring reactions involving oxidative coupling of alanine to the Trp-indole used to convert fumiquinazoline F to fumiquinazoline A are also detailed.

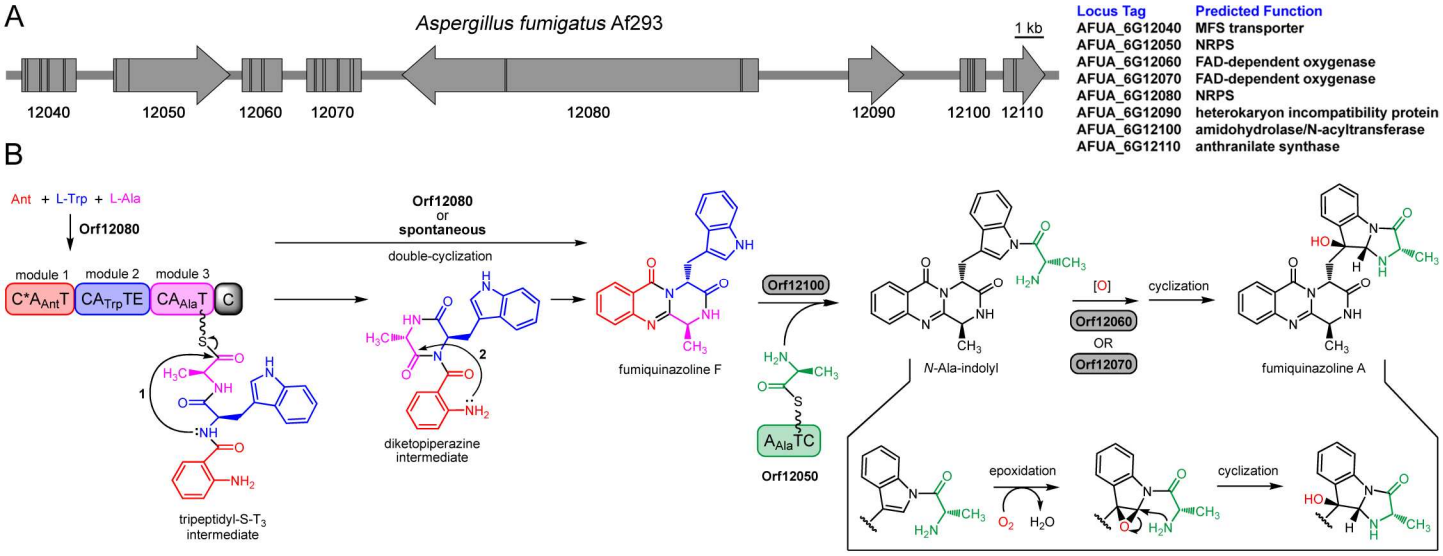


Figure S7. Michaelis-Menten plots constructed using data from the coupled PP_i-release assay for determination of kinetic parameters for AFUA_6g12080 C*AT with various aryl acid substrates and ATP.

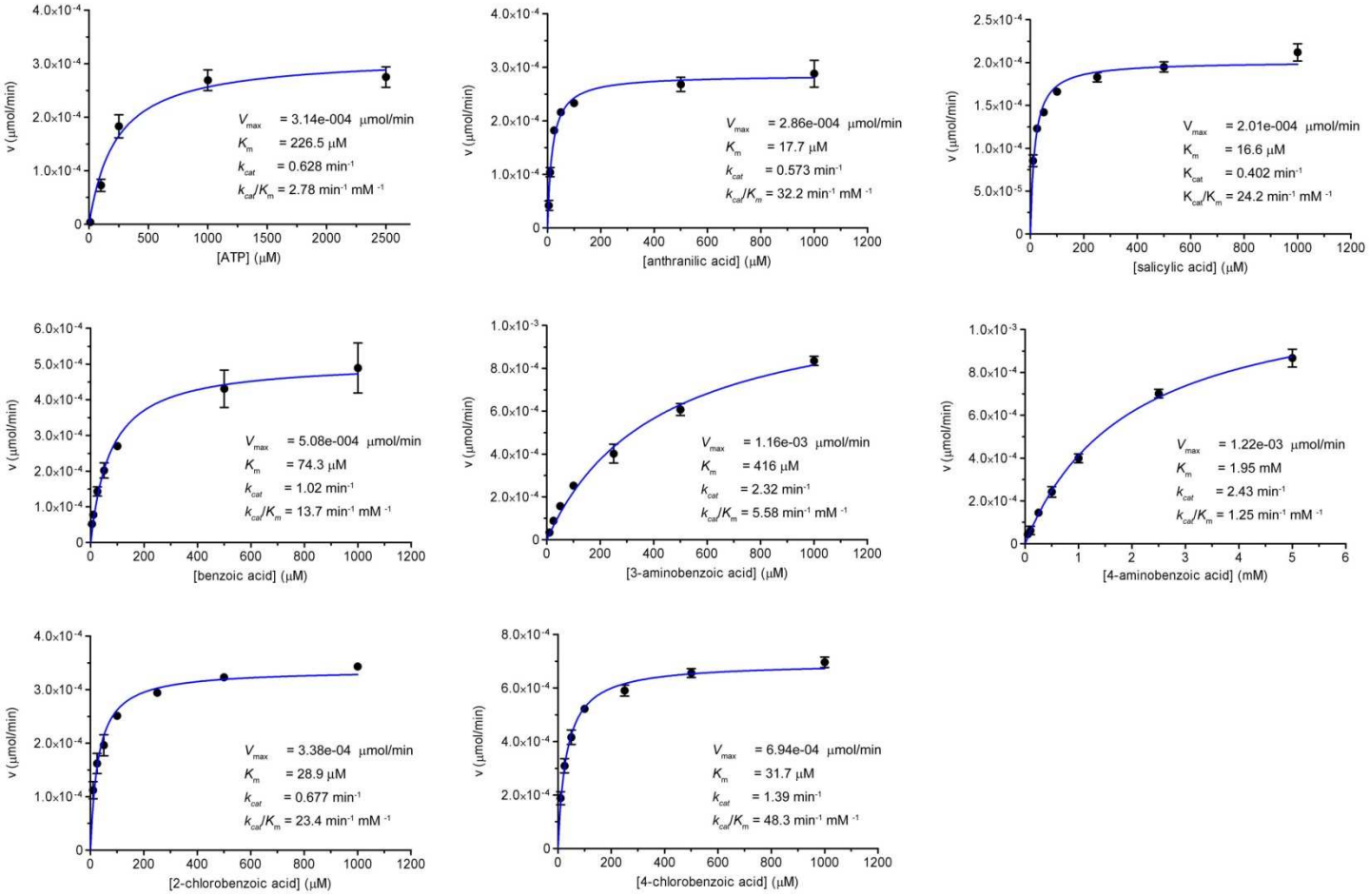


Figure S8. Loading of [^{14}C]anthranilate onto the holo-T-domain of AFUA_6g12080 module 1 C*AT. Left image, coomassie stained SDS-PAGE gel. Right image, autoradiograph of this gel. Lanes 1 and 2 differ in their treatment of the C*AT protein with Sfp and CoA (as indicated) for *in vitro* installment of the phosphopantetheine (PPT) prosthetic group (required for acyl-group transfer/thiolation activity (*12*)). Lane 1 lacks Sfp and CoA, while lane 2 includes Sfp and CoA in the reaction mixture. The autoradiograph suggests that some enzyme is phosphopantetheinylated during expression in *E. coli* (lane 1), but that *in vitro* phosphopantetheinylation by Sfp and CoA provides for greater loading of [^{14}C]Ant (lane 2).

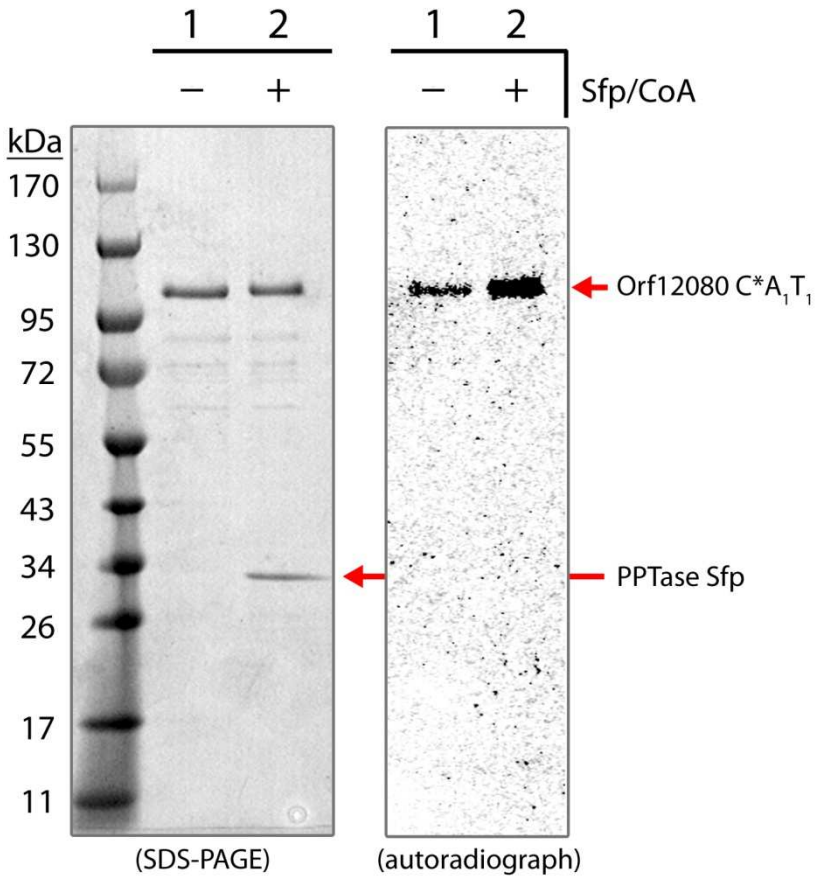
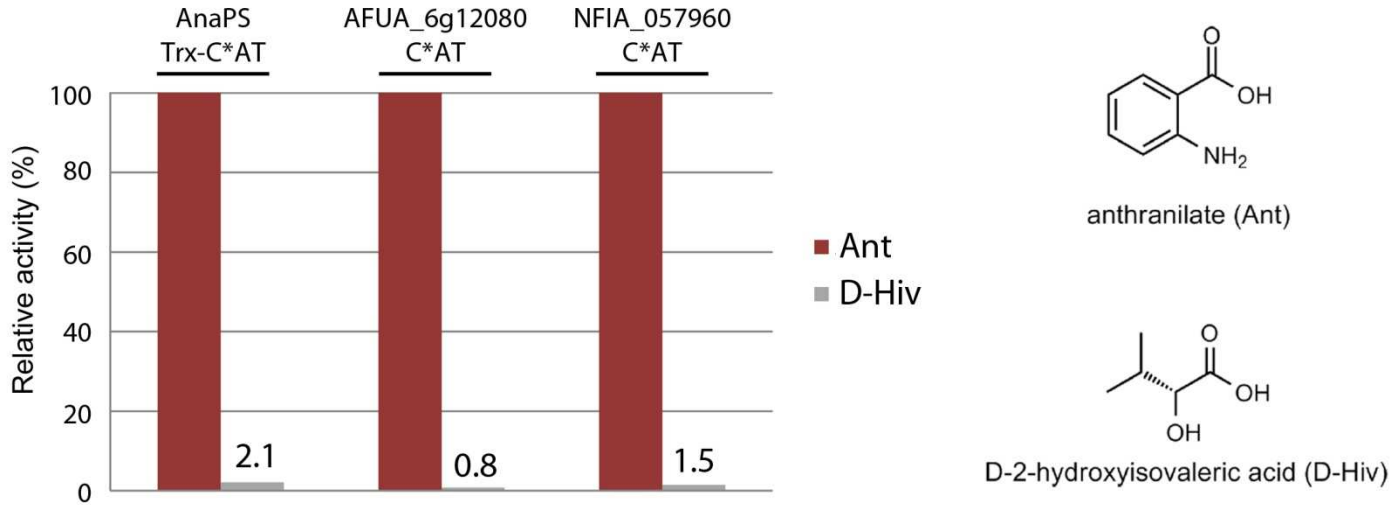


Figure S9. ATP-PP_i exchange assay of the three module 1 C*AT enzymes characterized in this work demonstrates anthranilate selectivity over D-2-hydroxyisovaleric acid. In terms of CPM's, the 100% relative activity for anthranilate corresponds to: 6200 CPM (AnaPS Trx-C*AT), 214400 CPM (AFUA_6g12080 C*AT), and 16200 CPM (NFIA_057960 C*AT).



1. Frisvad, J. C., Rank, C., Nielsen, K. F., and Larsen, T. O. (2009) Metabolomics of *Aspergillus fumigatus*, *Med. Mycol. 47*, 53-71.
2. Takahashi, C., et al. (1995) Fumiquinazolines A–G, novel metabolites of a fungus separated from a *Pseudolabrus* marine fish, *J. Chem. Soc., Perkin Trans. 1*, 2345 - 2353.
3. Nierman, W. C., et al. (2005) Genomic sequence of the pathogenic and allergenic filamentous fungus *Aspergillus fumigatus*, *Nature 438*, 1151-1156.
4. Stack, D., Neville, C., and Doyle, S. (2007) Nonribosomal peptide synthesis in *Aspergillus fumigatus* and other fungi, *Microbiology 153*, 1297-1306.
5. Loria, R., et al. (2008) Thaxtomin biosynthesis: the path to plant pathogenicity in the genus *Streptomyces*, *Antonie van Leeuwenhoek 94*, 3-10.
6. Wiest, A., et al. (2002) Identification of peptaibols from *Trichoderma virens* and cloning of a peptaibol synthetase, *J. Biol. Chem. 277*, 20862-20868.
7. Keating, T. A., et al. (2001) Chain termination steps in nonribosomal peptide synthetase assembly lines: directed acyl-enzyme breakdown in antibiotic and siderophore biosynthesis, *Chembiochem 2*, 99-107.
8. Lawen, A., and Traber, R. (1993) Substrate specificities of cyclosporin synthetase and peptolide SDZ 214-103 synthetase. Comparison of the substrate specificities of the related multifunctional polypeptides, *J. Biol. Chem. 268*, 20452-20465.
9. Larkin, M. A., et al. (2007) Clustal W and Clustal X version 2.0, *Bioinformatics 23*, 2947-2948.
10. Nicholas, K. B., Nicholas, H. B., and Deerfield, D. W. (1997) GeneDoc: analysis and visualization of genetic variation, *EMBNEW. NEWS 4*, 14.
11. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S, *EMBO J. 16*, 4174-4183.
12. Lambalot, R. H., et al. (1996) A new enzyme superfamily -- the phosphopantetheinyl transferases, *Chem. Biol. 3*, 923-936.