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  $C_a$  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<sub>1</sub>-Trp<sub>2</sub>-Ala<sub>3</sub>-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.



**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<sub>t</sub> 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 <sup>15</sup>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<sub>2</sub>, 1 mM DTT, 50 mM Hepes (pH 7.4), and Ant + ATP (Rxn1), or <sup>15</sup>N-Ant + ATP (Rxn2), or Ant + <sup>15</sup>N<sub>5</sub>-ATP (Rxn3), or <sup>15</sup>N-Ant + <sup>15</sup>N<sub>5</sub>-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<sup>2+</sup> ion (11).

			Pos1	2 3		4	
anaPS Al-Ant	:	FASIATHARSLCLSOOSRSLO	TASFCEG	ASLLEIWCTL	IVGGTLCIPSDHDRLNS	LGEFMAKMRINWAFI	TPTVLASI
grsA Al-Phe	:	ISNLKVFFENSLNVTEKDRIGQ	FASISED	ASVWEMEMAL	LT <mark>GASLYII</mark> LKDT <mark>IND</mark> FVK	FEQ <mark>YI</mark> NQKE <mark>ITVI</mark> TLE	PTYVV-HLD
tyrocidine1 A1-Phe	:	IANLQSFFQNSFGVTEQDRIGLE	FASMSFD	AS <mark>VWEMF</mark> MAL	LS <mark>GASLYIL</mark> SKQT <mark>I</mark> HDFAA	FEHYLSENELTIITL	PTYLT-HLT
tyrocidine2 A2-Phe	:	IVNSVTWNRDEFALSVRDSGTLS	SL <mark>SF</mark> A <mark>FD</mark>	AFALTEETL	VS <mark>GSTV</mark> VLMPDHEAK <mark>D</mark> PIA	LRNLIAAWECSYVVFV	/ <mark>PSM</mark> FQ-A <mark>I</mark> LECST
grsB A2-Val	:	VIRLVKNTNYVQVREDDRIIQ	rgaig <mark>fd</mark>	ALTFEVEGSL	LH <mark>GAEL</mark> YPVTKDVLL <mark>D</mark> AEK	LHKFLQANQ <mark>ITIM</mark> WLI	SPLFN-QLSQGT
srfA2 Al-Val	:	ILRLVKNAGYVPVTEEDAMAQ	rgavs <mark>fd</mark>	AGT <mark>FEVF</mark> GAL	L <mark>NGA</mark> ALYPVKKRH <mark>V</mark> LDAKQ	FAA <mark>FL</mark> REQS <mark>ITTMWLT</mark>	SPLFN-QLAAKD
tyrocidine3_A4-Val	:	VTRLVMHTNYVQVRESDRMIQ	[GAIG <mark>FD</mark>	AMTFEIFGAL	<mark>l</mark> h <mark>gas</mark> ly <mark>lv</mark> skdvll <mark>d</mark> aek	LGD <mark>FL</mark> RTNQ <mark>IT</mark> TMWLT	SPLFN-QLSQDN
srfA1_A1-Leu	:	ILRTVKETNYLSITEQDTILGI	LSNYVFD	AF <mark>MF</mark> DMFGSL	L <mark>NGAKLVLI</mark> PKET <mark>V</mark> L <mark>D</mark> MAR	LSRV <mark>I</mark> EREN <mark>ISIL</mark> MII	TALFH-LLVDLN
grsB_A4-Leu	:	IINCLQWRKEEYEFGPGDTALQ	/F <mark>SF</mark> A <mark>FD</mark>	GF <mark>V</mark> AS <mark>LF</mark> AP <mark>I</mark>	L <mark>agat</mark> sv <mark>l</mark> pkeeeak <mark>d</mark> pva	LKKLIASEE <mark>IT</mark> HYYGV	/ <mark>PSL</mark> FS-A <mark>I</mark> LDVSS
srfA3_A1-Leu	:	IQGLVKHVDYMAFSDQDTFLSV	/SNYAFD	AFTFDFYASM	LNAARLI IADEHTLLDTER	LTDL <mark>I</mark> LQENVNVMFAI	TALFN-LLTDAG
tyrocidine2_A1-Pro	:	MANLMHFTFDQTNIAFHEKVLQ	TTCSFD	VCYQ <mark>EIF</mark> ST <mark>L</mark>	L <mark>SG</mark> GQ <mark>L</mark> Y <b>LI</b> TNELRRHVEK	LFA <mark>FI</mark> QEKQ <mark>ISIL</mark> SLF	PV <mark>S</mark> F <b>L</b> K-F <mark>I</mark> FNEQDYA
pristinamycin3_A1-Pro	:	LVNLLAWHRREIPGEAGAPVAQ	FT <mark>T</mark> IG <mark>FD</mark>	VAAQ <mark>EI</mark> LATW	LHGKTLAVPSQEVRRSAEQ	LAAWLDEQH <mark>VS</mark> ELYAE	PNL <mark>VI</mark> EALAEAA-AEA
pristinamycin2_A1-Thr	:	VVRLFTSTDHWFGFGPDDVWTLH	TH <mark>SY</mark> AFD	FS <mark>VWE</mark> IWGAL	LHGGRLVVVPYHVSRSPGD	FLDLLAREK <mark>VTVL</mark> NQ <mark>I</mark>	PTAFHQLDAADRART
syringomycin_A1-Ser	:	VVNRLLWARDAYQVNSQDRVLQ	KTPCGED	LS <mark>VWE</mark> FFLPL	LTGAELVMAPPGGHQDPDY	LAQVMSDAGITLLHFV	PSMLD-VFLEHRS
srfA1_A1-Glu	:	VHHLVESLQQTIYQSP <b>TQ</b> TLPMAFI	_PPFHFD	ASVKQIFASL	L <mark>GQTLYIV</mark> PKKT <mark>V</mark> TNGAA	LTAYYRKNSIEATDG	PAHLQ-MLAAAG
tyrocidine3_A1-Asn	:	LVNYIWWANKVYVQGEAVDFPL	(S <mark>S</mark> ISFD	LTVTSIFTPL	LSGNTIHVYRGADKVQV	ILDIIKDNKVGIIKLI	PTHLK-LIEHIDG
tyrocidine3_A2-Gln	:	ICNHMLWMRETFPLTTEDAMLQ	KTPFSFD	ASVWEFYLPL	ITGGQLVLAKPDGHRDIAY	MTRLIRDEKITTIQMV	/ <mark>PSLL</mark> D-L <mark>V</mark> MTDPG
syringomycinA5-Arg	:	LFAVSAAWEQLYALHAPLNHLQ	1AGE-ED	VFSADLIRSL.	AF <mark>G</mark> GTLVLCPRETLMDPPA	LYRLLSEESIGFADEV	/LA <mark>VI</mark> N-A <mark>L</mark> LGWVEET
		B	lock 1	Block 2		Block 3 B	llock 4
		5 6		7 89			10
				1 11			
DC 31 3-+					DAMODOTDE		
anaPS_A1-Ant	:	↓ ↓ LFIAGEPIGERDIRTWAPR		↓ A <mark>YGLTE</mark> WAGV	FAVSRQIRT	FLPLSQIP	TTITCKADRRSLCRD
anaPS_A1-Ant grsA_A1-Phe	:	LFIAGEPIGERDIRTWAPR	ARLFQ VTYIN	↓ ↓↓ AYGLTEWAGV AYGPTETTIC.	FAVSRQIRT A <mark>T</mark> TWVA-TK	FLPLSQIP FIQLDKMP	TTITCKADRRSLCRD
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe		LFIASEPIGERDIRTWAPR	AR <mark>L</mark> FQ VTYIN LRYIN	AYGLTEWAGV AYGPTETTIC. AYGPTETSIC.	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATTIBETTOP	FLPLSQIP FIQUDKMP FVKLDKMP LLCDDAEB	TTITCKADRRSLCRD LTSNGKIDRKQLBEP LTPNDKIDRKALPEP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe greB_A2-Va1		LFIAGEPIGERDIRTWAPR LTASSATSPSLVNKWKEK MITASSASSPLVNKWKDK VMLGGEKLSPKLVQLCKAM-HPQ	ARLFQ VTYIN LRYIN MSVMN	↓ AYGLTEWAGV AYGPTETTIC. AYGPTETSIC. AYGPTESSVM CYCPTENTTE	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP	FLPISQIP FIQIDKMP FVKLDKMP LIQIDAFP FVOIDAFP	TTITCKADRESLCRD ITSNCKIDRKOLPEP ITPNDKIDRKALPEP ITPNCKVDRKALPEP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Va1 srfA2_A1-Va1		LFIAGEPIGERDIRTWAPR LTASSATSPSLVNKWKEK MITAGSASSAPLVNKWKEK VMLGEKLSPKLVQLKAM-HPQ LIVGGDALSPKHINNVKRKCPNL LTCGDALVPHTVSKVKASPSL	ARLFQ VTYIN LRYIN MSVMN TMWN	AYGLTEWAGV AYGPTETTIC. AYGPTETSIC. AYGPTESSVM GYGPTENTTF GYGPTENTTF	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDR	FLPISQIP FIQIDKMP FVKIDKMP LIQIDAFP FVQIEQMP LIQMDSIP	TTITCKADRRSLCRD LTSNGKIDRKOLPEP LTPNDKIDRKALPEP LTPNGKVDRKALPEP LTONGKVNRSALPEP LTONGKINKRELPAD
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Val srfA2_A1-Val tyrocidine3_A4-Val		LFIASEPIGERDIRTWAPR LTASSATSPSLVNKWKEK MITASSASSAPLVNKWKDK VMLGEKLSPKLVQLKAM-HPQ LIVGGDALSPKHINNVKKKCPNL LIJGDALVPHIVSKVKQASPSL LTVCGEALSPKHINRVKSALPDL	ARLFQ VTYIN LRYIN MSVMN TMWN SLWN	AYGLTEWAGV AYGPTETTIC. AYGPTETSIC. AYGPTESSWM GYGPTENTTF GYGPTENTTF GYGPTENTTF	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ	FLPISQIF FIQIDKMP FVKIDKMP LIQIDAFP FVQIEQMP LIQMDSLP YVOMEKLP	TTITCKADRRSLCRD LTSNGKIDRKQLPEP LTPNDKIDRKALPEP LTPNGKVDRKALPEP LTQNGKVNRSALPEP LTPNGKUDRALPOP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Va1 srfA2_A1-Va1 tyrocidine3_A4-Va1 srfA1_A1-Leū		LFIAGE PIGERDIRTWAPR LTAGSATSPSLVNKWKEK WIGGEKLSPKLVQLCKAM-HPQ LIVGSDALSPKHINNVKRKCPNL LIVGSDALSPKHINRVKSALPDL LIVGSEALSPKHINRVKSALPDL LIVGSEASVEHVRKALOTVGKG	AR <mark>L</mark> FQ VTYIN LRYIN MS <mark>V</mark> MN TMWN SLWN EIWN EIWN	AYGLTEWAGV AYGPTETTIC. AYGPTETSIC. AYGPTESSUM GYGPTENTTF GYGPTENTTF GYGPTENTTF MYGPSESTVF	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE	FLPISQIF FIQIDKMP FVKIDKMP LIQIDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIOMDELP	TTITGKADRSLCRD LTSNCKDRKOLPEP LTPNDKLDRKALPEP LTPNGKVDRKALPEP LTONCKVNRSALPEP LTANCKVDRKALPAP LTGNCKLDRALPIP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Va1 srfA2_A1-Va1 tyrocidine3_A4-Va1 srfA1_A1-Leu grsB_A4-Leu		LFIAGE PIGERDIRTWAPR LITAGSATS PSLVNKWKEK MIGGEKIS PKLVQLCKAM-HPQ LIVGSDALS PKHINNVKRKCPNL LIIGSDALVPHIVSKVKQASPSL LIVGGEALSPHINRVKSALPDL IMFGGERASVEHVRKALQTVGKG VTLGEEKIPAOIVKKIKEK-NKE	AR <mark>L</mark> FQ VTYIN LRYIN MSVMM TMWN SLWN EIWN KLLH IEVNN	AYGPTETTIC AYGPTETTIC AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF MYGPSESTVF EYGPTENSVV	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIOV	FLPLSQIP FIQLDKMP EVKLDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP	TTITKADRESLCRD LTSNGKIDRKQLPEP LTPNGKIDRKALPEP LTPNGKVDRKALPEP LTONGKVNRSALPEP LTANGKVDRKALPAP LTGNGKIDRALPIP LTGNGKUDRALPEP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Va1 srfA2_A1-Va1 tyrocidine3_A4-Va1 srfA1_A1-Leu grsB_A4-Leu srfA3_A1-Leu		LFIASEPIGERDIRTWAPR LITAGSATSPSLVNKWKEK MITASSASSAPLVNKWKEK VMLGEEKISPKLVQLCKAM-HPQ LIVGGDALSPKHINVVKKCPNL LIVGGDALSPKHINRVKSALPDL LIVGGEALSPKHINRVKSALPDL IMFGGERASVEHVRKALQTVGKG VTLCGEKLPAQIVKKIKEK-NKE IFCGERASVPHVRKALRIMGPG	AR <mark>L</mark> FQ VTYIN LRYIN MSVMM SLWM SLWN EIWN KLLH IEVNN KLIN	AYGLTEWAGV AYGPTETTIC. AYGPTETSIC. AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF EYGPTENTVF EYGPTENTVF	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD	FLPLSQIP FIQLDKMP FVKLDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP FTFLDELP	TTITGKADRESLCRD LTSNGKIDRKQLPEP LTPNGKIDRKALPEP LTPNGKIDRKALPKP LTQNGKVNRSALPKP LTANGKVDRRALPXP LTGNGKIDRRALPIP LTTNGKVNKRLLPKP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Va1 srfA2_A1-Va1 tyrocidine3_A4-Va1 srfA1_A1-Leu grsB_A4-Leu srfA3_A1-Leu tyrocidine2_A1-Pro		LFIAGEPIGERDIRTWAPR ITAGSATSPSLVNKWKEK MITAGSATSPSLVNKWKEK VMLGGEKLSPKLVQLCKAM-HPQ LIVGGDALSPKHINNVKRKCPNL LIGGDALVPHIVSKVKQASPSL LIVGGEALSPKHINRVKSALPDL IMFGGERASVEHVRKALQTVGKG VTLGGEKLPAQIVKKIKEK-NKE IFGGERASVPHVRKALRIMGPG ITAGECLVVTHELOXYLR-QHR	ARUFQ UTYIN LRYIN SUN TMWN SLWN SLWN KLIN KLIN KLIN	AYGLTEWAGV AYGPTETTIC. AYGPTESSC. AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF MYGPSESTVF CYGPTENSVV CYGPTEGTVF.	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG	FLPLSQIP FIQLDKMP FVKLDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP FTFLDELP FVTLERIP	TTITCKADRSLCRD LTSNGKIDRKQLPEP LTPNGKUDRKALPEP LTPNGKUDRKALPEP LTQNGKUDRSALPEP LTQNGKUDRSALPEP LTANGKUDRRALPEP LTPNGKUDRKALPEP LTTNGKUNKRLLPEP WTPNGKTDRRALPEP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Va1 srfA2_A1-Va1 tyrocidine3_A4-Va1 srfA1_A1-Leu grsB_A4-Leu srfA3_A1-Leu tyrocidine2_A1-Pro pristinamycin3_A1-Pro		LFIAGET GERDIRTWAPR MITAGSATSPSLVNKWKEK MITAGSATSPSLVNKWKEK VMLGEKLSPKLVQLCKAM-HPQ LIVGOALSPKHINNVKRKCPNL LIGGALVPHIVSKVKQASPSL LIVGGEALSPKHINRVKSALPDL IMFGGERASVEHVRKALQTVGKG VTLGEKLPAQIVKKIKEK-NKE ILFGERASVEHVRKALRIMGPG IITAGEQIVVTHELQKYLR-QHR IAQAGEALTLTRTVREFAAAVPG	ARUFQ UTYIN LRYIN TSVMN TMWN SLWN EIWN KLIN KLIN KLIN KLIN	AYGLTEWAGV AYGPTETTIC. AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF MYGPSESTVF CYGPTESSVV CYGPTEGTVF. HYGPSETHVV HYGPAETHV	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG TGTA-LPED	FLPLSQIP FIQLDKMP EVKLDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP FTFLDELP FVTLERIP IVALDALP	TTITCKADRRSLCRD LTSNGKIDRKQLPEP LTPNGKUDRKALPEP LTPNGKUDRKALPEP LTONGKUNRSALPEP LTONGKUNRSALPEP LTANGKUDRRALPEP LTGNGKIDRRALPEP LTTNGKUDRRALPEP LTTNGKIDRRALPEP LTPNGKIDRRALPEP
anaPS_Al-Ant grsA_Al-Phe tyrocidine1_Al-Phe tyrocidine2_A2-Phe grsB_A2-Val srfA2_Al-Val tyrocidine3_A4-Val srfA1_Al-Leu grsB_A4-Leu srfA3_A1-Leu tyrocidine2_Al-Pro pristinamycin3_A1-Pro pristinamycin2_Al-Thr		LFIAGET GERDIRTWAPR ITAGSATSPSLVNKWKEK MITAGSASSAPLVNKWKDK VMLGGEKLSPKLVQLCKAM-HPQ LIVGSPALSPKHINNVKRKCPNL LIGGDALVPHIVSKVKQASPSL LIVGGEALSPKHINRVKSALPDL IMFGGERASVEHVRKALQTVGKG VTLGGEKLPAQIVKKIKEK-NKE ILFGGERASVPHVRKALQTVGKG ITAGEQLVVTHELQKYLR-QHR ITAGEALTLTRTVREFAAAVPG VFGGEALDVARLADWYARRGTA	ARUFQ URYIN MSVMM SLWM SLWM SLWM KLLH KLLH KLIN KLIN VFLHN RQLHN ARUVN	AYGLTEWAGV AYGPTETTIC. AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF MYGPSESTVF. EYGPTENSVV CYGPTEGTVF. HYGPSETHVV MYGITETTV	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG TGTA-LPED VTHAPLGPG	FLPLSQIP FVKLDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP FTFLDELP FVTLERIP IVALDALP FTLLAALP	TTITCKADRRSLCRD LTSNGKIDRKQLPEP LTPNGKVDRKALPEP LTPNGKVDRKALPEP LTONGKVNRSALPEP LTONGKVNRSALPEP LTPNGKVDRKALPEP LTGNGKIDRRALPIP LTTNGKVDRKALPEP LTTNGKTDRRALPEP LTPNGKDRAALPAP LTANGKLDRAALPAP
anaPS_A1-Ant grsA_A1-Phe tyrocidine1_A1-Phe tyrocidine2_A2-Phe grsB_A2-Va1 srfA2_A1-Va1 tyrocidine3_A4-Va1 srfA1_A1-Leu grsB_A4-Leu srfA3_A1-Leu tyrocidine2_A1-Pro pristinamycin3_A1-Pro pristinamycin2_A1-Thr syringomycin_A1-Ser		↓	ARUFQ VTYIN LRYIN MSVMN TMWN SUWN EIWN KILH IEVNN KIIN VFLHN RQHN RQUN RQUN RQUN	AYGLTEWAGV AYGPTETTIC. AYGPTETSIC. AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF MYGPSESTVF EYGPTEGTVF. HYGPSETHVV MYGITETVH LYGPTEAALD	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG TGTA-LPED VTHAPLGPG VTAWECRPT	FLPLSQIP FIQLDKMP FVKLDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP FTFLDELP FVTLERIP IVALDALP FTTLAALP YVLLEAMP	TTITGKADRSLCRD LTSNGKLDRKQLPEP LTPNGKUDRKALPKP LTPNGKUNRSALPKP LTONGKUNRSALPKP LTANGKUDRKALPAP LTGNGKLDRALPIP LTGNGKUDRKALPEP LTTNGKUNKRLLPKP VTPNGKDRAALPAP LTNGKLDRAALPAP LTSNGKLDRAALPAP
anaPS_Al-Ant grsA_Al-Phe tyrocidine1_Al-Phe tyrocidine2_A2-Phe grsB_A2-Val srfA2_Al-Val tyrocidine3_A4-Val srfA1_Al-Leu grsB_A4-Leu srfA3_Al-Leu tyrocidine2_Al-Pro pristinamycin3_Al-Pro pristinamycin3_Al-Pro syringomycin_Al-Ser srfA1_Al-Glu		LFIAGE PIGERDIRTWAPR LITAGSATS PSLVNKWKEK MIGGEKIS PKLVQLCKAM-HPQ LIVGSDALS PKHINNVKRKCPNL LIVGSDALS PKHINRVKSALPDL INFGGERASVEHVRKALQTVGKG VILGGEKIPAQIVKKIKEK-NKE ILFGSERASVPHVRKALRIMGPG ITASEQIVVHELQKXLR-QHR IAQGEALTLTRTVREFAAAVPG VFGSEALDVARLADWYARRGTA VLCSEALPRALQRFEQQLKG MLIGSEGLSSVVADKLLKLFKEAG		AYGLTEWAGV AYGPTETTSIC. AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF GYGPTENTVF CYGPTESTVF. CYGPTEGTVF. HYGPSETHVV HYGPAETHVM MYGITETTVH LYGPTEAAID	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG TGTA-LPED VTHAPLGPG VTAWECRPT ASVHPVIPE	FLPLSQIP FIQLDKMP FVKLDKMP LIQLDAFP YVQTEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP FTFLDELP FVTLERIP IVALDALP FTLLAALP YVLLEAMP FTEVTEIP	TTITKADRSLCRD LTSNGKLDRKQLPEP LTPNDKLDRKALPEP LTPNGKVDRKALPEP LTPNGKVDRKALPEP LTANGKVDRKALPEP LTANGKVDRKALPEP LTGNGKLDRALPEP LTTNGKVDRKALPEP LTTNGKUDRALPEP LTTNGKDRRALPAP LTSNGKLDRALPAP LTSNGKDRKALPAP
anaPS_Al-Ant grsA_Al-Phe tyrocidine1_Al-Phe tyrocidine2_A2-Phe grsB_A2-Val srfA2_Al-Val tyrocidine3_A4-Val srfA1_Al-Leu grsB_A4-Leu srfA3_Al-Leu tyrocidine2_Al-Pro pristinamycin3_Al-Pro pristinamycin3_Al-Pro pristinamycin_Al-Ser srfA1_Al-Glu tyrocidine3_Al-Asn		LFIASE PIGERDIRTWAPR LITAGSATS PSLVNKWKEK MIGGEKIS PKLVQLCKAM-HPQ LIVGGDALS PKHINNVKRKCPNL LIVGGDALS PKHINRVKSALPDL INFGGERAS VEHVRKALQTVGKG VTLGSEKIPAQIVKKIKEK-NKE ILFGGERAS VPHVRKALRIMGPG ILFGGERAS VPHVRKALRIMGPG		AYGLTEWAGV AYGPTETTIC. AYGPTESSVM GYGPTENTTF GYGPTENTTF GYGPTENTTF GYGPTENTVF CYGPTEGTVF HYGPSETVV HYGPAETHVM HYGPTETTVH LYGPTETCVD EYGPTETCVD	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG TGTA-LPED VTHAPLGPG VTAWECRPT ASVHPVIPE CMIYLYDPQ	FLPLSQIP FIQLDKMP FVKLDKMP LIQLDAFP EVQIEQMP LIQMDSLP YVQMEKLP FIQDDSIP FIQLDSIP FTFLDELP FVTLERIP IVALDALP YVLLEAMP FTLLAALP FTLLAALP	TTITGKADRESLCRD LTSNGKIDRKOLPEP LTPNGKIDRKALPEP LTPNGKVDRKALPEP LTPNGKVDRKALPEP LTANGKVDRKALPEP LTGNGKIDRALPIP LTGNGKUDRKALPEP LTTNGKVDRKALPEP LTTNGKKDRAALPAP LTPNGKDRAALPAP LTSNGKLDRAALPAP LTSNGKVDRKKLPAP LTSNGKVDRKKLPAP
anaPS_Al-Ant grsA_Al-Phe tyrocidine1_Al-Phe tyrocidine2_A2-Phe grsB_A2-Val srfA2_Al-Val tyrocidine3_A4-Val srfA1_Al-Leu grsB_A4-Leu srfA3_Al-Leu tyrocidine2_Al-Pro pristinamycin3_Al-Pro pristinamycin2_Al-Thr syringomycin_Al-Ser srfA1_Al-Glu tyrocidine3_Al-Asn tyrocidine3_A2-Gln		LFIAGE PIGERDIRTWAPR LITAGSATS PSLVNKWKEK MITAGSATS PSLVNKWKEK VMLGEEKIS PKLVQLCKAM-HPQ LIVGGDALS PKHINNVKRKCPNL LIVGGDALS PKHINRVKSALPDL INFGGERAS VEHVRKALQTVGKG TLGGEKISVHVRKALQTVGKG IIFGGERASVPHVRKALRIMGPG IAGSEALTLTRTVREFAAAVPG VVFGGEALTLTRTVREFAAAVPG VCSGEALPVARLADWYARRGTA VLCSGEALPAAQRFEQQLKG MIIGGEGISVVADKLLKLFKEAG FIVGGEALPTKLAKQIYDHFGEN VFGGEALTPALVSRFYETQQ-		AYGLTEWAGV AYGPTETTIC. AYGPTESSV GYGPTENTTF GYGPTENTTF GYGPTENTTF EYGPTENTVF HYGPSESTVF. HYGPSETHVV HYGPSETHVV HYGPTEATIV YGPTETCVD GYGPTETCVD GYGPTETVVG LYGPTETTI	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG TGTA-LPED VTHAPLGPG VTAWECRPT ASVHPVIPE CMIYLYDPQ -ATYWPCPR	FLPLSQIP FIQLDKMP FVKIDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FIQLDSIP FTFLDELP FVTLERIP IVALDALP FTLLAALP FTLLAALP FTLLAALP FTLLAALP FTLLAALP	TTITGKADRESLCRD LTSNGKIDRKOLPEP LTPNGKIDRKALPEP LTPNGKIDRKALPEP LTPNGKIDRKALPEP LTANGKVDRKALPEP LTGNGKIDRALPIP LTGNGKIDRALPIP LTTNGKVDRKALPEP LTTNGKVDRKALPEP LTTNGKUDRALPAP LTPNGKLDRAALPAP LTSNGKLDRAALPAP LTSNGKUDRKALPAP LTPSGKVDRKKLPAP LTPSGKVDRKKLPAP
anaPS_Al-Ant grsA_Al-Phe tyrocidine1_Al-Phe tyrocidine2_A2-Phe grsB_A2-Val srfA2_Al-Val tyrocidine3_A4-Val srfA1_Al-Leu grsB_A4-Leu srfA3_Al-Leu tyrocidine2_Al-Pro pristinamycin3_Al-Pro pristinamycin3_Al-Pro syringomycin_Al-Ser srfA1_Al-Glu tyrocidine3_Al-Asn tyrocidine3_A2-Gln syringomycinA5-Arg		LFIAGE PIGERDIRTWAPR LITAGSATSPSLVNKWKEK MITAGSATSPSLVNKWKEK VMLGEKISPKLVQLCKAM-HPQ LIVGGALSPKHINNVKRKCPNL LIVGGALSPKHINRVKSALPDL INFGGERASVEHVRKALQTVGKG VTLGEKIPAQIVKIKEK-NKE ILFGGERASVPHVRKALRIMGPG ILFGGERASVPHVRKALRIMGPG IAQAGEALTLTRTVREFAAAVPG VVFGGEALDVARLADWYARRGTA VLCSGEALPRALQRFEQQLKG MIGGEGLSSVVADKLLKLFKEAG FIVGGEALTPALVSRFYETQQ VCGSDIWTAHSARQLRKLCGDH-		AYGLTEWAGV AYGPTETTIC. AYGPTESSV GYGPTENTTF GYGPTENTTF GYGPTENTTF GYGPTENTTF HYGPSESTVF. HYGPSETHVV HYGPSETHVV HYGPTEATIV YGPTETVVG LYGPTETVVG LYGPTETTID AYGVTEASID	FAVSRQIRT ATTWVA-TK ATIWEAPSN ATYLRDTQP STCF-LIDK STSF-LIDR STCY-LIEQ ATYH-PVDE TTIMRDIQV ATAH-VVHD TTCTMD-PG TGTA-LPED VTHAPLGPG VTAWECRPT ASVHPVIPE CMIYLYDPQ -ATYWPCPR STCFEFEAT	FLPLSQIP FIQLDKMP FVKLDKMP LIQLDAFP FVQIEQMP LIQMDSLP YVQMEKLP FIQMDELP FTFLDELP FVTLERIP IVALDALP FTLLAALP FTLLAALP FTLLAALP FTLLAALP FTLRLAEIP FVFLEQLP FVRMAALP	TTITKA ADRESLCRD LTSNGK LDRKALPEP LTPNGK /DRKALPEP LTPNGK /DRKALPEP LTPNGK /DRKALPEP LTANGK /DRKALPEP LTGNGK LDRKALPEP LTTNGK /DRKALPEP LTTNGK /DRKALPEP LTTNGK /DRKALPEP LTTNGK /DRKALPAP LTSNGK LDRAALPAP LTSNGK LDRAALPAP LTSSK /DRKALPAP LTSSK /DRKKLFAL LTANGK / VRKKLFAP LTSNGK / VRKKLFAP

Figure S5. Sequence alignment comparing the substrate-binding/selectivity-determining region of the fungal Ant-activating Adomains 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 2,3dihydroxybenzoic 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<sub>3</sub>-3-OH-Ant, 4-CH<sub>3</sub>-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 arvl 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 AAactivating enzymes or the Phe-Gly pair of AnaPS A<sub>1</sub>), as well as conserved insertions/deletions around the specificitydetermining 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.

	GrsA/PheA DhbE		β9 β12	2222 2222 α8	80 0000000 000000 80 000000	β10 β13	α9 2020 2	α10 202020200 . 20202020 α10	20 - 222 -	β11 β14	α11 2000 200000 α1	20000000		β	12	α12 000000 0000 α12	ε <u>0000</u> α13	
AA	GrsA/PheA SrfAB_A5	224 1679	DRIGQFASISFD DKTVLLSSYAFD	ASV-W LGY-1	-EMFMALLTO -CMFPVLLGO	GASLYII GGELHIV	LKDTINI QKETYTA	FVKFEQYIN PDEIAHYIK	QKEITV EHGITY	I <mark>T</mark> LP I <mark>K</mark> LT	PTYVVH PSLFHT	LD IVNTASFA	-PERIL FDANFE	SIQTL SLRLI	ITAGS# VL <mark>G</mark> GEH	ATSPSLVNK KIIPTDVIA	WKEK FRKMYG	316 1780
2-OH Aryl	DhbE VibH YbtE MbtA	224 228 211 247	TVYLAALPMAHN TRYLCVLPAAHN SVYLAVLPVAHN DVYLVVLAAGHN	YPLSS FPLSS FPLAC FPLAC	PGVLGVLYAC PGALGVFWAC PGILGTLACC PGLLGAMTVC	GGRVVLS GGCVVLS GGKVVLT GATAVFA	PSPSP QDASP DSASC PDPSP	-DDAFPLIE -QHAFKLIE -DEVMPLIA -EAAFAAIE	REKVTI QHKITV QERVTH RHGVTV	TALV TALV VALV TALV	PPLAMV PPLALL PALAQL PALAKL	WMDAASSR WMDHAEKS WVQAREWE WAQSCEWE	-RDDLS -TYDLS -DSDLS -PVTPK	SLQVL SLHFV SLRVI SLRVI	2VGGAI 2VGGAI 2AGGAI 2VGGSI	KFSAEAARR KFSEAAARR RLDPTLAEQ KLEPEDARR	VKAVFG LPKALG VIATFD VRTALT	323 327 310 346
S. Ant	AnthrOrf21 Sibiromycin_SibE Tomaymycin_TomA	201 189 214	PQDRFLQLAQPSFA PDDCFLQLAPYSFA PPDGLTQSAAFSFA	AST-1 AST-1 AST-1	-DIWTCLLRO -DIWLSLLHO -EIWLAFLHO	GGRLSVA GARVVVL GATLLPM	PQELP-P PSQLP-S PPGLP-S	LGDLARLIV LPKLAHTIK LPVLREAVE	RERTTV EYGVTF ERGATV	'L <mark>N</mark> LP' 'LNLP( 'L <mark>S</mark> LP(	VGLFNL GGLMNL CGLFNA	LVEHHPQT LIDAHPEA LVDQEPEC	LA FA LR	QTRSV KVRTV SVRIVI	IV <mark>S</mark> GDI IVSGDI LLSGDI	FPSAAHLER FPSAPHLAR FPSPDHLRR	ALAVVG VMKAVP ALAHTD	299 288 312
Ant	AnaPS_A1 AFUA_6g12080_A1 NFIA_057960_A1 NFIA_043670_A1 NFIA_043670_A1 ACLA_017890_A1 ACLA_076770_A1 ACLA_095980_A1 ANID_09226.I_A1 ATEG_07358.1_A1	413 449 402 268 283 445 411 426 413 426	QQSRSLQFASFCFG PTSRVLQFASFTFG NRSRVLQFASTFFG SESRVLQFASTSFG PNSRVYQFASHAFG STSRVLQFASFSFG FASRVLQFASTFG FASRVLQFASTFG DSRVLQFASTFGG TKSRVLQFASTFGG	ASL-I VSL-I ISL-I MAL-I VSL-I VSL-I VSL-I VSL-I ISF-I	-EIWCTLIVC -EVWCTLAAC -EIWCTLAAC -DIYCTLAAC -DIYCALTLC -EIWCTLSAC -EIYCTLAAC -EIYCTLAAC -EIYCTLAAC -EIYCTLAAC	GTLCIP GTVCLP GTICMP GATICIP GATICIP GTICMP GTVCVP GTVCVP GTVCIA	SDHDR SDSDR SEDDR SKKDR SEAER SDSDR SADDR SEDDR SEDDR SEHER	LINSLGEFMA VSRLADAIR VNRLAGAIT LINNLSNAIK TGSLAQSMT VSRLGDAVQ INALDRAMN TSRLSGVMN LINALSSILL LINALDAAIR	KMRINW DMQADW KMRITW RMQTTW TMGINW DMQITW IDMQVNW SMQVTW RMEVNW	AFIT ICILT IALMT IALLT IAFLT IAFLT ISIMT IAILT IALIT	PTVLAS PTVLAT PTALQV PSTAIS PSIIQS PTVLAT PSTTTS PTVIDS PTVIDS PTLAQS	ISPDNFNN LEPEAVPN LSPHQVPS LGG-SVEC ICPDDLPH LHPDMVPG IMG-PVET LTPESVPL IADAVVC- LSPEEIPT		-LH-L -LRTI -LRTI -LKTL -LQTL -LQTL -LQTL -LQTL -LQTL -LGTV -LNTL -LRKL	FIAGEI LVAGEI GMAGEI VLAGEI VTAGEI LVAGEI LVAGEI VAGEI FLGGEI	PIGERDIRT PLKKAQFSL PPRKAQIEM PMGIEHIQQ PLPRELMSL ALKKAQISV PMGLNHLHT PLKAPQIDL ALTMDRLHS APNDDLISR	WAPRAR WAERAR WAKAVY LASRLE WAQRVH WASRTS WIDRVE WASNVN LADKTE WQSKAS	506 543 496 361 377 539 504 520 506 520
			Pos1 235	Pos2 236	Pos3 239				Ē	Pos4 278				Pos5 299	Pos 30	<u>66</u> 1		
	GrsA/PheA DhbE		β13 η2	β	اط د	معووه	β15 →						η5 222 222					
AA	GrsA/PheA SrfAB_A5	317 1781	-VTYINAYGPTETT HTEFIN <mark>H</mark> YGPTEAT	ICATI ICATI IGAIA	WVATKEN GRVDLYEPDA	α14 FIGHSVP AFAKRPT	IGAPIQN IGRPIAN	1 354 1 1822		50 19	07 LDK 074 VDA	MPLTSNG <mark>P</mark>	α18 IDRKQ TDRNA	522 1990				
2-OH Aryl	DhbE VibH YbtE MbtA	324 328 311 347	-CTLQQ <mark>V</mark> FGMAEG- -CQLQQVFGMAEG- -CTLQQVFGMAEG- -PGLQQ <mark>V</mark> FGMAEG-	LV-NY LV-NY LL-CF LL-NF	TRLDDPH TRLDDSA TRLDDPH TRIGDPH	EEIIVNT AELIATT HATILHS PEVVEHT	QGKPMSE QGRPISA QGRPLSE QGRPLCE	9 359 3 363 9 346 9 382		50 51 51 53	9 VES 4 IDC 4 IDC 82 MPA	FPQTGVGP LPKTSVGP LPKTSVGP LPKTSVGP	VSKKA IDKNA IDKNA IDKRA	524 529 529 547				
S. Ant	AnthrOrf21 Sibiromycin_SibE Tomaymycin_TomA	300 289 313	-GDLFN <mark>A</mark> FGCTENS -GTVYNAFGCTENS -AVVYN <mark>G</mark> YGCTENS	ALTAV ALTAV SITAI	YHKITPADI YHPMTPEDI .HPLTSPDDVI	LSGTDIP LQLGVVP DGTGVVP	VGRPMPI IGLPLPG IGRPLPI	338 326 353		49 50 47	0 LDS 07 LEK 17 LAE	FPLNANG <mark>P</mark> MPTNVNGP LPVNVNG <mark>P</mark>	(IDRRE (IDRTA (LDRMA	505 523 493				
Ant	AnaPS_A1 AFUA_6g12080_A1 NFIA_057960_A1 NFIA_058030_A1 ACLA_017890_A1 ACLA_076770_A1 ACLA_095980_A1 ANID_09226.I_A1 ATEG_07358.1_A1	507 544 497 362 378 540 505 521 508 521	LFQAYGLTEWA LFQAYGFTEWA LIQAFGFTEWA LUQAFGFTEWA LUQGYGLTEWA LUQGFGFTEWA LUQGFGFTEWA LUQGFGFTEWA VIQAFGLTEWA LFQVFGTTEMA : :* :*	GVFAV GICCV GICAV GICCV GVCCV GVCCV GVCCV GLFSI GICCV GVTMV	/SRQIR7 /SPQIR7 /SQPIK5 /SPPIS1 /SPIG5 /SRRIH5 /SQRIH5 /SSEIT5	FPEDRKS SIGDVGI FPESNSI SEADIRS PADSPGN FIADIGI SERDMRL SRSDLSS SETDLRV STAQRKI	IGSPVNA IGTPANA IGRCAGG IGVSPTA VGRLPTS IGTPANA IGSSPTA IGSPIGG IGRSHTA VGFPANS *	4 540 4 578 5 531 4 375 5 413 4 575 4 539 5 555 4 539 5 555 4 532 5 555		73 76 72 57 59 77 71 73 69 75	30 LSQ   33 LKE   22 TRY   20 LHS   20 LTF   21 IRS   23 VKK   24 LEF   25 LEF   26 LTT	IPTTITG TPVTITG VPVTVSG VPVTITG VPLTPTG MPLTVTG VPLTVSG LPLTISG IPKTISR VPRTVTG *	ADRRS IARQK VDRR VDRRA RSRRD IDRLR TDRRA IARQK VDRKA IDRSG	746 778 737 586 612 787 727 754 712 766				
			Pos7 Pos8 322 330	Pos 331	<u>9</u> I							Pos 51	s10 7					

**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<sub>i</sub>-release assay for determination of kinetic parameters for AFUA\_6g12080 C\*AT with various aryl acid substrates and ATP.



**Figure S8.** Loading of [<sup>14</sup>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 [<sup>14</sup>C]Ant (lane 2).



**Figure S9.** ATP-PP<sub>i</sub> 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.