Biochemistry Supporting Information

Unraveling terminal C-domain-mediated condensation in fungal biosynthesis of imidazoindolone metabolites

Stuart W. Haynes[1§], *Brian D. Ames*[1§], *Xue Gao*[2], *Yi Tang*[2] *and Christopher T. Walsh*[1]*

[1]Department of Biological Chemistry & Molecular Pharmacology, Harvard Medical School, 240 Longwood Avenue, Boston, MA 02115

[2]Department of Chemical and Biomolecular Engineering, Department of Chemistry and Biochemistry, University of California Los Angeles, 420 Westwood Plaza, Los Angeles, CA 90095.

Figure S1. Alignment of TqaH and Af12060 generated using ClustalW (56/76% identity/similarity).

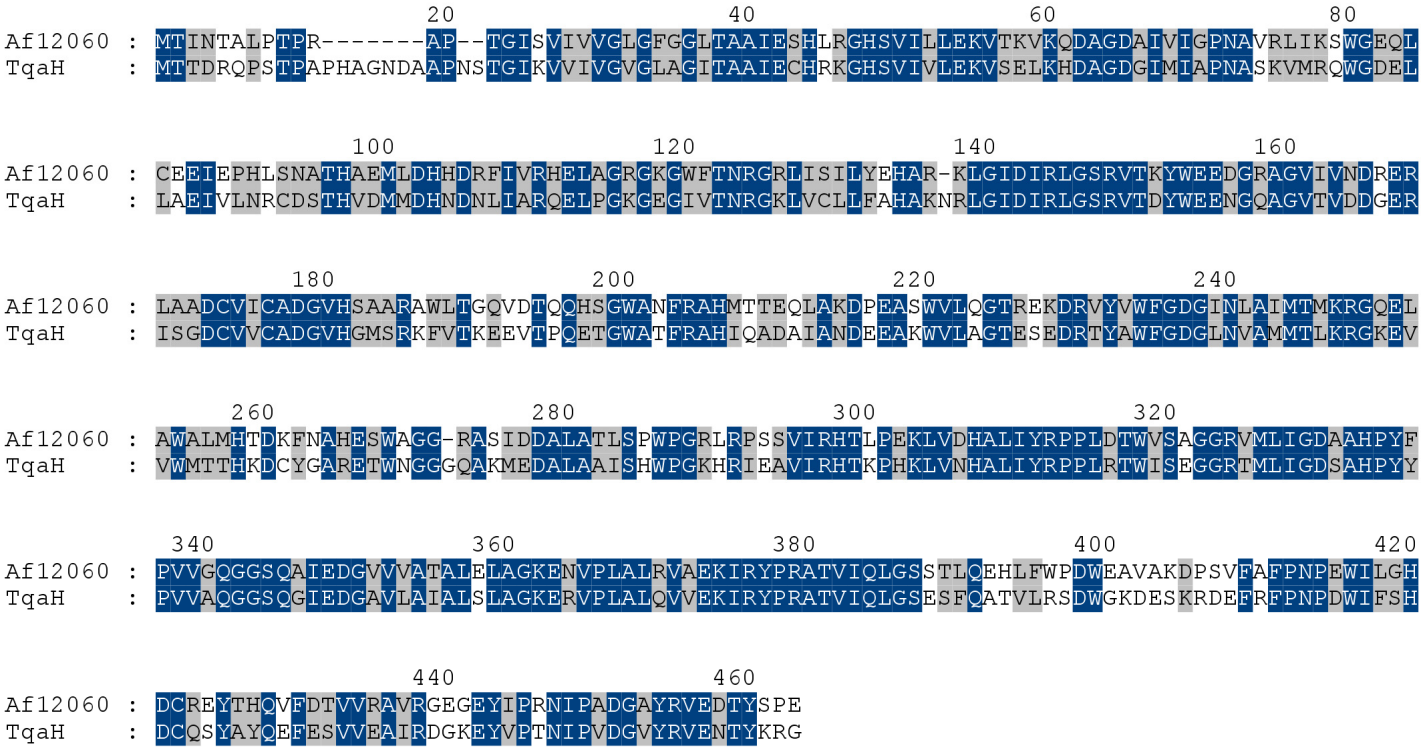


Figure S2. Alignment of TqaB and Af12050 generated using ClustalW (53/75% identity/similarity). The sequence region chosen for cloning and protein expression of the standalone T-domains is indicated by a green bar. Site chosen as the start of the standalone C-domain, and as the site for C-domain swapping to generate the chimeric constructs, is indicated by a red asterisk. The active site motifs for the T- and C-domains are boxed.

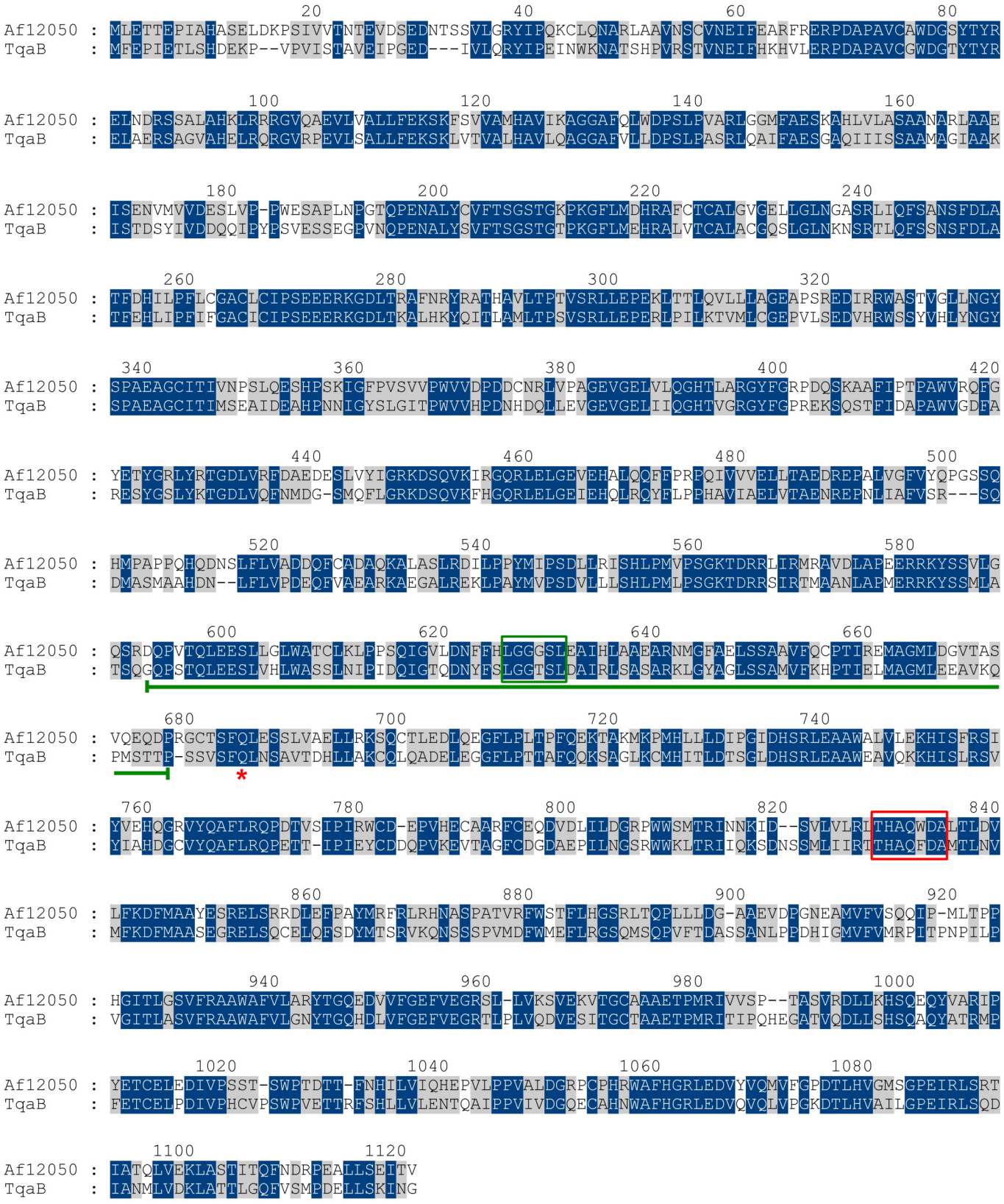


Figure S3. Comparison of the loading of [^{14}C]-acetyl CoA onto the standalone apo-T-domains from Af12050 and TqaB. The percentage of loaded T-domain was calculated as a ratio of the nanomoles of ^{14}C-labelled protein and the total nanomoles of T-domain included in the incubation.

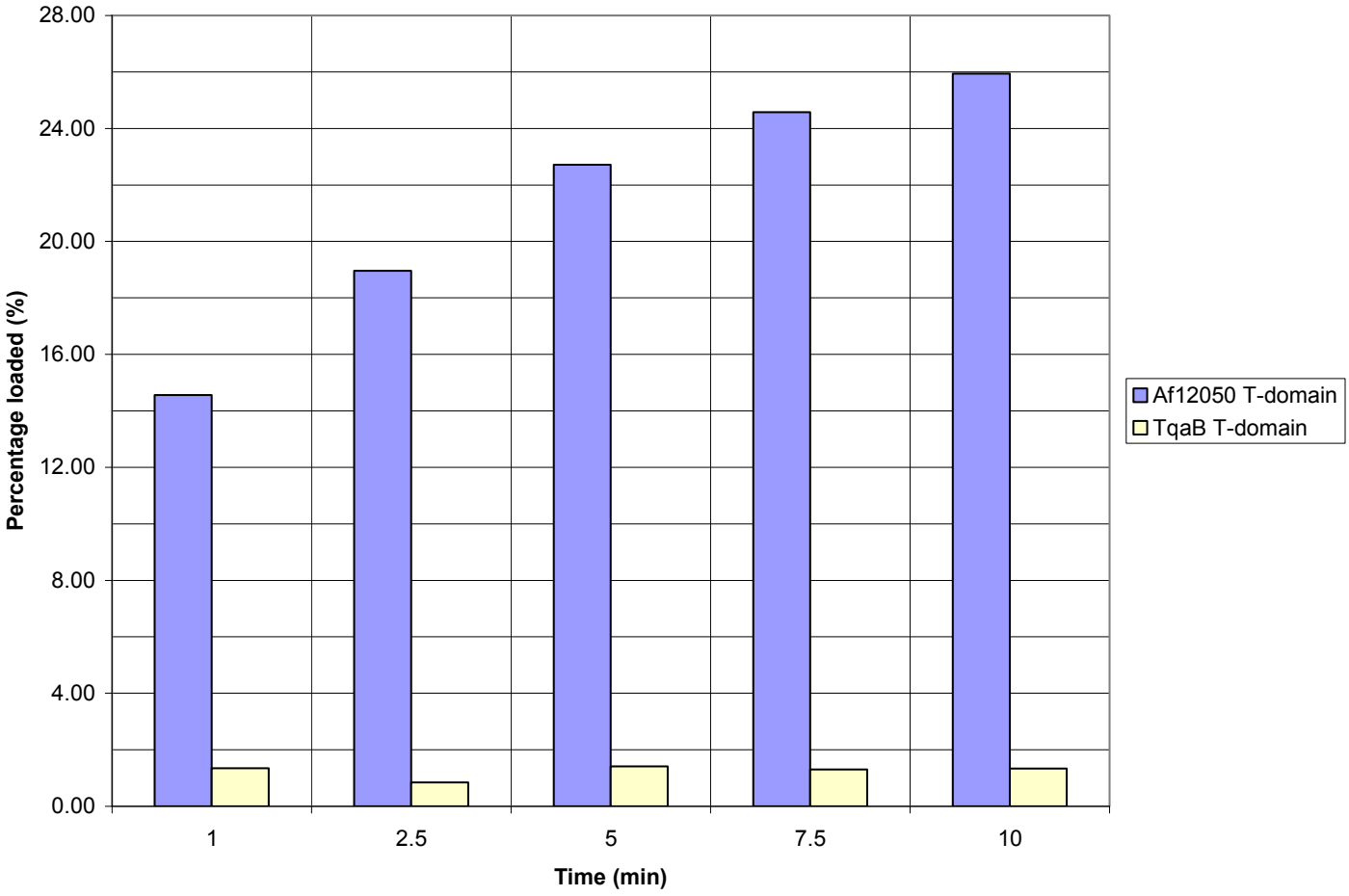
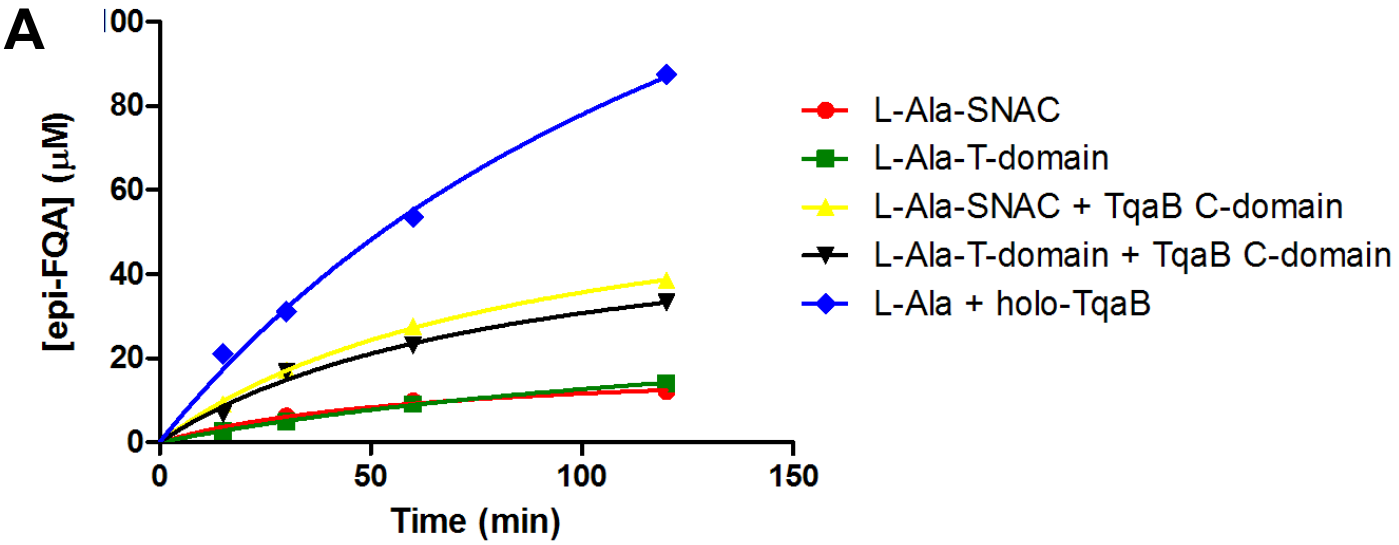


Figure S4. Timecourse HPLC-based assays for 2′-*epi*-FQA formation (A) Observed concentration of 2′-*epi*-FQA over time when using L-Ala SNAC and L-Ala-T-domain plus and minus TqaB C-domain, in comparison incubation of L-Ala with full-length holo-TqaB. (B) Initial rate data of 2′-*epi*-FQA formation, in the presence and absence of TqaB C-domain.



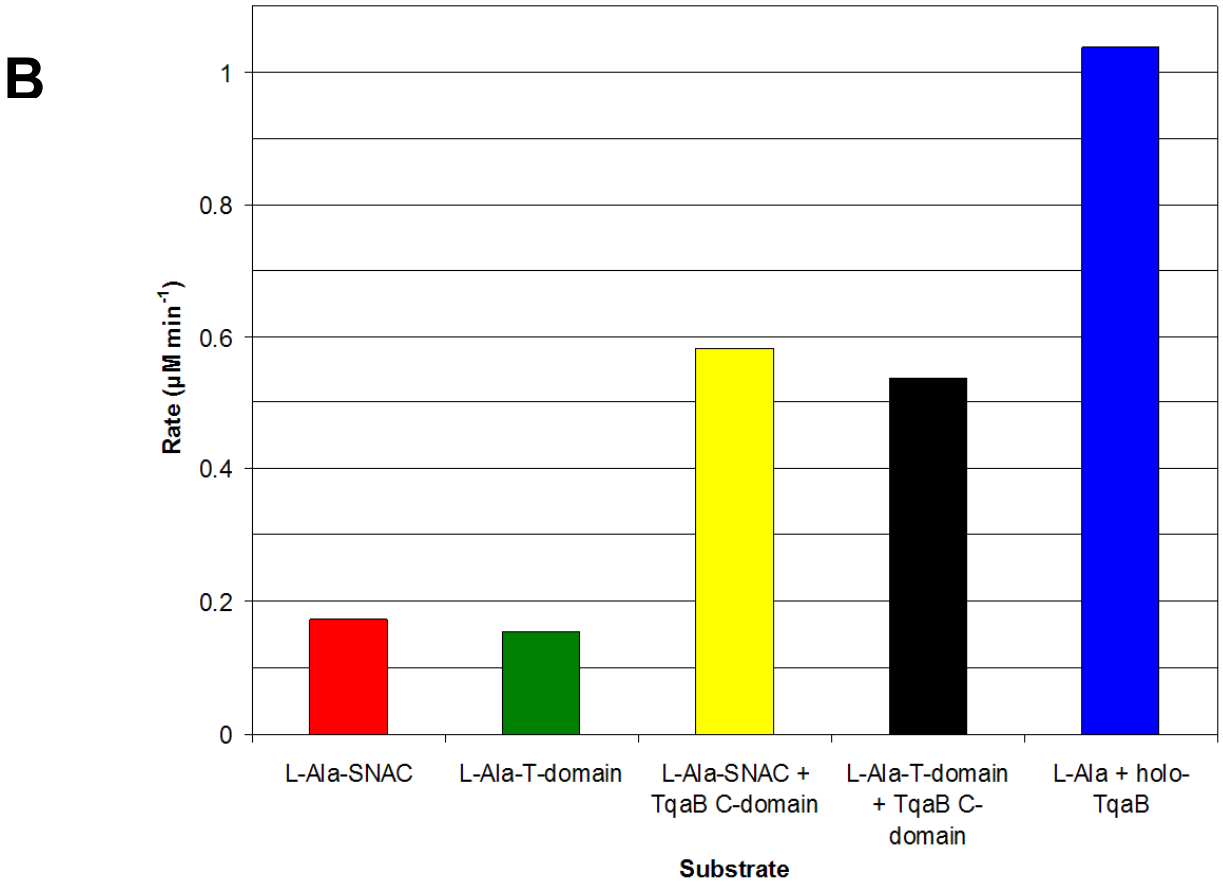
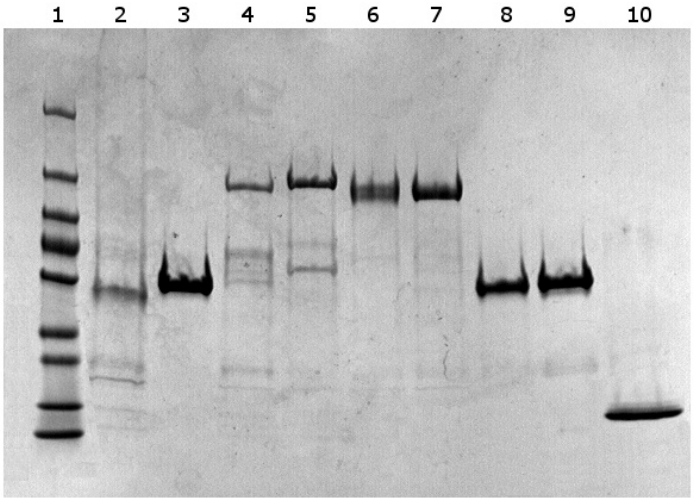


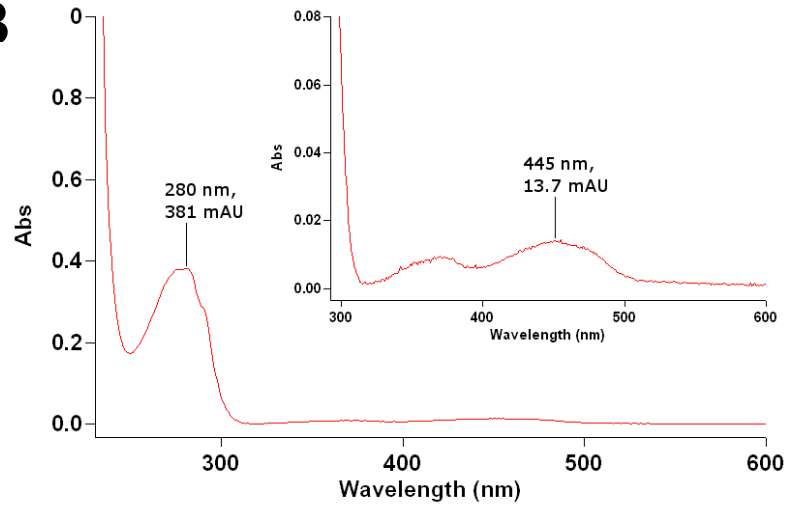
Figure S5. (A) SDS-PAGE analysis of proteins using during these investigations (B) UV spectroscopic traces of FAD-monoxygenases Af12060 and TqaH. Absorbances at 280 nm were used to calculate the enzymes concentration (using the extinction coefficient estimated from protein primary sequence at 280 nm), while absorbances at 445 nm were used to calculate the concentration of FAD (using the known extinction coefficient for FAD at 445 nm of 11, 300 M^{-1} cm^{-1}). Ratios of absorbances for each protein at 280 nm and 445 nm were used to estimate the percentage of protein with FAD bound in each case.

A



Lane	Sample
1	Ladder
2	TqaH
3	Af12060
4	TqaB
5	Af12050
6	Chimera 1
7	Chimera 2
8	TqaB C-domain
9	Af12050 C-domain
10	Af12050 T-domain

B

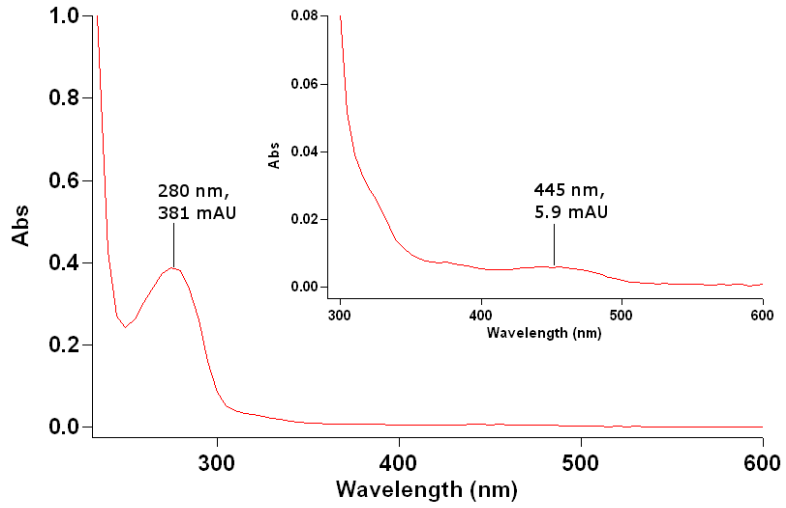


Af12060:

$E_{280\ nm}$ = 91,900 M^{-1} cm^{-1}

Af12060 conc. = 4.15 µM
FAD conc. = 1.21 µM

Protein bound with FAD: 29.2 %



TqaH:

$E_{280\ nm}$ = 78,380 M^{-1} cm^{-1}

TqaH conc. = 4.86 µM
FAD conc. = 0.52 µM

Protein bound with FAD: 10.6 %

Figure S6. UV spectroscopic analysis of compounds key to these investigations recorded via HPLC equipped with a DAD in a mixture of acetonitrile and water plus (0.1% trifluoroacetic acid).

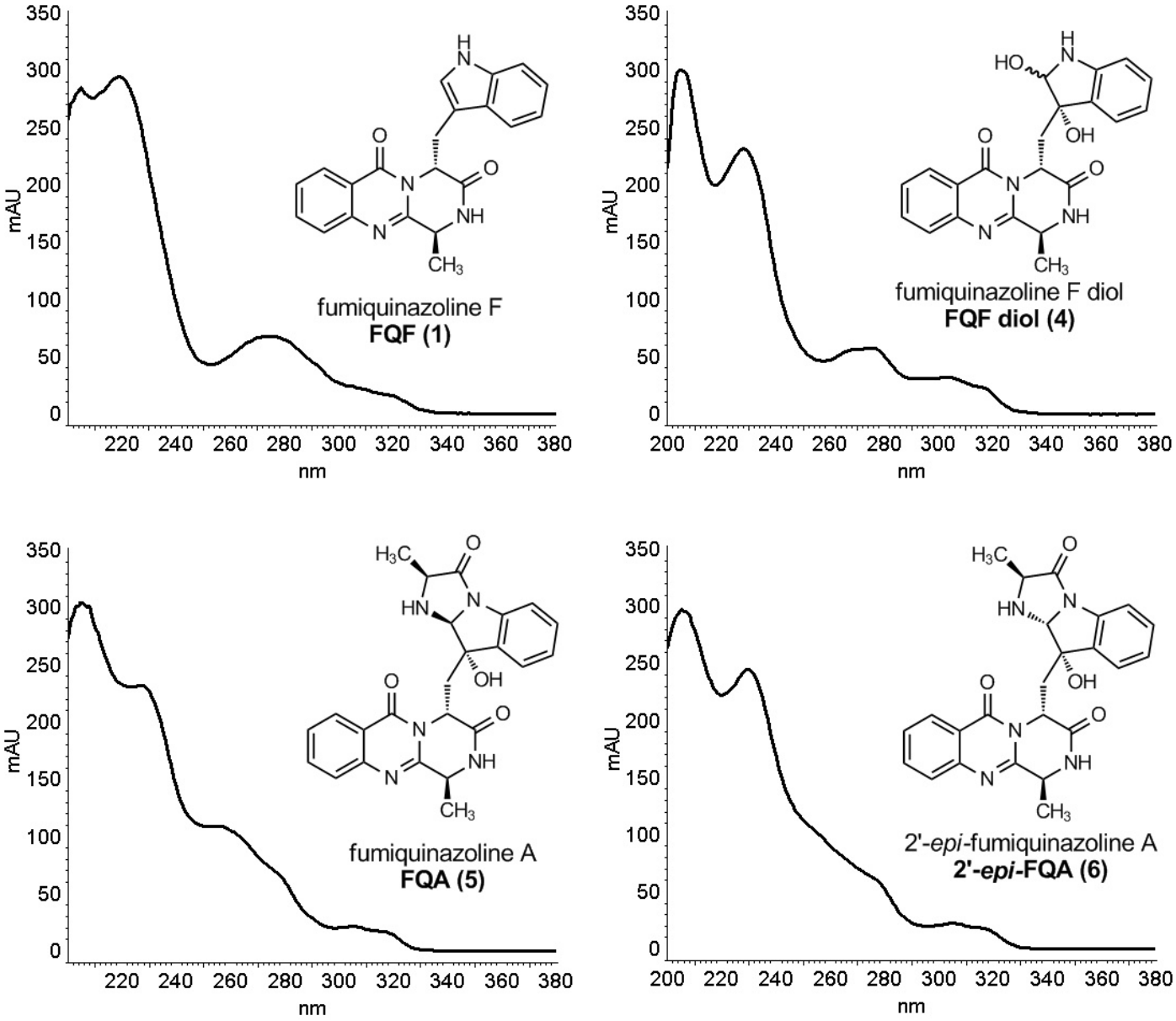
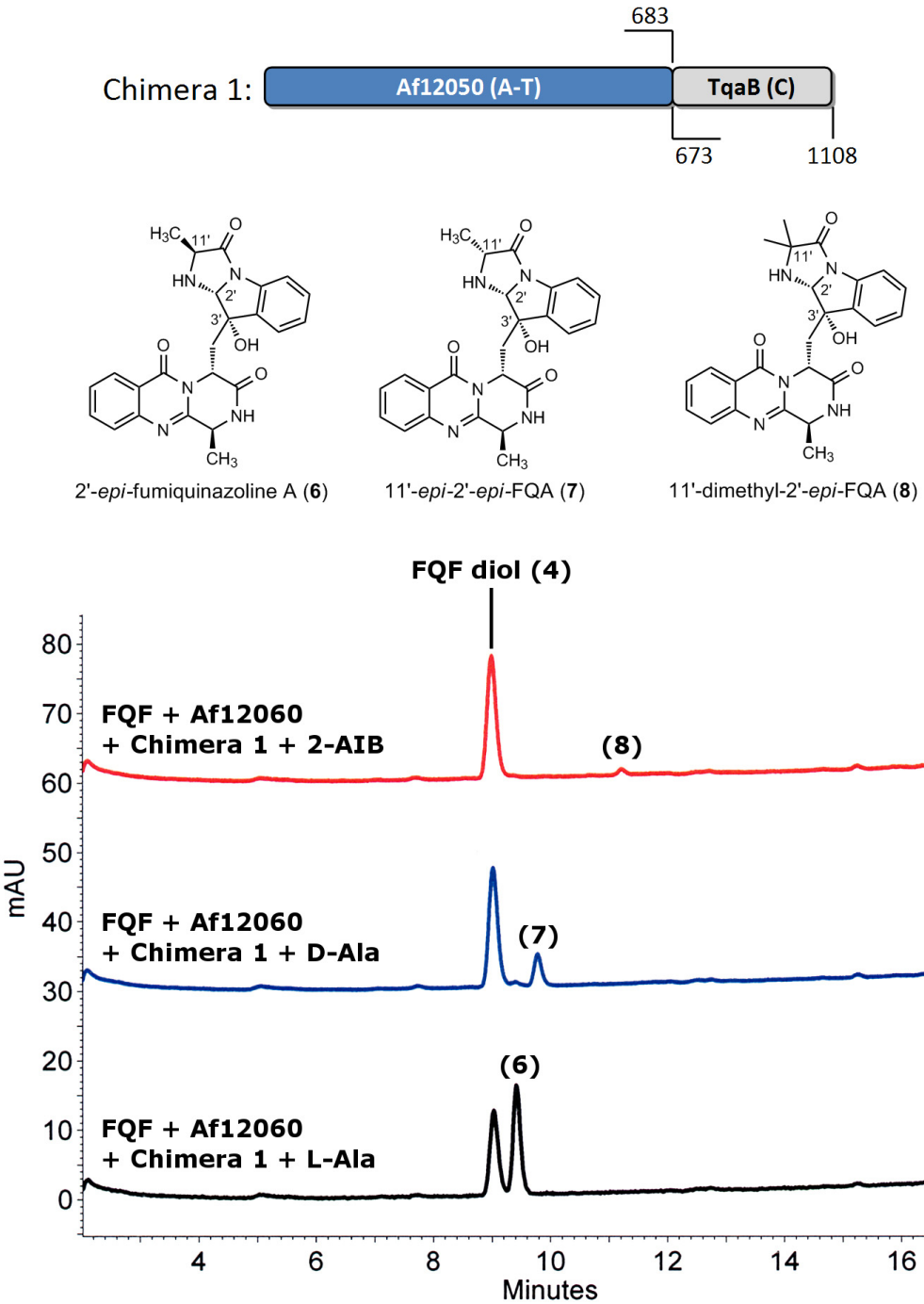


Figure S7: Analysis of substrate specificity of chimera 1 with respect to L-alanine, D-Ala and dimethyl analog 2-aminoisobutyic acid (2-AIB). Analyses were preformed using HPLC detection at 254 nm and identities of 2′-*epi*-FQA analogues were corroborated by UV and MS analysis.

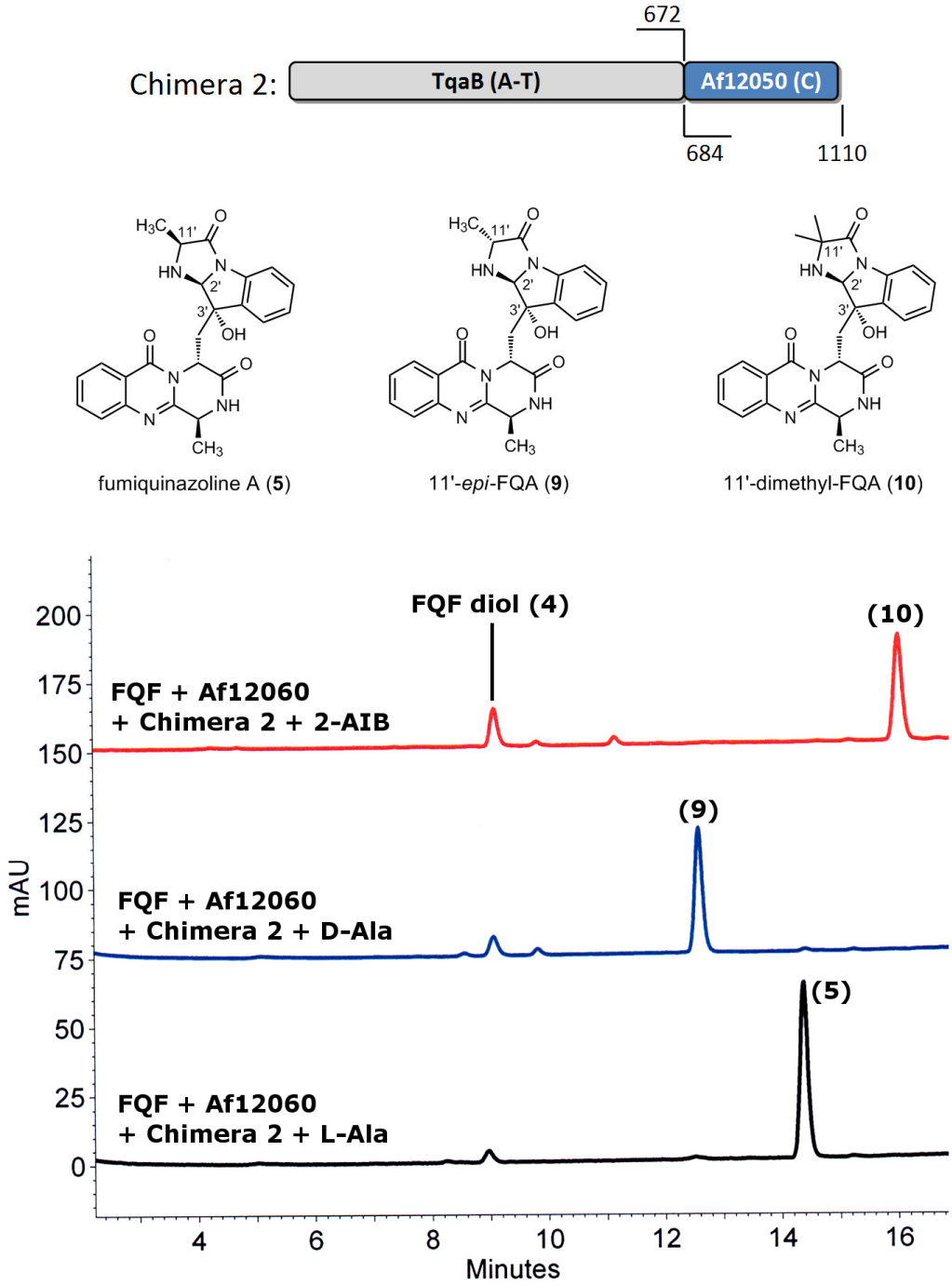


HRMS: Chimera 1 + 2-AIB (4) m/*z* calculated for $C_{21}H_{20}N_4O_4$: 375.1452 $[M-(H_2O) + H]^+$. Found 375.1453
(8) m/*z* calculated for $C_{25}H_{25}N_5O_4$: 460.1979 $[M + H]^+$. Found 460.1980

Chimera 1 + D-Ala (4) m/*z* calculated for $C_{21}H_{20}N_4O_4$: 375.1452 $[M-(H_2O) + H]^+$. Found 375.1454
(7) m/*z* calculated for $C_{24}H_{23}N_5O_4$: 446.1823 $[M + H]^+$. Found 446.1826

Chimera 1 + L-Ala (4) m/*z* calculated for $C_{21}H_{20}N_4O_4$: 375.1452 $[M-(H_2O) + H]^+$. Found 375.1449
(6) m/*z* calculated for $C_{24}H_{23}N_4O_4$: 446.1823 $[M + H]^+$. Found 446.1827

Figure S8: Analysis of substrate specificity of chimera 2 with respect to L-alanine, D-Ala and dimethyl analog 2-aminoisobutyic acid (2-AIB). Analyses were preformed using HPLC detection at 254 nm and identities of FQA analogues corroborated by UV and MS analysis.



HRMS: Chimera 2 + 2-AIB (4) m/*z* calculated for $C_{21}H_{20}N_4O_4$: 375.1452 $[M-(H_2O) + H]^+$. Found 375.1449
(10) m/*z* calculated for $C_{25}H_{25}N_5O_4$: 460.1979 $[M + H]^+$. Found 460.1977

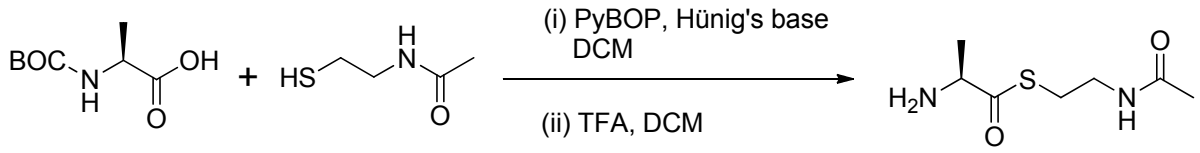
Chimera 2 + D-Ala (4) m/*z* calculated for $C_{21}H_{20}N_4O_4$: 375.1452 $[M-(H_2O) + H]^+$. Found 375.1451
(9) m/*z* calculated for $C_{24}H_{23}N_5O_4$: 446.1823 $[M + H]^+$. Found 446.1823

Chimera 2 + L-Ala (4) m/*z* calculated for $C_{21}H_{20}N_4O_4$: 375.1452 $[M-(H_2O) + H]^+$. Found 375.1452
(5) m/*z* calculated for $C_{24}H_{23}N_4O_4$: 446.1823 $[M + H]^+$. Found 446.1825

Figure S9: (A) Synthesis of L-Ala SNAC via PyBOP mediated coupling of *N*-BOC L-Ala with *N*-acetylcysteamine (B) Structure and synthetic route to L-Ala SNAC analogs used attempting to probe the route of dual N-C bond formation in FQA biosynthesis.

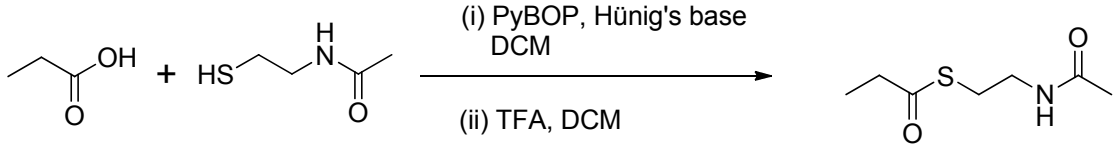
A

L-Ala SNAC

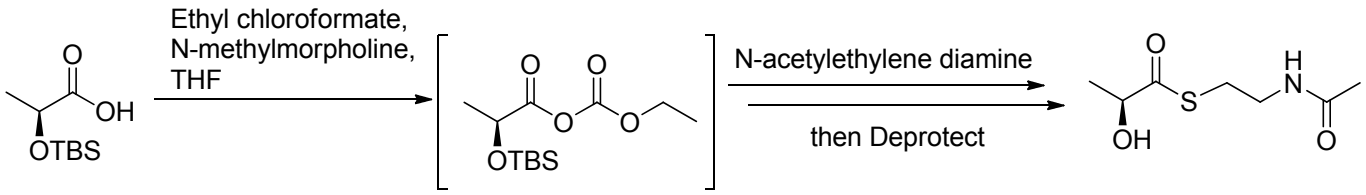


B

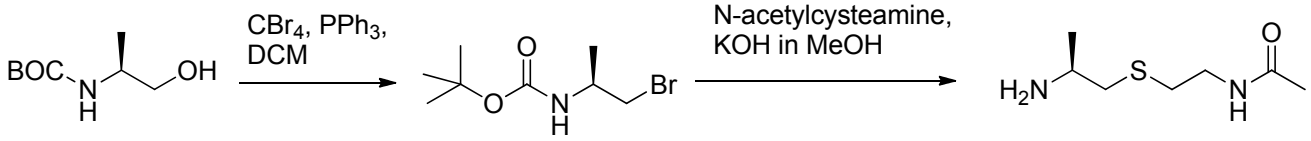
Propyl SNAC



L-Lactic acid SNAC



Alaninol SNAC



L-Ala amide

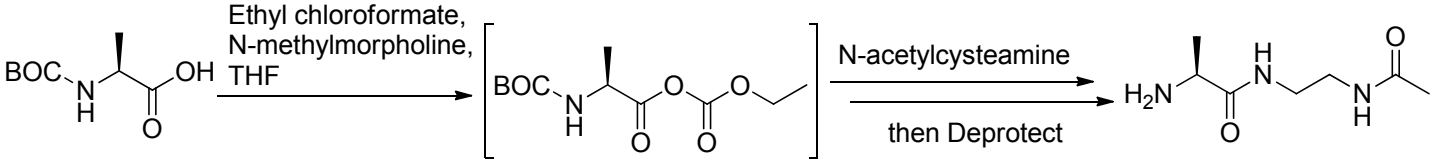
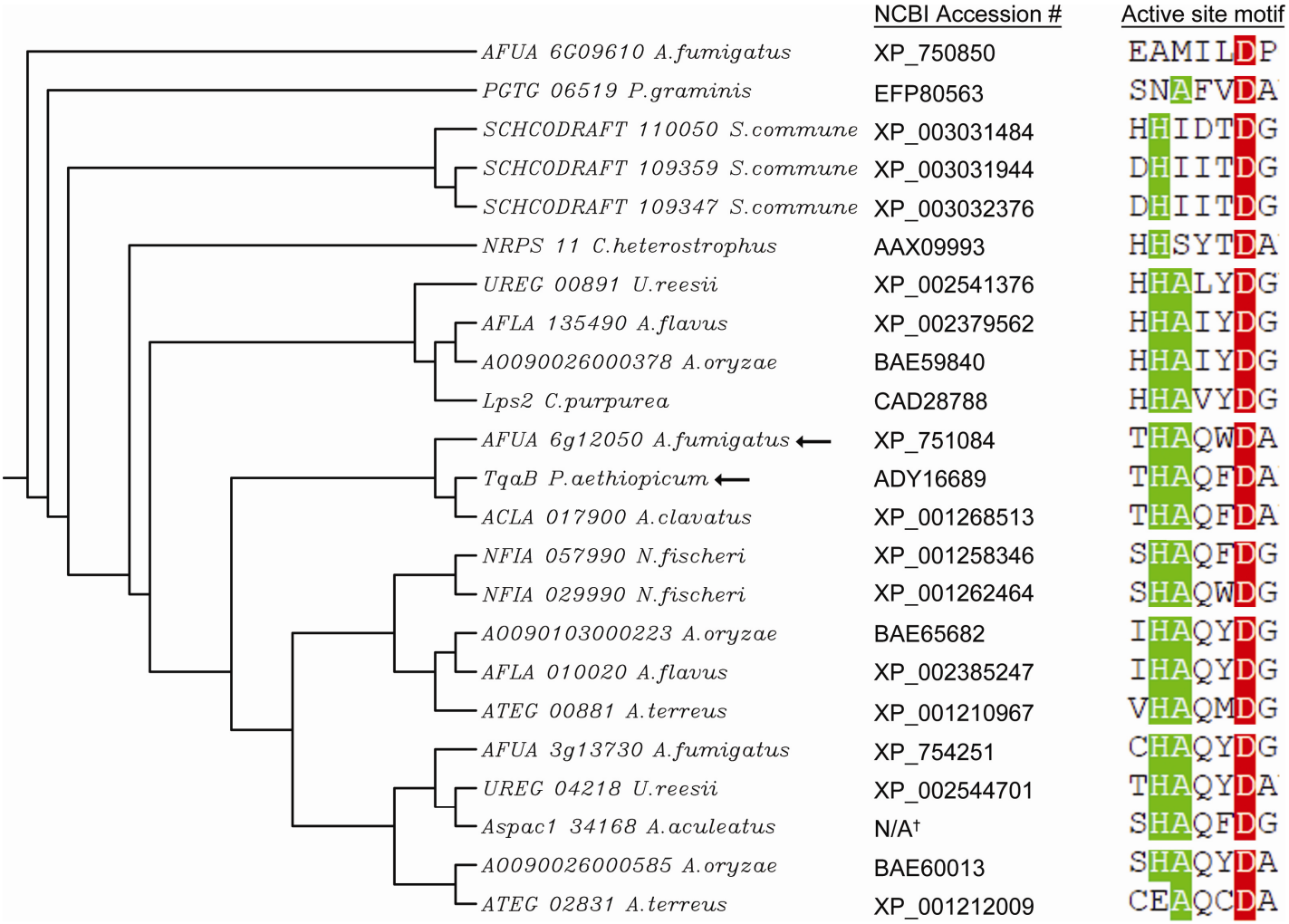


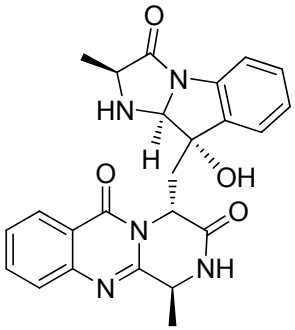
Figure S10: Neighbor-joining phylogenetic tree and active site motif for the C-domains of fungal monomodular NRPSs of domain organization A-T-C identified in this work. Arrows highlight the two C-domains characterized in this work. Sequence alignment and tree construction were performed using Clustal W (*10*).



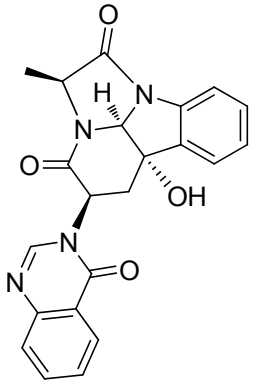
†The sequence for Aspac1_34168 from *Aspergillus aculeatus* was obtained from http://genome.jgi-psf.org/ with the FASTA header: jgi|Aspac1|34168|e_gw1.16.3.1. This sequence was not present in the NCBI database at the time of manuscript submission.

Figure S11: Stereochemical differences between fungal metabolites formed by proposed to be formed by the catalytic action of terminal C-domains, as highlighted by comparison with fumiquinazoline A and tryptoquialanine (derived from 2′-*epi*-FQA).

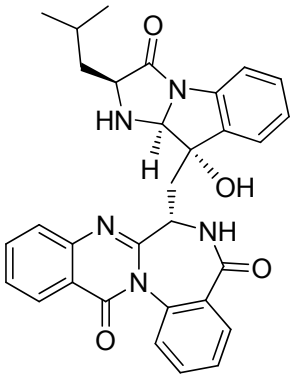
(A) C-N bond stereochemistry '*trans*-' to orientation of oxy-indole derived hydroxyl.



Fumiquinazoline A

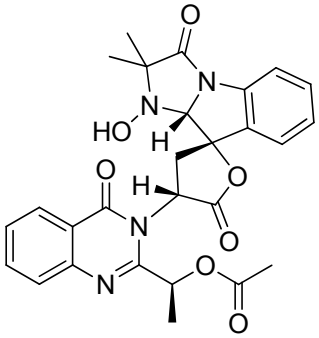


Chaetominine

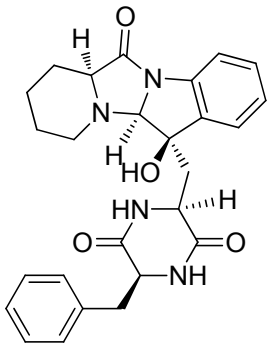


Asperlicin

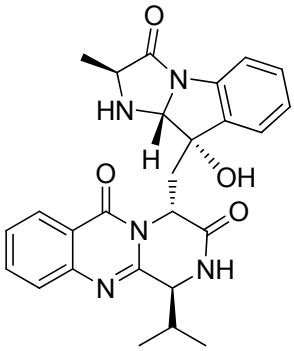
(B) C-N bond stereochemistry '*cis*-' to orientation of oxy-indole derived hydroxyl.



Tryptoquialanine A



Lumpidin



Fiscalin

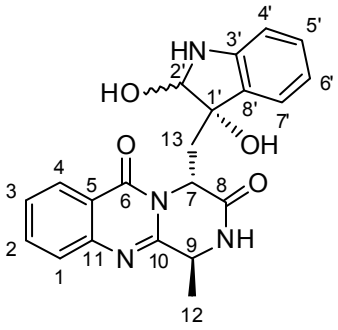
Table S1: Cloning primers used in this work for the generation of expression vectors

Primer	Sequence
Af12050_C-v1_for*	5′- GGTATTGAGGGTCGCcagctggagtcgtcactggtgg -3′
Af12050_C-v1_rev*	5'- AGAGGAGAGTTAGAGCCtcacaccgtaatctcagataacagagc -3′
TqaB_C-v1_for*	5′- GGTATTGAGGGTCGCcagctgaactcagctgtgacagatc -3′
TqaB_C-v1_rev*	5′- AGAGGAGAGTTAGAGCCttacccgttgatcttggagagcaattc -3′
Af12050_T_for*	5′- GACGACGACAAGATGgaccaaccGgtaacccagctggaag -3′
Af12050_T_rev*	5′- GAGGAGAAGCCCGGTTAgcgggggtcctgttcttggac-3′
TqaB_T_for*	5′- GACGACGACAAGATGggccagccGtcaacccagctc -3′
TqaB_T_rev*	5′- GAGGAGAAGCCCGGTTAcgatggtgtggttgacatggg -3′
Chimera-1_for*†	5′- CAGGGACCCGGTatgttggagacgacggagcc -3′
Chimera-1_int_rev†	5′- ctgtcacagctgagttcagctggaaggaggtgcagccgcggg -3′
Chimera-1_int_for†	5′- cccgcggctgcacctccttccagctgaactcagctgtgacag -3′
Chimera-1_rev*†	5′- GGCACCAGAGCGTTcccgttgatcttggagagcaattc -3′
Chimera-2_for*†	5′- CAGGGACCCGGTatggctagcatgactggtggacag -3′
Chimera-2_int_rev†	5′- ccaccagtgacgactccagctggaatgataccgacgatggtgtgg -3′
Chimera-2_int_for†	5′- ccacaccatcgtcggtatcattccagctggagtcgtcactggtgg -3′
Chimera-2_rev*†	5′- GGCACCAGAGCGTTcaccgtaatctcagataacagagc -3′
TqaH_NdeI_for:	5′- AAAAAACATatgacaaccgaccgtcaaccat-3′
TqaH_EcoRI_rev:	5′- TTTTTTGAATTCtcacccgcgcttgtatgtattc-3′

*Lowercase type signifies nucleotides complementary to the gene of interest. Uppercase type represents the LIC overhangs used for cloning (5’ end) or bases to be mutated for codon optimization.

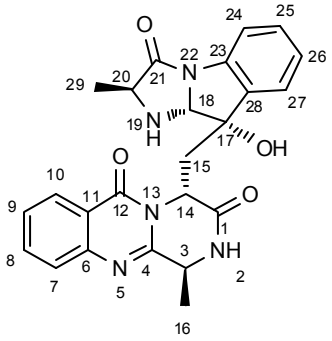
†Underlined portion is Af12050-specific, plain-type is TqaB specific.

Table S2: NMR spectroscopic characterization and assignments of FQF diol. Compound observed as a 1:1 mixture of diastereomers at C2′. Diastereomer for which each signal is responsible is indicated by subscript A or B.



	^{1}H	^{13}C
1_A	7.70	127.49
1_B	7.67	127.49
2_A	7.83	136.38
2_B	7.83	136.38
3_A	7.56	128.71
3_B	7.53	128.71
4_A	8.16	127.26
4_B	8.10	127.26
5_A	-	120.82
5_B		120.82
6_A	-	Not seen
6_B		
7_A	5.75	54.37
7_B	5.53	53.81
8_A	-	171.82
8_B		171.20
9_A	4.93	49.83
9_B	4.88	49.96
10_A	-	153.92
10_B		153.53
11_A	-	147.34
11_B		147.34
12_A	1.63	16.86
12_B	1.61	16.86
13_A	2.54 and 2.57	33.90
13_B	2.39 and 2.28	38.67
$1'_A$	-	81.51
$1'_B$		80.03
$2'_A$	4.97	89.87
$2'_B$	4.98	87.67
$3'_A$	-	149.13
$3'_B$		148.10
$4'_A$	6.67	111.59
$4'_B$	6.63	111.46
$5'_A$	7.12	131.22
$5'_B$	7.07	130.71
$6'_A$	6.72	120.01
$6'_B$	6.72	120.43
$7'_A$	7.18	124.08
$7'_B$	7.28	125.14
$8'_A$	-	131.72
$8'_B$		130.99

Table S3: NMR spectroscopic characterization and assignments of 2′-*epi*-FQA



	^{1}H (multiplicity, coupling constant, integral value)	^{13}C
1	-	168.95
2 (-NH)	6.03 (s, 1 H)	-
3	4.80 (q, *J*=6.75 Hz, 1 H)	48.87
4	-	150.11
5	-	-
6	-	146.29
7	7.71 (d, *J*=8.22 Hz, 1 H)	127.04
8	7.79 (m, 1 H)	134.26
9	7.53 (m, 1 H)	127.01
10	8.32 (d, *J*=7.92 Hz, 1 H)	126.57
11	-	119.85
12	-	160.03
13	-	-
14	5.89 (dd, *J*=8.22, 3.96 Hz, 1 H)	51.33
15_A 15_B	2.75 (dd, *J*=14.53, 3.96 Hz, 1 H) 2.46 (dd, *J*=14.53, 8.36 Hz, 1 H)	39.05
16	1.81 (d, *J*=6.75 Hz, 3 H)	16.83
17	-	73.88
18	5.44 (s, 1 H)	80.54
19	Not seen	-
20	4.00 (q, *J*=7.04 Hz, 1 H)	59.92
21	-	173.59
22	-	-
23	-	137.47
24	7.53 (m, 1 H)	115.42
25	7.33 (t, *J*=7.63 Hz, 1 H)	129.63
26	7.11 (t, *J*=7.63 Hz, 1 H)	124.79
27	7.30 (d, *J*=7.63 Hz, 1 H)	123.47
28	-	136.56
29	1.55 (d, *J*=7.04 Hz, 3 H)	18.30
-OH	4.85 (br. s, 1 H)	-