UCLA
UCLA Electronic Theses and Dissertations

Title

Investigation, Characterization and Engineering of Fungal Natural Product Biosynthesis

Permalink

Author

Gao, Xue

Publication Date

2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Investigation, Characterization and Engineering of Fungal

Natural Product Biosynthesis

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Chemical Engineering

By

Xue Gao

2013

ABSTRACT OF THE DISSERTATION

Investigation, Characterization and Engineering of Fungal Nature Products Biosynthesis

By

Xue Gao

Doctor of Philosophy in Chemical Engineering

University of California, Los Angeles, 2013

Professor Yi Tang, Chair

Nature products from filamentous fungi are extremely important sources of bioactive and structurally diverse compounds for agricultural and pharmaceutical applications. Understanding the biosynthetic machinery of fungal natural products will not only reveal novel enzymes that catalyze complicated chemical reactions, but also enable us to use these enzymes as powerful tools to make fine chemicals and high value drugs. Here, we revealed the biosynthetic pathways of a variety of fungal peptidyl alkaloids, including the anthranilate-containing tryptoquialanines (*tqa*), fumiquinazolines (*fqa*), asperlicins and ardeemin. By systematically inactivating every biosynthetic gene in the gene clusters, followed by isolation and characterization of the intermediates, we were able to establish the biosynthetic sequence of each pathway. Notably, the *tqa* pathway has been confirmed to go through an intermediate common to the *fqa* pathway, fumiquinazoline F (FQF), which originates from a fungal trimodular nonribosomal peptide synthetase (NRPS). Furthermore,

cyclization of linear peptidyl precursors produced by NRPSs is an important step in the biosynthesis of bioactive cyclic peptides. Whereas bacterial NRPSs use thioesterase domains to perform the cyclization, fungal NRPSs producing macrocyclic peptides terminate with a condensation-like (C_T) domain. To probe the role of such a C_T domain, we reconstituted the activities of the *Penicillium aethiopicum* trimodular NPRS TqaA in *Saccharomyces cerevisiae* and *in vitro*. Extensive biochemical and mutational studies confirmed the essential role of the C_T domain in catalyzing cyclization in a thiolation domain–dependent fashion. After the formation of FQF, TqaH in *tqa* pathway and Af12060 in *fqa* pathway were demonstrated to be identical and perform indole ring oxidation. Annulation of the indole side chain of FQF with different stereochemistry was controlled by the condenstation domain of monomodular NRPS enzymes (Af12050 and TqaB) to form fumiquinazoline A (FQA) and *epi*-FQA, respectively. The bimodular NRPS AspA in the asperlicin biosynthetic pathways is also characterized and synthesize multicyclic fungal alkaloids asperlicin C and D *in vitro*. The first module of AspA iteratively activates two molecules of anthranilate, while the C_T domain cyclizes the linear peptide into a macrolactam which can undergo further transannular cyclization. Finally, we employed directed evolution to engineer an enzyme, LovD which was identified from the lovastatin biosynthetic pathway and can be used in the synthesis of the cholesterol-lowing drug, Simvastatin (Zocor®). The best LovD mutant displayed an 11-foldincrease in an *Escherichia coli*-based biocatalytic platform. Seven X-ray crystal structures were determined to understand the structural basis of LovD enzymology.

The dissertation of Xue Gao is approved by

Todd O. Yeates

James C. Liao

Tatiana Segura

Yi Tang, Committee Chair

University of California, Los Angeles

2013

Table of Contents

List of Figures

viii

List of Schemes

List of Tables

Acknowledgments

Section 2.4 contains reprinted material with permission from Haynes, S. W., Gao, X., Tang, Y., Walsh, C. T. "Assembly of asperlicin peptidyl alkaloids from anthranilate and tryptophan: a two enzyme pathway generates heptacyclic scaffold complexity in asperlicin E." *J. Am. Chem. Soc.,* 2012, 134, 17444-17447. Copyright © 2012 American Chemical Society.

Section 2.4 contains reprinted material with permission from Gao, X., Jiang, W, Jiménez-Osés, G, Choi S. M., Houk K.N., Yi Tang, Walsh C.T., "An iterative, bimodular nonribosomal peptide synthetase that converts anthranilate and tryptophan into tetracyclic asperlicins." *Chem. Biol., accepted in principle.*

Section 2.5 contains reprinted material with permission from Haynes, S. W., Gao, X., Tang, Y., Walsh, C. T., "Complexity generation in fungal peptidyl alkaloid biosynthesis: a two enzyme pathway to the hexacyclic mdr export pump inhibitor ardeemin." *ACS Chem. Biol.* 2013, DOI: 10.1021/cb3006787. Copyright 2013, American Chemical Society.

Section 2.6 contains reprinted material with permission from Gao, X., Xie, X., Pashkov, I., Sawaya, R. M., Laidmen, J., Zhang, W., Cacho, R., Yeates, T. O., Tang, Y. "Directed evolution and structural characterization of a simvastatin synthase." *Chem. Biol.*, 2009, 16, 1064-1074. Copyright © 2009, Elsevier Ltd.

answering my fundamental questions. Dr. Stuart W. Haynes, Dr. Brian D. Ames and Dr. Wei Jiang (Harvard Medical School) are thanked for helpful discussion during our collaborations on fungal natural product biosynthesis. They set great examples for my near future postdoctoral research life. Dr. Gonzalo Jiménez-Osés is thanked for his discussion and contribution to the computational analysis of enzymes and small molecules. I want to greatly thank Dr. Michael R. Sawaya, Dr. Duilio Cascio and Inna Pashkov for their help on the crystallization of LovD and TqaM. Dr. Yitheng Chooi is thanked for establishing the fungal transformation system. Yeting Lai is thanked for assistance with gel filtration FPLC. Yuewei Sheng is thanked for her helpful discussions on the enzyme metal contents. Dr. Jacklyn Winter is thanked for revising my research manuscripts and proposals.

Personally, I would also like to thank my advisor Prof. Yi Tang for his understanding and support to me and my husband Peng Wang. It was the opportunity that he provided us to meet in USA. In particular, I want to greatly thank him and his wife Michelle for their hospitalities to invite us to celebrate all the Chinese festivals and let us feel like home. I want to thank my Master advisor Prof. Yingjin Yuan from Tianjin University for his strong recommendation and support for me to study at UCLA. Also, I want to thank Prof. Yufeng Lu for his support and care since very beginning. I am thankful to Dr. Yanran Li, Anuradha Biswas and Angelica Zabala for being great colleagues, best friends and lab sisters. Whenever I need help, they are always there for me.

I would like to thank my parents and family for their tremendous support and let me pursue my dreams no matter the distance. Thank my mom and dad for always being so patient and caring to listen to all my happiness and concerns everyday and letting me know they will

always be my side. Very especially, I want to thank my dear husband Peng Wang. Thank him for listening and discussing to my research ideas, staying with me in the lab until 3 am for numerous nights, getting dry ice whenever I need it, proof-reading my manuscripts when I was tired and became impatient, calming me down when I was angry and cheering me up when I was depressed. Without him, I won't be able to go through all the difficulties. I feel incredibly lucky and happy to be with him in the past and for the rest of my life.

Finally, I am indebted for the past six years at UCLA. This experience is the most valuable treasure of mine and I will cherish it for my life time.

VITA

2001-2005   B. S., Pharmaceutical Engineering

Tianjin University

Tianjin, PRC

2005-2007   M. S., Pharmaceutical Engineering

Tianjin University

Tianjin, PRC

2007-2013   Research Assistant

Department of Chemical and Biomolecular Engineering

University of California, Los Angeles, USA

2010-2011   Teaching Assistant

Department of Chemical and Biomolecular Engineering

University of California, Los Angeles, USA

2012-2013   Dissertation Year Fellowship

University of California, Los Angeles

Los Angeles, California, USA

Publications and Presentations

1. **Gao, X.**, Jiang, W, Jiménez-Osés, G, Choi S. M., Houk K.N., Tang, Y., Walsh C.T., "An iterative, bimodular nonribosomal peptide synthetase that converts anthranilate and tryptophan into tetracyclic asperlicins." ***Chem. Biol., in press.***

2. **Gao**, **X**., Haynes, S. W., Ames, B. D., Wang, P., Vien, L. P., Walsh, C. T., Tang, Y., "Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain." ***Nat. Chem. Biol.,*** 2012, 8, 823-830.

3. **Gao**, **X**., Chooi, Y., Ames, B. D., Wang, P., Walsh, C. T., Tang, Y., "Fungal quinazoline alkaloid biosynthesis: Genetic and biochemical investigation of the tryptoquialanine pathway in *Penicillium aethiopicum*." ***J. Am. Chem. Soc.,*** 2011, 133, 2729-2741. (***Highlighted by Nat. Chem. Biol.***)

4. **Gao**, **X**., Xie, X., Pashkov, I., Sawaya, R. M., Laidmen, J., Zhang, W., Cacho, R., Yeates, T. O., Tang, Y., "Directed evolution and structural characterization of a simvastatin synthase." ***Chem. Biol.***, 2009, 16, 1064-1074. (***2012 Presidential Green Chemistry Challenge Award***)

5. **Gao**, **X**., Wang, P., Tang, Y., "Engineered polyketides biosynthesis and biocatalysis in *Escherichia coli*." ***Appl. Microbiol. Biotechnol.,*** 2010, 88, 1233-1242.

6. Wang, P., **Gao**, **X**., Tang, Y., "Redox enzymes to generate complexity of natural products." ***Curr. Opin. Chem. Biol.,*** 2012, 16, 362-369.

7. Haynes, S. W., **Gao, X.,** Tang, Y., Walsh, C. T., "Complexity generation in fungal peptidyl alkaloid biosynthesis: a two enzyme pathway to the hexacyclic mdr export pump inhibitor ardeemin." ***ACS Chem. Biol***. **2013**, DOI: 10.1021/cb3006787.

8. Haynes, S. W., **Gao**, **X**., Tang, Y., Walsh, C. T., "Assembly of asperlicin peptidyl alkaloids from anthranilate and tryptophan: a two enzyme pathway generates heptacyclic scaffold complexity in asperlicin E." ***J. Am. Chem. Soc.,*** 2012, 134, 17444-17447.

9. Wang, P., **Gao**, **X**., Deng, Z., Tang, Y., "Genetic characterization of enzymes involved in the priming steps of oxytetracycline biosynthesis in *Streptomyces rimosus*." ***Microbiology***, 2011, 157, 2401-2409.

10. Xie, X., Pashkov, I., **Gao**, **X**., Guerrero, J., Yeates, T. O., and Tang, Y., "Rational improvement of simvastatin synthase solubility in *Escherichia coli* leads to higher whole-cell

biocatalytic activity." ***Biotechnol. Bioeng.***, 2009, 102, 20-28.

11. Ames, B. D., Haynes, S. W., **Gao**, **X**., Tang, Y., Walsh, C. T., "Complexity generation in fungal peptidyl alkaloid biosynthesis: oxidation of fumiquinazoline A to the heptacyclic hemiaminal fumiquinazoline C by the flavoenzyme Af12070 from *Aspergillus fumigatus*." ***Biochemistry***, 2011, 50, 8392–8406.

12. Haynes, S. W., Ames, B. D., **Gao**, **X**., Tang, Y., Walsh, C. T., "Unraveling terminal C-domain-mediated condensation in fungal biosynthesis of imidazoindolone metabolites." ***Biochemistry,*** 2011, 50, 5668-5679.

13. Walsh, C. T.*, Haynes, S. W., Ames, B. D., **Gao, X**., Tang, Y., "Short pathways to complexity generation: fungal peptidyl alkaloid multicyclic scaffolds from anthranilate building blocks." ***ACS Chem. Biol.****, doi: 10.1021/cb4001684.*

14. Wang, P., Kim, W., Pickens, L. B., **Gao**, **X**. and Tang, Y., "Heterologous expression and manipulation of three tetracycline biosynthetic pathways." ***Angew. Chem. Int. Ed. Engl.****, doi: 10.1002/anie.201205426*

15. Wang, P., Bashiri, G., Sawaya, M. R., **Gao, X.**, Yeates T. O., Tang, Y., "Uncovering the enzymatic basis of the final steps in oxytetracycline biosynthesis." ***J. Am. Chem. Soc.****, doi: 10.1021/ja403516u.*

16. Jiménez-Osés, G., Osuna S., **Gao, X.**, Yeates T. O., Tang, Y., Houk K. N., "From induced-fit to preorganization: unveiling the evolutionary pathways in engineered simvastatin synthases." *Manuscript in Preparation*.

17. Directed evolution and structural characterization of a simvastatin synthase (Poster, 2012), *Zing conferences-biocalalysis,* Xcaret, Cancun, Mexico.

18. Heterologous expression and manipulation of three tetracycline biosynthetic pathways (Orall presentation, 2012), *America Institute of Chemical Engineers annual meeting*, Pittsburgh, PA, USA.

19. Biochemical investigation of the cyclization of fungal nonribosomal peptides catalyzed by a terminal condensation-like domain (Oral presentation, 2012), *243st ACS National Meeting & Exposition*, San Diego, CA, USA.

20. Genetic and biochemical investigation of tryptoquialanine biosynthesis in *Penicillium aethiopicum* (Poster, 2011), *241st ACS National Meeting & Exposition*, Anaheim, CA, USA

1   Introduction

Fungal secondary metabolites produced by the filamentous fungi consist of a large variety of bioactive natural products, which include both valued pharmaceuticals and detrimental mycotoxins [1]. Among these low-molecular-weight natural products, fungal secondary metabolites can also be divided to several classes: polyketdies, nonribosomal peptides and terpenes [1, 2]. Over the past decade, genome sequencing and molecular biology techniques have greatly facilitated our studying of fungal secondary metabolism pathways. Numerous gene clusters have been identified by heterologous expression or genome mining [3, 4]. Unlike primary metabolism, the biosynthetic genes of fungal secondary metabolites are normally clustered together either by horizontal transfer in fungi or uptake from prokaryotic organisms, which greatly facilitate our detection by bioinformatics algorithms [5, 6]. Genome sequencing of several *Aspergillus* and *Penicillium* strains in the past few years have indicated two dozen to three dozen nonribosomal peptide synthetase (NRPS)-associated gene clusters per fungal genome [7, 8]. Only a fraction of these gene clusters have been matched to known metabolites, reflecting cryptic biosynthetic capability under laboratory growth conditions [9, 10]. An equivalent number of polyketide synthase gene clusters are also found in these fungal genomes as the second major class of natural product encoding genes [10]. Understanding the biosynthetic mechanism will enrich our knowledge on the direct prediction of the end fungal product structures based on the genetic sequences. Meanwhile, revealing the novel enzymes involved in fungal secondary metabolism pathways can also shed lights on the efficient synthesis of chemicals and clinical drugs [11].

Fungal natural product including many clinically useful drugs such as the antibiotic penicillin [12], the immunosuppressant drug cyclosporin [13] and the anti-hypercholesterolemia lovastatin [14]; nevertheless, many other metabolites are potentially toxic called mycotoxins, including gliotoxin and aflatoxins [15, 16] (**Scheme 1**). Fungi, especially Ascomycota, are prolific producers of peptidyl alkaloids, often isolated because of their toxic activities toward mammalian cells and tissues [17]. A characteristic feature of many of these conditional peptidic metabolites is the multi-cyclic, constrained architectures that can lead to high affinity for biological targets. The bioactive properties of fungal metabolites have triggered a growing interest in understanding the genetic basis of these structurally diverse compound biosynthesis pathways in the last decade.

Penicillin G

Cyclosporine A

Lovastatin

Gliotoxin

Aflatoxin B1
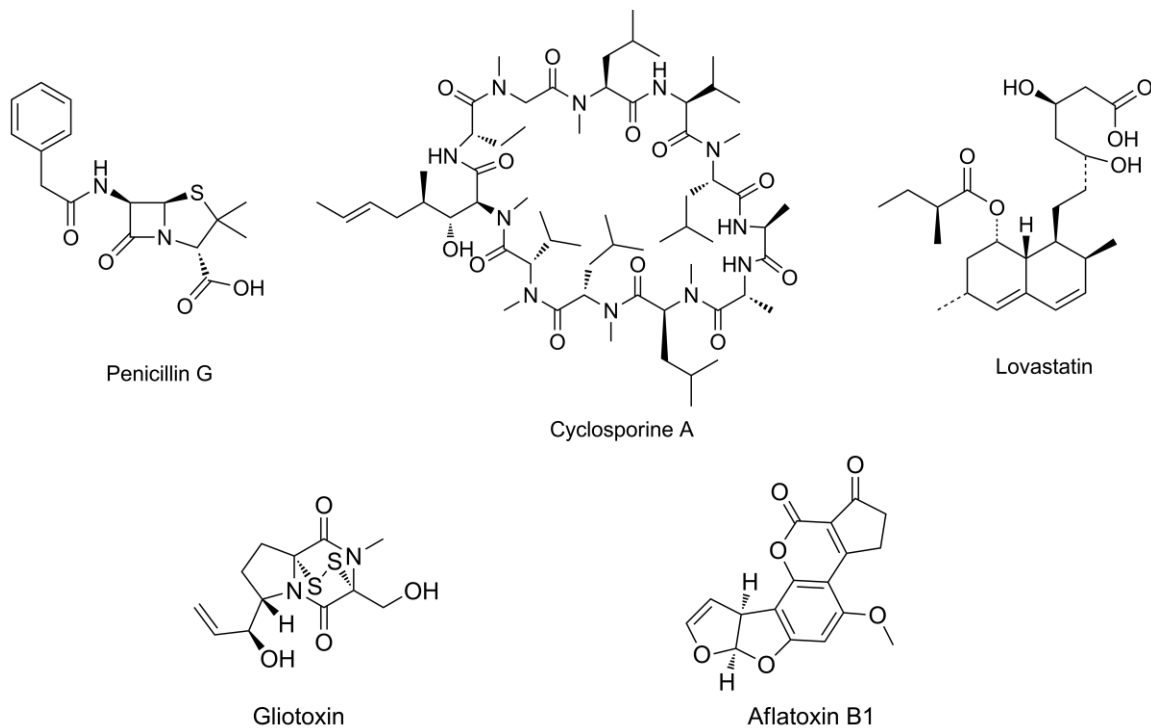
Scheme 1. Representative of fungal natural products.

2

1.1 The Biosynthesis of Fungal Nonribosomal Peptides

1.1.1 Fungal Nonribosomal Peptide Synthetases

Nonribosomal peptides (NRPs) are composed of an extremely important group of natural products, e.g. the immunosuppressant cyclosporin [18] and also echinocandin B [19], the scaffold for clinical antifungal drugs [20]. The core structures of these large peptidyl scaffolds are normally synthesized by NRPSs. NRPSs are generally mulitimodular megasthases minimally
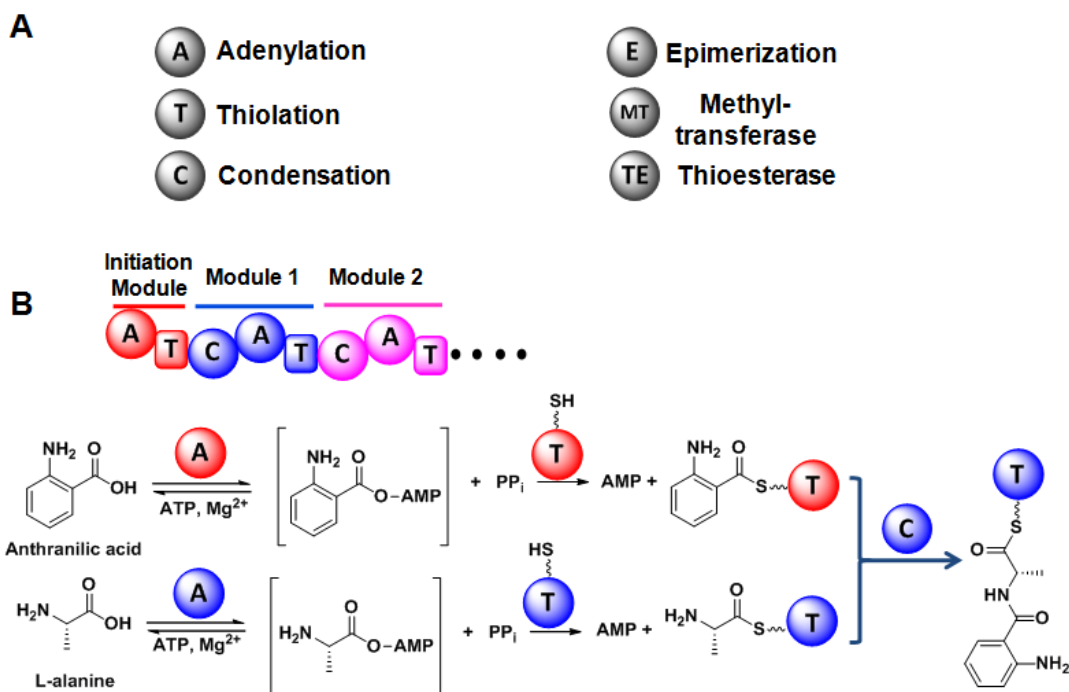
Figure 1. Representative NRPS domains (A) and mode of NRPS assembly line chain elongation (B). NRPSs are able to select, activate, and incorporate proteinogenic (L-alanine) or nonproteinogenic (anthranilic acid).

consist of adenylation (A), thiolation (T), and condensation (C) domains [21]. Generally, the process of building the peptidyl frameworks starts with each A domain using ATP and activating one amino acid to form the acyl-AMP intermediate. Meanwhile, T domains are subjected to the modification of phasphopanthetheinyl transferases to generate the active protein machinery [22]. The acyl-group of the acyl-AMP is then transferred and attached to the the 4'-phosphopantetheine (ppt) prosthetic arm of the T-domain to form an acyl-thioester intermediate. Two acyl-tethered intermediates will be delivered to the active site of the C domain which will catalyze the peptide bond formation to finish one round of peptidyl chain elongation (**Figure 1**). NRPS assembly lines are freed from the restrictions to use proteinogenic amino acid building blocks, mandated by the inventory of aminoacyl-tRNAs and their cognate aminoacyl-tRNA synthetases that enable ribosomal protein synthesis [23].

The fungal NRPS genes typically encode from one to six NRPS modules, (although the cyclosporin synthase involves eleven modules [24, 25]), correlating to assembly lines that can activate and modify single amino acids to those that can assemble hexapeptide scaffolds. Some of these fungal NRPS assembly lines release the free acid form of peptide products such as ACV synthetase [12] which yields the aminoadipyl-cysteine-valline tripeptide that is the immediate substrate for isopenicillin N synthetase and is the gateway to penicillins and subsequently cephalosporins. Many of the well-studied fungal NRPs arise from bimodular NRPS assembly lines; the nascent products released in those cases are the cyclic diketopiperazines (DKPs). The phenylalanine(Phe)-serine(Ser)-DKP and anthranilate(Ant)-tryptophan(Trp)-DKP as the building blocks for gliotoxin and acetylaszonalenin maturation, respectively (**Figure 2**) [16, 26]. The DKP formed between L-Trp and L-proline (Pro) is brevianamide F [27] that gets further

4

elaborated to the trypostatins and fumitremorgins by a series of prenylations and oxygenations effected by dedicated tailoring enzymes that modify the indole side chain of the Trp residue [28]. A variety of other Trp-X-DKPs, including the symmetric Trp-Trp-DKP [29] are similarly precursors for further elaboration of the nascent DKPs.
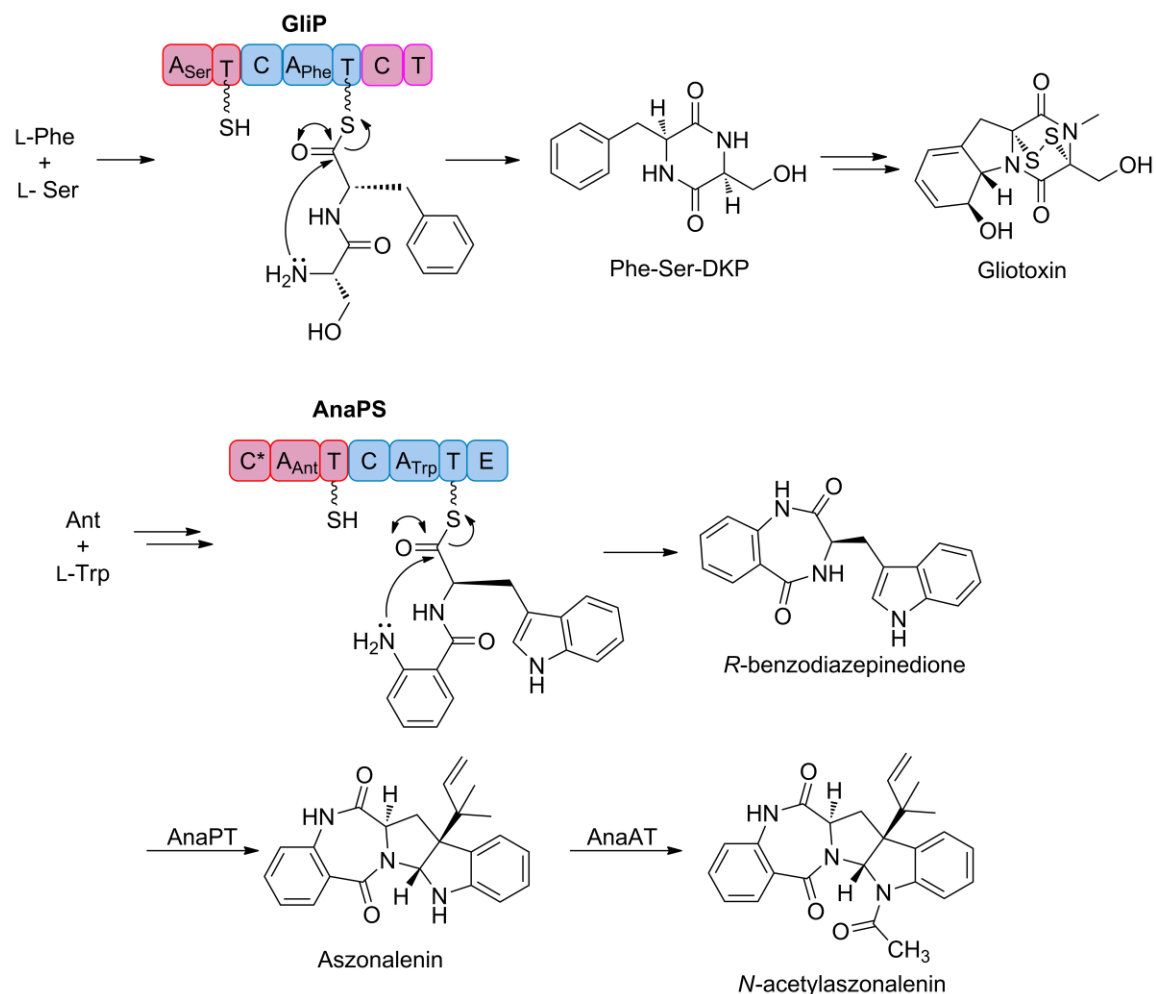
Figure 2. Two examples of fungal NRPs biosynthesis: gliotoxin and acetylaszonalenin. A = adenylation, C = condensation, T = thiolation, E = epimerization domain.

One of the hallmarks of NRPSs is the ability to select, activate, and incorporate nonproteinogenic amino acids as building blocks. The use of aminoadipate for penicillin biosynthesis by ACV synthetase [30] is among the most celebrated examples of this catalytic

versatility. A subset of fungal metabolites containing from two to seven fused nitrogen heterocycles (**Scheme 2**) have incorporated the noncanonical amino acid anthranilate (*ortho*-aminobenzoate) as a key building block. In this ariticle, we focused on the NRPS assembly lines that release the nascent products as bicyclic to tetracyclic nitrogen-containing frameworks (**Scheme 2**). The key building blocks are Ant [31], a nonproteinogenic amino acid, and its primary metabolite derivative L-Trp [32, 33]; the products are fused nitrogen heterocycles, not diketopiperazines. These bi- to tetracyclic nascent products then undergo tailoring reactions of oxygenation and/or prenylation that create additional fused ring systems, ranging from bicyclic to heptacyclic frameworks (also shown in **Scheme 2**).

Acetylaszonalenin (**Figure 2**) is the first well-studied Ant-containing peptidyl alkaloid and the key to decipher the chemical logic and enzymatic machinery for the whole class of anthranilate-derived multicyclic alkaloid scaffolds in fungal metabolism. The availability of the genome sequence of *Neosartorya fischeri* [34] enables the investigation of the regiospecificity and mechanism of the class of Mg^{2+}-independent prenyltransferases that act on distinct sites on the indole ring of tryptophan-based fungal metabolites [35-37]. The heterologously overexpression of both the prenyltransferase (AnaPT) and the neighboring acetyltransferase from *N. fischeri* in *E. coli* and using synthetic *R*- and *S*-benzodiazepinedione, which were proposed to be the prenyltransferase substrates if this were the aszonalenin pathway. AnaPT was observed to efficiently convert *R*-benzodiazepinedione to aszonalenin [26]. Also, the acetyltransferase (AnaAT) would then acetylate the indole nitrogen and yield the known acetylaszonalenin [26]. The aszonalenin prenyltransferase shows specificity for transfer to C_3 of the indole moiety of the *R*-benzodiazepinedione substrate. The resultant prenylated indole is now electron deficient at C_2

and is captured intramolecularly by one of the two amide nitrogens of the diazepinedione ring. This leads to construction of the fused 6-7-5-5-6 five ring system characteristic of the aszonalenin framework (**Figure 2**). Such prenyl transfers are paradigmatic for complexity generation not only in the molecules described below but also in Trp-X-DKP derivatives noted above.
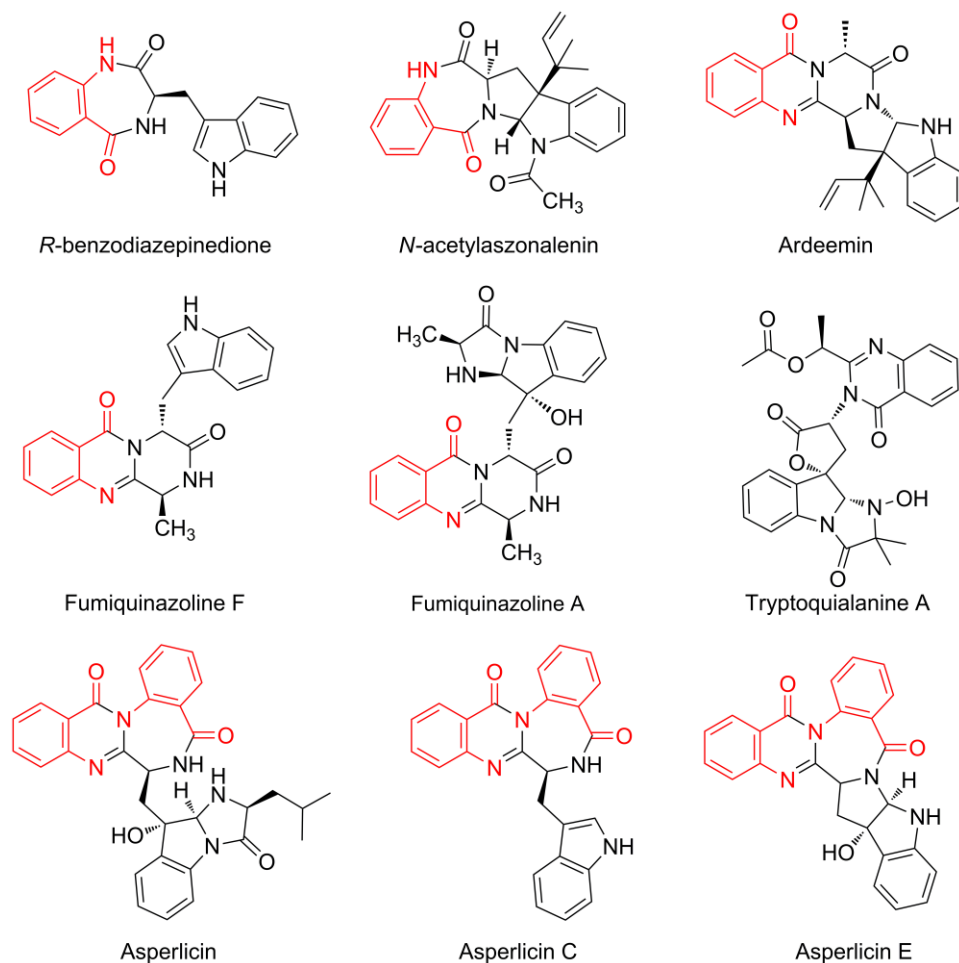
Scheme 2. Bicyclic to heptacyclic peptidyl alkaloid scaffolds from fungi utilizing anthranilate (Ant) as building block(s) (in red).

Followed by the biosynthesis of the benozdiazepinedione, later study was focused on the biomodular NRPS, AnaPS. To investigate whether the bicyclic benzodiazepinedione could arise

from either a tethered Trp_1-Ant_2-S-enzyme, released by attack of the NH_2 of the Trp moiety on the Ant thioester carbonyl, or from an Ant_1-Trp_2-S-enzyme, released by attack of the Ant-NH_2 on the activated Trp-thioester carbonyl, bioinformatic analysis of AnaPS (C^*-A_1-T_1-C_2-A_2-T_2-E_2) was performed and it suggested that module 2 contained a canonical amino acid activating Adenylation domain A_2 for activation of Trp, but the corresponding A_1 domain had an anomalous signature [38]. The potential Ant-activating domain was heterologously expressed C^*-A_1-T_1 in *E.coli* and indeed demonstrated reversible activation of Ant as anthranilyl-AMP and its subsequent covalent tethering in thioester linkage to the pantetheinyl arm posttranslationally introduced on T_1 (**Figure 1B**). A 10-residue code for Ant activation in fungal adenylation domains of NRPS enzymes [38] was proposed which opened the door for all the subsequent biosynthetic genetic and enzymatic studies noted in this article. To make the pentacyclic scaffold of aszonalenin from the two primary metabolites anthranilate and L-Trp only two enzymes are required, AnaPS and the prenyltransferase AnaPT. On the one hand it is a dramatic lesson in chemical efficiency and on the other hand presages the utility of Ant as a key unit in more complex multicyclic scaffolds.

Cyclization and Product releasing of Nonribosomal Peptides

A version of this section was published as: Gao, X., Haynes, S. W., Ames, B. D., Wang, P., Vien, L. P., Walsh, C. T., Tang, Y. "Cyclization of fungal nonribosomal peptides by a terminal condensation-Like domain." *Nat. Chem. Biol.,* 2012, 8, 823-830. Copyright © 2012, Rights Managed by Nature Publishing Group.

After maturation from the NRPS assembly line, linear peptide chain will generally go through the cyclization and product releasing process to form biologically active final products.

Cyclic NRPs from filamentous fungi represent a group of pharmaceutically important natural products, including the immunosuppressive cyclosporine A and the antifungal family of lipopeptides represented by echinocandin B [13, 20]. Biosynthesis of the linear precursors of these molecules by their corresponding nonribosomal peptide synthetases (NRPSs) parallels that of bacterial NRPs, such as the well-studied tyrocidine A and gramicidin [22, 39]. Following the sequential addition of each amino acid catalyzed by a NRPS module, which consists minimally of condensation (C), adenylation (A) and thiolation (T) domains, cyclization and product release take place through nucleophilic attack on the thioester carbonyl by a free amine from either the *N*-terminus or side chain of the peptide [39, 40]. Given the importance of cyclization to the biological activities of these molecules, understanding the enzymatic basis of this final step in NRPs biosynthesis is a significant goal from both mechanistic and biocatalytic perspectives [41, 42].
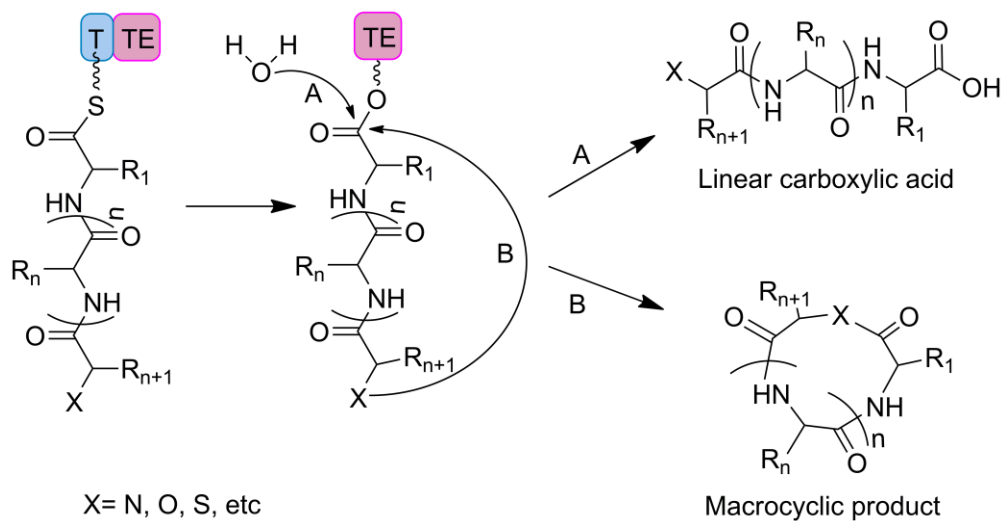
Figure 3. The general mechanism for TE catalyzed NRP releasing. The peptidyl chain could be hydrolyzed by water to produce linear carboxylic acid (route A) or intramolecular cyclized to form macrocyclic compound.

The classic and well-studied examples are the thioesterase (TE) domain catalyzed product releasing and macrocyclization in bacterial NRPS assembly lines. TE domains are

usually located at the C-terminus of the modular NRPSs, however, N-terminus TE was also found in the past few years [43]. The catalytic triad of TE domain consists of Ser-histidine (His)-aspartic acid (Asp). The general process starts with transferring the peptidyl chain from the last T domain. Asp residue possibly stabilizes the His residue which functions as the general base to deprotonate the hydroxyl of the active site Ser. The hydroxyl oxygen of the active site Ser acts as the nucleophile to attack the carbonyl of the T domain tethered peptidyl thioester and transfer the peptidyl to form the peptidyl ester on TE domain [44]. The TE domain bound peptidyl-*O*-Ser intermediates might be released from the enzyme by different methods. Firstly, water can serve as the nucleophile and hydrolyze the peptidyl chain from NRPS and release it as free carboxylic acid. Otherwise, an amine, a hydroxyl, or a thio group of the peptidyl chain will perform the intramolecular attack and lead to the formation of macrocyclic product. Such TE domains in the surfactin and fengycin biosynthesis have been biochemically and structurally well elucidated [45, 46]. Unlike the canonical TE domains mentioned above, some other bacterial TE domains with new characterstics have been studied recent years. One interesting example is the TE domain has the dual functions to catalyze both peptide dimerization and cyclization in a sequential manner. The T-TE didomain was excised from the NRPS TioS, which was involved in the biosynthesis of thicoraline [47]. It was demonstrated *in vitro* that the first half of the peptidyl chain was transferred to the active site Ser of TE domain, and the second half which has the exact same structure was synthesized and re-loaded to the upstream T domain. TE domain ligates these two identical parts and the full length of the peptidyl chain is subsequently attached to its active site Ser. Macrocyclization is also performed by the TE domain to release the final product,

thiocoraline (**Figure 4**) [47]. Similar mechanistic fashion was also observed in the GrsB TE domain for the biosynthesis of gramicidin S [48].
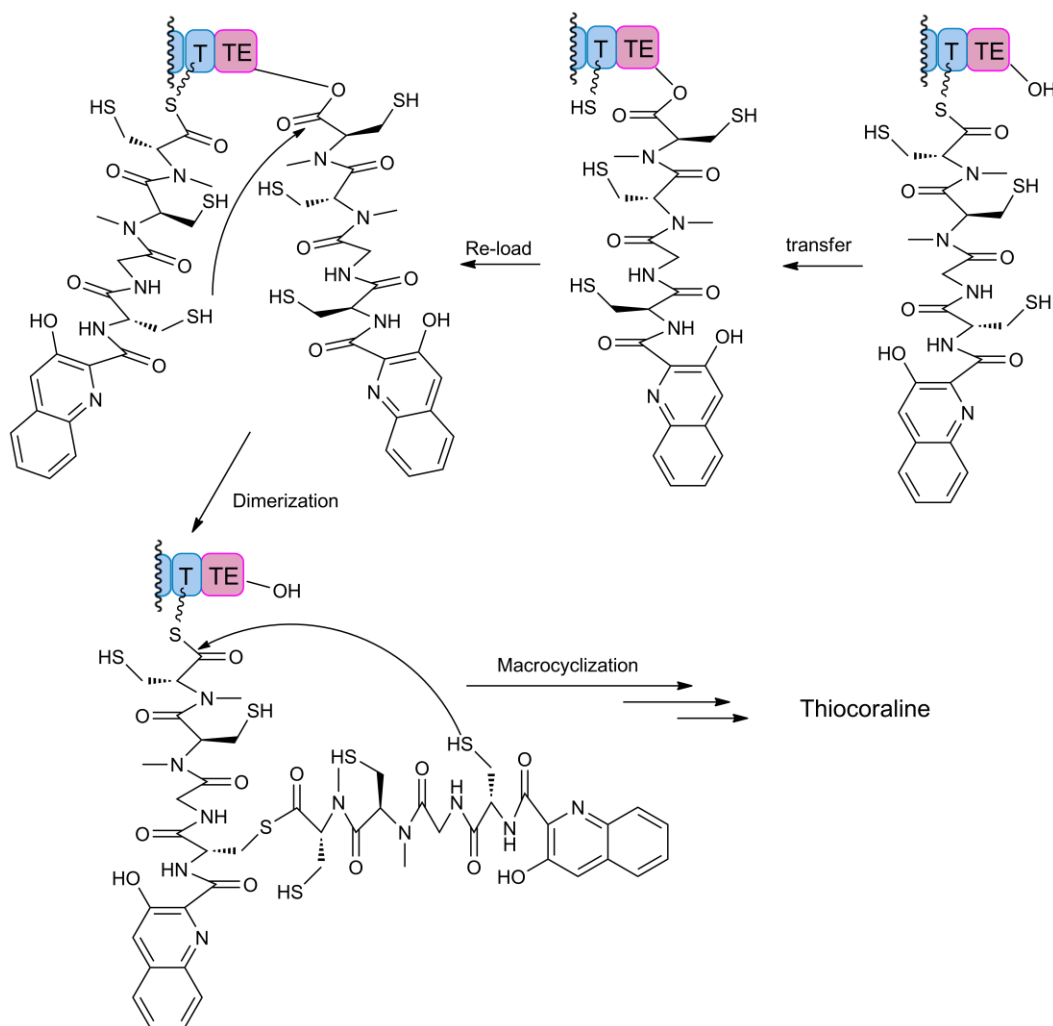
Figure 4. The TE-catalyzed dimerization and macrocyclization in thiocoraline biosynthesis.

In contrast with the extensively studied bacterial NRPSs product releasing mechanisms, fungal NRPSs are not that well characterized due to their relatively large size for heterologous expression and *in vitro* characterization, e.g. the cyclosporine synthetase is ~1,700 kDa. Whereas bacterial NRPSs use TE domains to perform the cyclization, fungal NRPSs have several different strategies for product release following completion of peptide elongation [44, 49]. From the

11

verified fungal NRPS gene clusters to date, none of the NRPSs associated with cyclic NRPs contain terminal TE domains, nor a free-standing TE-like domain encoded in the vicinity of the gene cluster [50]. TE domains have been found to be responsible for the hydrolysis of full length peptidyl-NRPS adducts as in the ACV synthetase that is responsible for the production of the ACV peptide precursor to penicillin [40]. NADPH-dependent reductive (R) domains have been found to be associated with the family of peptaibol linear peptides, in which the C-terminal thioester is reduced to an amino alcohol [51]. R and R-like domains are also found at the C-termini of PKS-NRPS hybrid megasynthetases which can catalyze reductive chain release or Dieckmann cyclization, respectively [52, 53].

1.1.2 Tailoring Enzymes Involved in Nonribosomal Peptide Biosynthesis

A version of this section was published as: Wang, P., Gao, X., Tang, Y. "Redox enzymes to generate complexity of natural products." *Curr. Opin. Chem. Biol.,* 2012, 16, 362-369. Copyright © 2012, Elsevier Ltd.

The structural complexity observed among natural products is a testament to nature's amazing ability to perform chemistry with biological catalysts. NRPs are a group of structurally and functionally diverse nature products which are synthesized by NRPSs [54]. In the pre-modification step, simple primary metabolites such amino acids are modified into more elaborate precursors by dedicated enzymes [54, 55]. These precursors are incorporated into the nascent natural product scaffold by highly programmed machineries, NRPSs [40]. The third level of complexity generation is performed by post-modification enzymes to complete the bioactive products [56]. While many different types of enzymes are recruited in the pre- and post-modification steps, redox enzymes such as FAD-dependent oxygenases, oxidoreductases,

cytochrome P450 oxygenases, and NADPH-dependent reductases are ubiquitous and arguably catalyze the most diverse and sophisticated transformations. Especially in the post-modification steps, efficient redox cascades can regioselectively and stereoselectively morph the relatively simple structures into highly complex final products. Redox enzymes such as FAD-dependent and cytochrome P450 oxygenases play indispensible roles in generating structure complex during natural product biosynthesis.

Pre-assembly modification in nonribosomal peptide biosynthesis

In the pre-assembly steps, redox enzymes can convert garden variety primary metabolites into unique starter and extender building blocks. Much of the structural complexity seen in NRPs is generated through the incorporation of nonproteinogenic amino acids, many of which are hydroxylated versions of the natural amino acids. Hydroxylation of amino acids can either place on the free amino acid, or on the activated aminoacyl moiety that is attached to the T domain (**Scheme 3**). For example, N^5-hydroxyornithine is found in siderophores coelichelin and erythrocelin, and is the product of L-ornithine N-hydroxylation by FAD-dependent monooxygeneases CchB and EtcB, respectively [57, 58]. In contrast, epoxidation of the succinamoyl building block by the α-ketoglutarate dependent DdaC to afford epoxysuccinamoyl takes place after N_β-fumaramoyl-L-2,3-diaminopropionate is transferred to the NRPS DdaD [59]. T-domain dependent oxidation is also observed in the formation of 3, 5-dihydroxy-4-methylanthranilate found in sibiromycin. Following NRPS-independent modification of 3-hydroxykynurenine to yield 3-hydroxy-4-methylanthranilic acid (3H4MAA), 3H4MAA is transferred to the NRPS SibE and is hydroxylated by the FAD-dependent hydroxylase SibG to give 3, 5-dihydroxy-4-methylanthranilate [60].
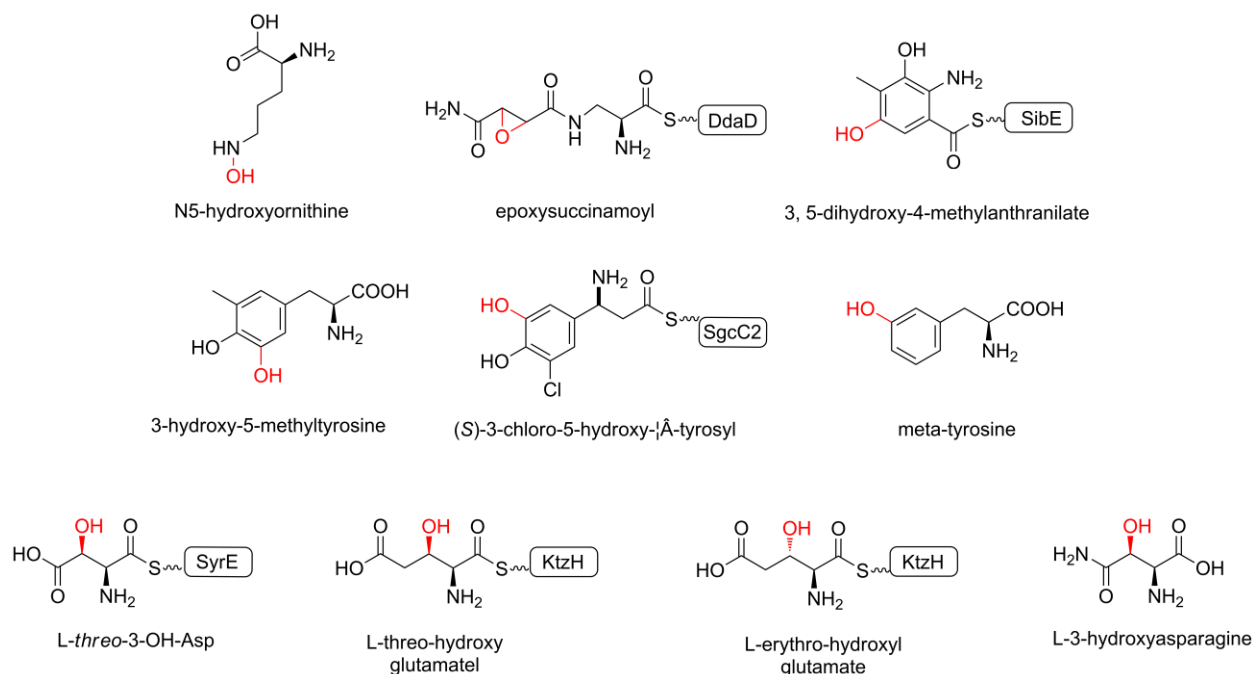
Scheme 3. Representative chemical structures generated by pre-assembly modifications in nonribosomal peptide biosynthesis.

Recently, oxygenases using different mechanisms were identified to catalyze the aromatic hydroxylation of tyrosine (Tyr) and Phe. SfmD was characterized as a heme containing peroxidase which regioselectivly hydroxylate 3-methyltyrosine to 3-hydroxy-5-methyltyrosine in the biosynthesis of saframycin A, using H_2O_2 as oxidant [61]. SgcC is a monooxygenase involved in the biosynthesis of the antitumor enediyne C-1207 and was shown to catalyze the hydroxylation of β-Tyr in a T-domain dependent fashion [62]. SgcC is able to hydroxylate (*S*)-3-chloro-β-tyrosine tethered to SgcC2 at the C-5 position to afford *(S)*-3-chloro-5-hydroxy-β-tyrosyl-Sgc2 in presence of O_2 and $FADH_2$. In additional to that, SgcC can also hydroxylate *meta*-substituted bromo-, iodo-, fluoro-, and methyl-β-tyrosyl-Sgc2. 2-hydroxyl-phenylalanine was found as one of the building blocks in the pentapeptide antibiotics pacidamycins [62]. A phenylalanine hydroxylase (PecX) was demonstrated to be an Fe(II)-dependent enzyme that

14

regioselectively synthesize meta-tyrocine from L-Phe using tetrahydrobiopterin as a cofactor [63, 64]. Comparing with another group of Phe 4-hydroxylases (Phe4Hs) which catalyzes the formation of L-Tyr, Cys187 and Thr202 in the binding pocket of PacX are replaced by Phe and Gly, respectively. A PacX C187F/T202G mutant was shown to have dramatically altered hydroxylation regiospecificity to afford >90% L-Tyr instead.

SyrP is an Fe(II) and α-ketoglutarate dependent hydroxylase that acts on an aspartic residue tethered to the eighth module of SyrE to generate L-*threo*-3-OH-Asp in the biosynthesis of syringomycin E [64, 65]. In the biosynthesis of the antifungal hexadepsipeptide kutznerides, both L-*threo*-hydroxyl and L-*erythro*-hydroxyl glutamate derivatives can be incorporated [66]. After L-glutamic acid is transferred to the T domain of KtzH, the oxygenases KtzO and KtzP can catalyze β-hydroxylation to form **15** and **16**, respectively. Unlike the above examples, AsnO was demonstrated to catalyze the hydroxylation of free L-asparagine to yield L-3-hydroxyasparagine, which is activated and incorporated by the NRPS to form the calcium dependent antibiotic (CDA) peptidyl chain [67]. Based on the AsnO crystal structure, Marahiel and coworkers engineered a D241N AsnO mutant of Asn [67].

Post-assembly modification in nonribosomal peptide biosynthesis

In the post-assembly tailoring steps, redox cascades can transform nascent scaffolds into structurally complex final products. Epoxidation of the indole ring in the synthesis of notoamides, a family of prenylated fungal indole alkaloids containing the bicyclo [2.2.2] diazaoctane core, is catalyzed by a FAD-dependent monooxygenease NotB (**Scheme 4**) [68]. Formation of the 2, 3-epoxide initiates either a Pinacol-like rearrangement to afford notoamide C or annulation of C-2 by the tryptophan amide to yield the hexacyclic notoamide D.
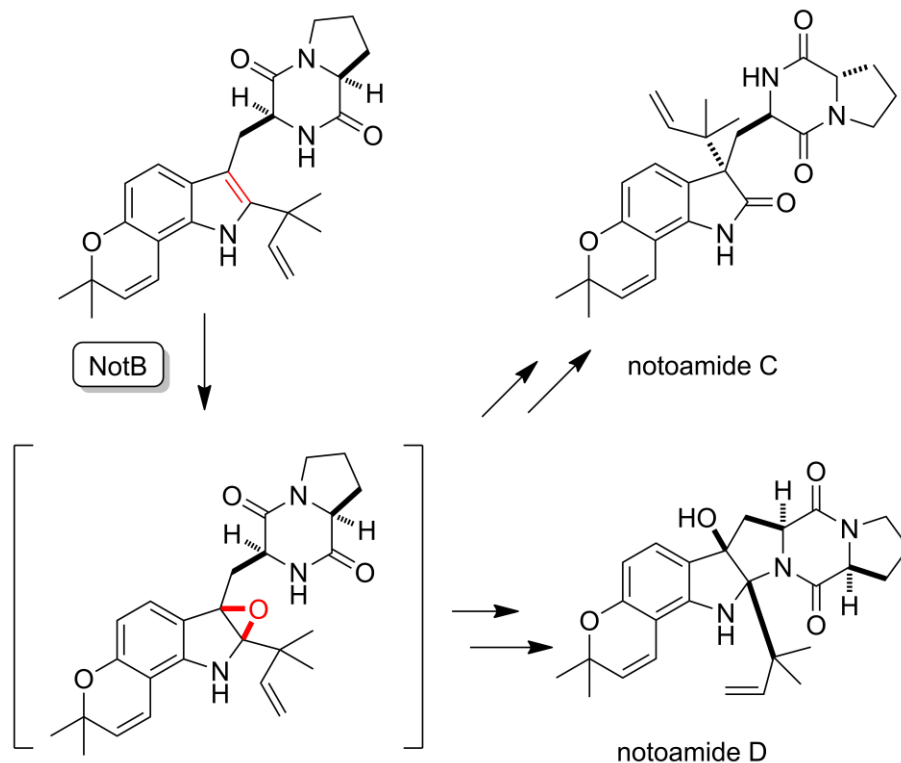
Scheme 4. The oxidative modification in notoamide D biosynthesis.

Whereas the above examples illustrate how post-NRPS oxidation can lead to dramatic change in alkaloid structure, reductive modifications have also been noted in generation of unusually complex alkaloid-like natural products, as exemplified in the formation of the pentacyclic tetrahydroisoquinoline saframycin A [69]. A single R domain in the NRPS SfmC reductively released three different peptidyl thioesters in a NADPH-dependent fashion to generate the corresponding aldehyde intermediates. Through Pictet-Spengler reactions catalyzed by the C domain, the aldehydes are cyclized with peptidyl thioesters reloaded on SfmC T-domain to afford the unique saframycin ring systems.

1.2   Engineered Simvastatin Biocatalysis

A version of this section was published as: Gao, X., Wang, P., Tang, Y. "Engineered polyketides biosynthesis and biocatalysis in *Escherichia coli.*" *Appl. Microbiol. Biotechnol.,*

Biocatalysts are gaining more and more attention in the past decades due to their intrinsic regioselectivities and stereoselectivities [70]. Moreover, the reaction conditions for biocatalysis are generally mild and more environmental friendly. To develop a powerful and industry desirable biocatalyst, directed evolution as a key technology combined with efficient screening method was frequently applied to improve enzyme solubility, thermal stability, and catalytic efficiency [71, 72]. On the other hand, crystallographic study provides detailed and fundmental insights on enzyme's three-dimensional structure and catalytic mechanism. Comparison between wild type enzyme and mutant structures might reveal how the amino acid mutations promote the improved activities, which might provide a deep insight into the poorly understood structure-activity relationship (SAR). Understanding of SAR in return will lead to rationally design the enzymatic activity and specificity, or even design enzymes de novo.

Polyketides constitute an important family of structurally diverse natural products, which include many clinically useful drugs such as the antibiotics erythromycin and tetracycline, anticancer epothilone, and the anti-hypercholesterolemia lovastatin. The pharmaceutical values of many polyketides have led to intense efforts in recent decades towards understanding and engineering the corresponding biosynthetic pathways. These concerted efforts from disciplines such as biochemistry, structural biology, genetics and metabolic engineering have enabled many examples of rational and combinatorial biosynthesis of "unnatural" natural products, as well as enzymatic tools that can be used in the biosynthesis of natural product-derived targets [73-75].

Tailoring enzymes found in natural product biosynthetic pathways catalyze a wide array of reactions, including acyltransfer [76], glycosylation [77], hydroxylation [78] and

17
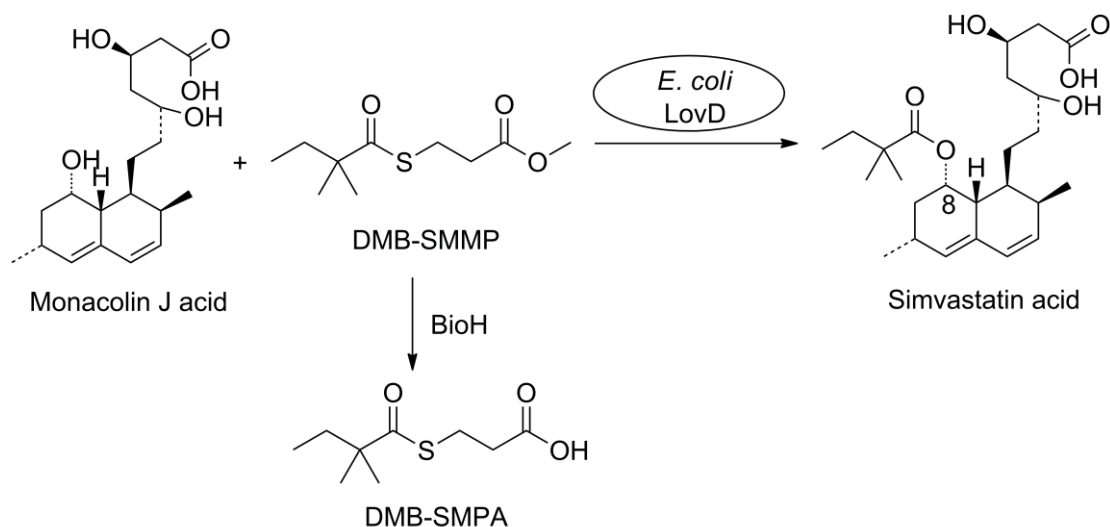
Figure 5. Synthesis Simvastatin by using *E.coli* as the whole-cell biocatalyst

halogenations [79]. A number of these enzymes decorate biologically inactive precursors into pharmaceutically active molecules via regioselective and stereoselective transformations. As a result, tailoring enzymes are attractive candidates as biocatalysts towards synthesis of semisynthetic derivatives and drug libraries [80].

E. coli has been engineered to catalyze the semisynthesis of important, polyketide-derived drug molecules. One example is the biocatalytic synthesis of simvastatin, an important cholesterol-lowing drug which is the active ingredient of the blockbuster drug Zocor[®]. Simvastatin is a statin, which is a family of compounds that that can inhibit the 3-hydroxy-3-methylglutaryl-coenzyme A (HMG-CoA) reductase, the enzyme that catalyzes the rate-limiting step in the biosynthesis of cholesterol [81]. The first statin approved by the FDA, lovastatin (Mevacor), is a fungal polyketide produced by *A. terreus* [82, 83]. Simvastatin is a semi-synthesized derivative of lovastatin and is a more potent inhibitor of HMG-CoA reductase, while exhibiting considerable less side effects in treating the hypercholesterolemia [84]. The only

difference between lovastatin and simvastatin is the substitution of α-*S*-methylbutyrate with α-dimethylbutyrate side chain at C-8 position of the decaline core (**Figure 5**). Currently, simvastatin is produced by a multistep chemical synthesis starting from lovastatin.

LovD is an acyltransferase found in the lovastatin biosynthetic pathway and is responsible for catalyzing the last step of lovastatin biosynthesis by acylating monacolin J with the α-*S*-methylbutyryl side chain synthesized by an iterative type I PKS LovF [85]. Therefore, LovD was considered as a potential enzyme that can transfer a dimethylbutyryl side chain to monacolin J and afford simvastatin via the biocatalytic approach. To confirm the activities of LovD, Xie *et al*. cloned and overexpressed LovD in *E. coli* [86]. LovD was shown to regioselectively transfer α-*S*-methylbutyryl acyl group to monacolin J to produce lovastatin via a ping-pong mechanism. The authors demonstrated that LovD had broad substrate specificities toward the acyl carrier, the acyl substrate and the decalin acyl acceptor. For example, it was able to accept membrane permeable SNAC thioesters, such as α-dimethyl-butyryl-SNAC (DMB-SNAC) as the acyl donor. Using *E. coli* as a whole-cell biocatalytic host, monacolin J was converted to simvastatin using a LovD expression strain and feeding of DMB-SNAC.

To improve the efficiency of the acyltransfer reaction and decrease the cost of the a-dimethylbutyryl acyl donor, further optimization of the substrate was performed [87]. Among the various thioesters assayed in this effort, the authors determined that α-dimethylbutyryl-*S*-methyl-3-mercaptopropionate (DMB-SMMP) is a superior substrate considering both the high initial turnover rate and the low raw material cost (**Figure 5**). When DMB-SMMP was supplemented to the *E. coli* host, nearly quantitative conversion (>99%) of monacolin J to simvastatin was achieved at 6 g/L. Subsequently, an engineered *E. coli* strain YT2, which contains a Δ*bioH*

knockout, was selected as the whole-cell host [88]. BioH was previously reported as a carboxylesterase involved in the biotin biosynthetic pathway [89]. It was found that BioH rapidly hydrolyzed DMB-SMMP into dimethylbutyryl mercaptopropionic acid (DMB-SMPA) (**Figure 5**), which significantly lowered the overall turnover rate of the transformation. By using the YT2 strain as the host, the rate of simvastatin synthesis was significantly increased.

2.   Results

2.1   Genetic Investigation of the Biosynthetic Pathway of Fungal Indole Alkaloid, Tryptoquialanine

A version of this section was published as: Gao, X., Chooi, Y., Ames, B. D., Wang, P., Walsh, C. T., Tang, Y. "Fungal quinazoline alkaloid biosynthesis: Genetic and biochemical investigation of the tryptoquialanine pathway in *Penicillium aethiopicum*." *J. Am. Chem. Soc.,* 2011, 133, 2729-2741. Copyright © 2011, American Chemical Society.

2.1.1   Introduction

Tremorgenic mycotoxins are a group of indole alkaloids that are capable of eliciting intermittent or sustained tremors in vertebrate animals by acting on the central nervous system (CNS) [90]. Grains, forages and animal feeds contaminated with the tremorgen-producing molds are one of the major sources of mycotoxin intoxications in cattle, sheep and dogs, where the clinical symptoms include diminished activity and immobility, followed by hyperexcitability, muscle tremor, ataxia, titanic seizures and convulsions [90, 91]. Based on their structural features, the tremorgenic agents can be divided into the indole-diterpenoids (e.g. penitrems and paspalitrems), the prenylated indole-diketopiperazines (e.g. fumitremorgens and verruculogens), and the quinazoline-containing indole alkaloids related to tryptoquivaline **2** (**Scheme 5**) [92, 93].

20

Tryptoquialanine **1** is highly similar to **2** and differs only in the alkyl substitution in the quinazoline ring. The mode of action of these tremorgens is not well understood, but they are thought to interfere with neurotransmitter release [94, 95]. Some of the tremorgens also exhibit
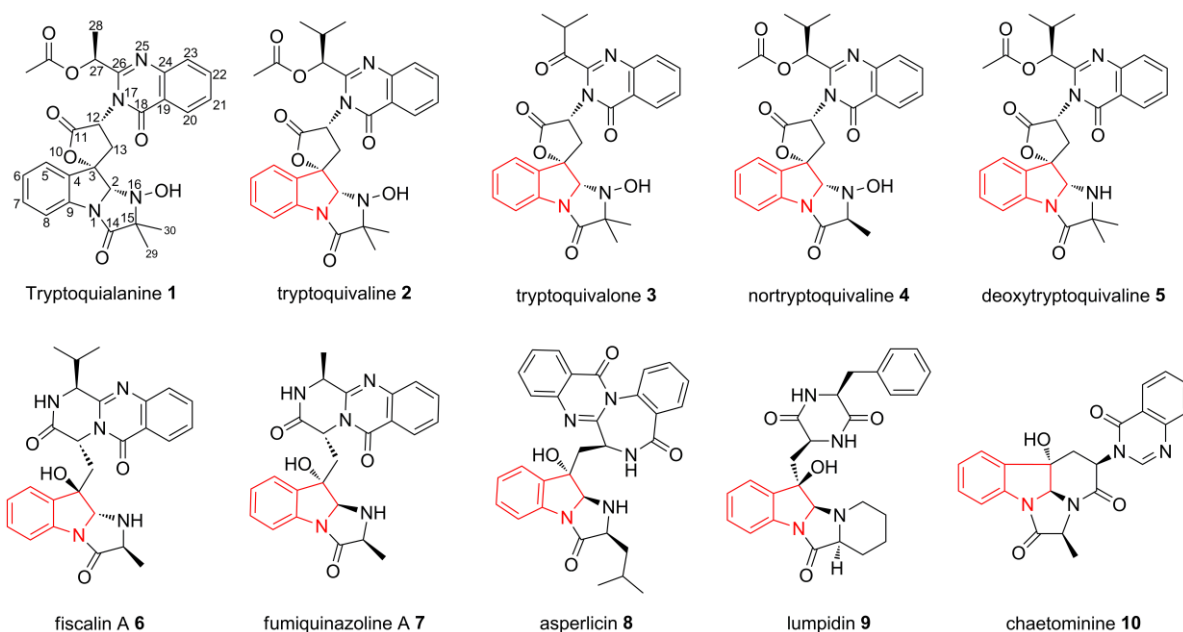
Scheme 5. Tryptoquialanine **1**, tryptoquivaline **2** and related fungal indole alkaloids

useful biological activities, for example, fumitremorgin C is a potent and specific inhibitor of breast cancer resistance protein (BCRP) [96], while the penitrems are shown to exhibit potent insecticidal activity [97-99]. Biosynthesis of the tremorgenic indole-diterpenoids and prenylated indole-diketopiperazines are currently subjects of intensive studies [100-103]. Comparatively, the biosynthesis of the tremorgenic quinazoline alkaloids related to **1** and **2** has not been elucidated.

The structurally related **1** and **2** are produced by several fungi in the *Penicillium* spp. and *Aspergillus clavatus,* respectively [104-106]. Both **1** and **2** are multicyclic compounds that exhibit structural features not observed among other indole alkaloids (**Scheme 5**). Common to both compounds is an acetylated quinazoline ring connected to a 6-5-5 imidazoindolone ring

system via a 5-membered spirolactone. The imidazolidone ring is heavily modified, containing the N16 hydroxylamine and the C15 *gem*-dimethyl group. The structural difference between **1** and **2** is thought to arise from the incorporation of alanine or valine, respectively. The structures of **1** and **2** are also related to the pyrazino[*2,1-b*]quinazoline alkaloids such as fiscalin A **6** and fumiquinazoline A **7** [107, 108]. These multicyclic scaffolds are assembled from various proteinogenic and nonproteinogenic amino acids by the actions of short nonribosomal peptide synthetase (NRPS) assembly lines [109, 110]. Common building blocks shared by many compounds in this family are an anthranilic acid and a tryptophan [38, 105].

Recently, the anthranilate-activating adenylation (A) domains of several fungal NRPS have been characterized and the modular assimilation of the amino acids to synthesize **7** in *Aspergillus fumigatus* has been partially reconstituted *in vitro* [38, 111]. The formation of the imidazoindolone moiety in **7** has been shown to involve a two-step oxidative-acylation of the indole ring by a single module NRPS and a flavin-dependent monooxygenase. A similar mechanism is likely involved in the biosynthesis of **1** and **2**, as well as other imidazoindolone-containing alkaloids, such as **9** and **10** [112, 113].

Nevertheless, it is not known whether a pyrazinoquinazoline intermediate analogous to **7** is involved in the biosynthesis of **1** and **2**. Isolation of metabolites related to **2**, such as tryptoquivalone **3**, nortryptoquivaline **4** and deoxytryptoquivaline **5** have provided hints regarding possible

Viridicatumtoxin **11**

Griseofulvin **12**: R=Cl
Dechlorogriseofulvin **13**: R=H
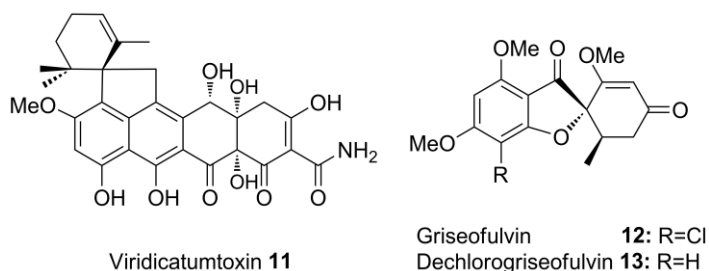
Scheme 6. Other metabolites produced by *P. aethiopicum*

22

biosynthetic intermediates and the origins of unique structural features [105, 114, 115]. For example, the isolation of **4** suggests that the *gem*-dimethyl group present in **2** may arise from the *gem*-methylation of a monomethylated intermediate. However, to obtain a comprehensive understanding of the biosynthetic mechanisms of these complex fungal alkaloids, a combination of both genetic and biochemical approaches are needed, starting from the identification of the respective biosynthetic gene clusters.

Recently, we used 454 sequencing technology to gain partial genome information of *P. aethiopicum* and determined the gene clusters involved in biosynthesis of the aromatic polyketides viridicatumtoxin **11** and griseofulvin **12** (**Scheme 6**) [116]. *P. aethiopicum* and *P. digitatum* have been reported previously to produce **1**; and a compound that has identical UV absorbance and mass to **1** were indeed detected by us in the extracts of *P. aethiopicum*. Therefore, the availability of the genome sequence data presented an excellent opportunity to study the biosynthesis of this family of fungal indole alkaloids. In this report, we present the identification and verification of the *tqa* gene cluster; functional assignment of the individual genes through genetic and biochemical approaches; and insights into the origins of the unique structural features of **1**.

2.1.2 Results and Discussions

Isolation of 1 and Verification of Structure.

Compound **1** was isolated from a four-day culture of *P. aethiopicum* grown on YMEG medium
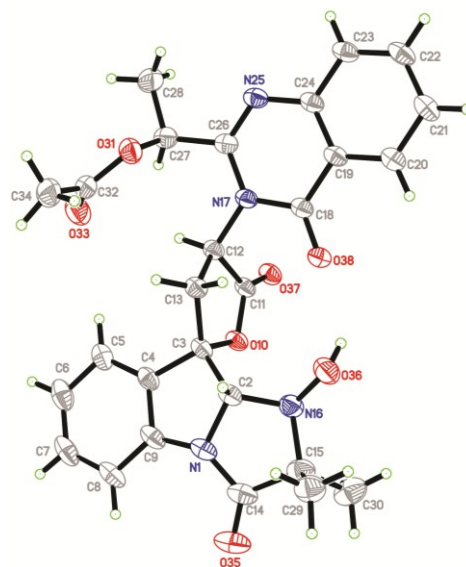
Figure 6. A perspective drawing of **1**.

23

at a final titer of 8 mg/L. Proton and carbon NMR spectra of the purified compound matched the previous published data (**Table S3, Figure S6**) [104]. To verify the three dimensional structure of **1** as that shown in **Scheme 5**, especially that of the substituents on the imidazoindolone rings, **1** was crystallized from a methylene chloride/heptane mixture and the X-ray structure was solved as shown in **Figure 7**. All of the relative configurations of **1** matched that of the solved structures of **2** and **4** [104, 105, 114] including positions C2, C3, C12 and C27.
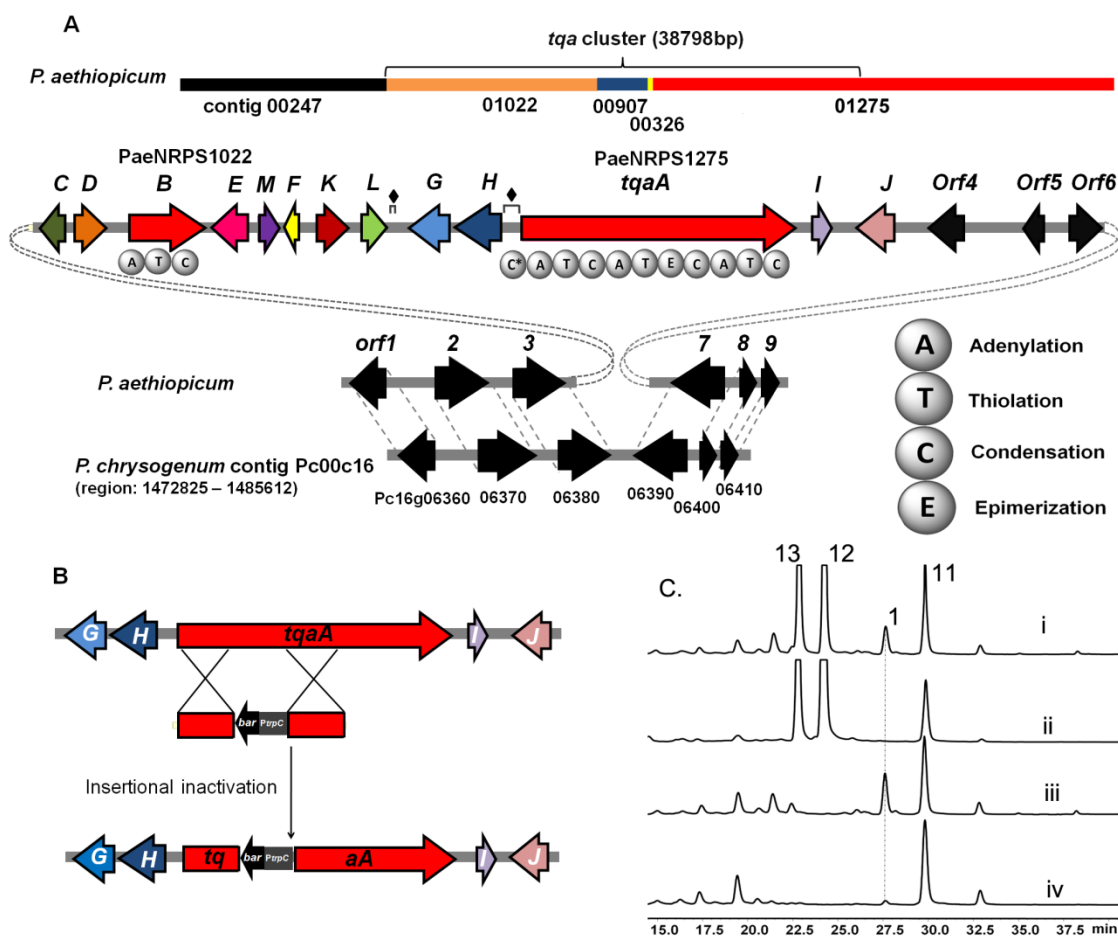
Figure 7. Organization of the *tqa* gene cluster and genetic verification of involvement in **1** biosynthesis. (A) The *tqa* gene cluster; (B) Knockout strategy used to inactivate *tqaA*; (C)HPLC (280 nm) traces of metabolic extracts from single gene deletion strains of *P. aethiopicum*. Trace i: wild type strain producing **1**, **11-13**; trace: ii, Δ*tqaA*; trace iii, Δ*gsfA*. This strain was constructed to eliminate the high titer metabolites **12** and **13**; and trace iv, Δ*gsfA*/Δ*tqaK*. ■ indicates repeating sequences.

Notably, the *syn* stereochemical configuration across C2 and C3 of the indole ring is confirmed. The *R* configuration at position C12 is consistent with incorporation of a D-tryptophan moiety that likely arises through epimerization of L-tryptophan during NRPS assembly. The crystal structure is also consistent with the absolute configurations of **1** determined by NOE and of that of **4** from X-ray crystallography [104, 114].

Identification and Verification of Gene Cluster and Analysis

Having verified the structure of **1**, we scanned the sequenced genome of *P. aethiopicum* for possible gene clusters that are responsible for biosynthesis. The NRPS (AnaPS) from *Neosartorya fischeri* NRRL 181, which synthesizes acetylaszonalenin, has been previously identified [26]. Using the adenylation (A) domains of AnaPS, which activates an anthranilate and a tryptophan, the NRPS genes in *P. aethiopicum* were identified from the local genome database by TBLASTN program (**Table S1**). By eliminating the common NRPS genes (>89% identity) in *P. aethiopicum* and *P. chrysogenum*, the number of candidate NRPS genes was narrowed down from 16 to 10 (**Table S1**). Further bioinformatic analysis of functional domains along with a specific search of common NRPS homologs present in the genomes of both *P. aethiopicum* and the **2** producer *A. clavatus* NRRL1, led to the identification of a candidate trimodule NRPS on contig 1275 (PaeNRPS1275, 67% identity to ACLA017890) and a single module NRPS on contig 1022 (PaeNRPS1022, 64% identity to ACLA017900) (Figure 2A and Table 1). Sequence analysis of PaeNRPS1275 revealed high overall sequence identity (54%), and identical domain arrangement to the recently identified AFUA6G12080 (abbreviated as Af12080), which is proposed to synthesize fumiquinazoline F **14**, an intermediate on the way to fumiquinazoline A **7** (**Table 1**, **Figure S7**) [111]. The three A domains of PaeNRPS1275 are predicted to activate

anthranillic acid, L-tryptophan, and L-alanine sequentially. The substrate specificity of the first A domain of Af12080, which activates anthranilic acid, has also been confirmed [38]. The presence of an epimerization domain (E) following the second module, which is proposed to activate L-tryptophan, is also consistent with the presence of D-tryptophan in the scaffold of **1**.

To verify the involvement of the PaeNRPS1275 NRPS in **1** biosynthesis, a double recombination cassette was constructed as shown in **Figure 7B** and transformed into protoplasts of *P. aethiopicum*. Following selection of glufosinate and PCR verification, seventeen clones were identified to contain a Δ*tqaA* knockout (**Figure S1**). None of these clones produced **1** (**Figure 7C**) while the biosynthesis of other metabolites such as **11-13** were unaffected, confirming the essential role of PaeNRPS1275 (renamed as TqaA) in **1** biosynthesis. To eliminate the production of **12** and **13**, which are present at very high levels and can complicate detection and purification of compounds related to **1**, we constructed a Δ*gsfA* mutant of *P. aethiopicum* using the zeocin selection marker. The Δ*gsfA* strain was not longer able to synthesize **12** and **13** and is used in subsequent genetic analysis of the *tqa* cluster.

To complete the *tqa* gene cluster, a combination of fosmid sequencing and primer walking was performed to link different contigs with contig1275. The putative *tqa* gene cluster is shown in **Figure 7A**. To determine the putative boundary of the gene cluster, a comparative analysis with the sequenced *P. chrysogenum* genome was performed. Interestingly, the upstream *orf1-3* and downstream *orf7-9* flanking the *tqa* cluster are highly conserved and syntenic in *P. chrysogenum* (**Table 1 and Figure 7A**). We assumed that these conserved syntenic genes do not participate in biosynthesis of **1** but are involved in *Penicillium* housekeeping roles. Although *orf4* is not syntenic, it is highly similar to an ortholog in *P. chrysogenum* (96% identity, **Table 1**).

26

The similarity of *orf5* and *orf6* to the possible orthologs in *P. chrysogenum* is significantly lower (37% and 25% identity respectively). To exclude the possible involvement of *orf4-6* in biosynthesis of **1**, single gene deletions were performed for these three genes on Δ*gsfA* background. As expected, production of **1** was unaffected in the Δ*orf4,* Δ*orf5* and Δ*orf6* mutants (**Figure S2**).

Based on the results from genetic knockouts and comparative genomic analysis, the *tqa* cluster embedded within the conserved syntenic regions is proposed to span ~32 kB and contains 13 genes (named *tqaA – tqaM*). The putative assignments of gene functions are shown in **Table 1**. The gene cluster encodes one transcriptional regulator TqaK. TqaK is a basic-region leucine zipper (bZIP) DNA-binding protein, and shared 27% protein identity with RadR, which regulates radicicol biosynthesis [117]. Deletion of *tqaK* using the *bar* selection marker did not completely abolish production of **1**, as observed for *radR*, but led to substantial attenuation of **1** titer to less than one-twentieth of the wild type strain (**Figure 7C**). This confirms the role of TqaK as a positive transcription regulator.

Functions of the Two NRPSs in *tqa* Gene Cluster

The *tqa* gene cluster contains a monomodule (A-T-C) NRPS TqaB (PaeNRPS1022), which shares high sequence similarity to Af12050 that acylates L-Ala to the oxidized indole ring of **14** to yield **7**. The *tqa* gene cluster also contains two flavin-dependent oxidoreductases TqaH and TqaG, which are homologous to Af12060 and Af12070 found in the FQA gene cluster in *A. fumigatus*, respectively (**Figure S3**). While Af12060 is responsible for oxidation of the indole ring of **14** prior to *N*-acylation, Afl12070 is likely involved in the oxidative rearrangement of **7** towards other natural fumiquinazolines, such as fumiquinazoline C and D [111]. The presence of

these enzymes, along with the similarity between TqaA and Afl2080, hints that the biosynthesis of **1** may proceed first via the pyrazinoquinazolinone intermediate **14** and then the C2-epimer of **7.** Depending on the timing of the introduction of the C15-*gem*-dimethyl group, either 15-dimethyl-2-*epi*-fumiquinazoline A **18** or 2-*epi*-fumiquinazoline A **19** may be a biosynthetic precursor of **1**.

To identify these possible intermediates in the *tqa* pathway, we constructed single-gene knockouts of these three genes based on the Δ*gsfA* strain and analyzed the subsequent metabolite profiles (**Figure 8A**). Inactivation of TqaH led to the synthesis of a single metabolite at titers of 6 mg/L. The compound has the mass (m/z = 358) and UV absorption pattern consistent with that of **14**. Purification and NMR characterization confirmed the compound is indeed **14** (**Table S4**), and points to the analogous role of TqaH in oxidizing **14** as previously demonstrated for Afl2060 (**Scheme 7**) [111]. The Δ*gsfA*/Δ*tqaB* knockout strain no longer produced either **1** or **14**, but instead afforded a more polar metabolite with mass (m/z = 374). A possible structure of this compound is **15**, which might be the 2, 3-epoxidized version of **14** (**Figure S8**). This compound is highly unstable during purification and could not be isolated for further spectroscopic analysis. Finally, to probe the pathway shown in **Scheme 7** and the role of TqaG as an enzyme that can possibly modify the fumiquinazoline-like intermediate, we analyzed the extract of Δ*gsfA*/Δ*tqaG*. This strain produced a predominant compound with mass (m/z = 459) and UV pattern suggestive of fumiquinazolines. This compound was purified and the structure was determined based on extensive NMR data to be that of **18** (**Table S5, Figure S9**). To verify the *syn* stereochemical configuration across C2 and C3 of the indole ring, as well the relative stereochemistry of other chiral carbons, the X-ray structure of **18** was determined and shown in **Figure 8B**.

28

More detailed examination of the $\Delta gsfA/\Delta tqaG$ extract revealed the presence of another quinazoline compound (RT = 19.3 min) with mass ($m/z = 445$) corresponding to that of **18** with one fewer methyl group. The most likely candidate compound is therefore **19** in which the C15 position is occupied by a single methyl group which can be introduced from the side chain of L-Ala. Although this compound was present in a significantly lower titer, it was purified and thoroughly characterized by NMR to be indeed **19**. The loss of ^{1}H signal of the *gem*-dimethyl at $\delta=0.93$ ppm and ^{13}C signal at $\delta=23.5$ ppm, and the accompanying appearance of additional C$\underline{H}$ at $\delta=3.65$ ppm and C29 methyl at $\delta=17.2$ ppm, are consistent with the structural difference between **18** and **19** (**Table S5, Figure S10**).

Figure 8. Biosynthesis of **18** as an intermediate in the *tqa* pathway. (A) HPLC analysis (280 nm) of intermediates accumulated in the knockout strains constructed starting from $\Delta gsfA$. Trace i: $\Delta gsfA$; trace ii: $\Delta gsfA/\Delta tqaB$; trace iii: $\Delta gsfA/\Delta tqaH$; and trace iv: $\Delta gsfA/\Delta tqaG$. (B) A perspective drawing of **18**.

Origin of *gem*-dimethyl Quaternary Carbon

The isolation of both **18** and as a minor component, **19**, from the $\Delta gsfA/\Delta tqaG$ strain provides clues to the timing and source of the *gem*-dimethyl incorporation. Since **18** is a relatively early intermediate in the pathway that leads to **1**, the *gem*-dimethyl can be introduced via one of the following different routes: a) activation of the nonproteinogenic amino acid 2-aminoisobutyrate

(AIB) by TqaB; b) activation of ʟ-alanine by TqaB and α-methylation while attached to the thiolation domain as an activated aminoacyl thioester; or c) direct α-methylation of **19** to yield **18**. The last alternative should be a difficult methylation reaction since generation of the nucleophilic enolate at C15 of **19** can be considerably more difficult for the amide carbonyl under biological settings.

Table 1. The *tqa* gene cluster and gene functions assignment

Gene	Size (bp/aa)	BLASTP homolog accession number	Identity/ similarity (%)	Putative function	*E*-value	Related metabolite produced after KO
tqaA	12310/4095	ACLA_017890 AFUA_6G12080	64/77 54/69	NRPS (C*ATCATECATC)	0 0	No product
tqaB	3327/1108	ACLA_017900 AFUA_6G12050	67/80 56/72	NRPS (ATC)	0	**15, 30**
tqaC	1106/363	ACLA_061530	59/72	Short-chain dehydrogenase	7e-93	**25**
tqaD	1542/513	ACLA_061540	50/64	Acetyltransferase	2e-113	**26, 27**
tqaE	1620/466	ACLA_017910 ADM34142 (*notI*) ADM34135 (*notB*)	61/73 45/63 43/65	FAD-dependent oxidoreductase	8e-173 2e-101 6e-94	**24, 28**
tqaF	717/238	AO090701000440	70/84	Haloalkanoic acid dehalogenase	4e-98	**1**
tqaG	1723/489	ACLA_017880 AFUA_6G12070	72/83 46/61	FAD-dependent oxidoreductase	0 7e-114	**18, 19**
tqaH	1545/463	ACLA_017920 AFUA_6G12060	65/82 54/70	FAD-dependent oxidoreductase	1e-177 3e-134	**14**
tqaI	810/269	ACLA_017930	54/73	Trypsin-like serine protease	2e-68	**23**
tqaJ	1852/587	ACLA_098230	55/76	MFS toxin efflux pump	7e-145	-
tqaK	1498/416	UREG_02305	33/46	bZIP DNA-binding protein	4e-54	**1**
tqaL	1116/371	NCU01071 ACLA_063370	62/76 60/76	Unknown function	8e-112	**20**
tqaM	1057/311	NCU01072 ACLA_063360	70/82 50/67	Class II aldolase	2e-124 3e-76	**20**
orf4	1302/434	Pc12g07140	96/98	Unknown function	0	**1**
orf5	819/273	PMAA_037110	49/68	RTA1-like transmembrane protein	7e-58	**1**
orf6	1387/417	NECHADRAFT_80860	57/75	Zn2Cys6 transcription factor	3e-74	**1**

To uncover the origin of the *gem*-dimethyl group, we first examined the A-domain specificity of TqaB. The uninterrupted *tqaB* was cloned by using splice-by-overlap extension PCR, expressed from *Escherichia coli* in both apo- and holo forms, and purified to single-band purity using Ni-NTA affinity chromatography (**Figure 9A**). ATP-[^{32}P]PP$_i$ exchange assay was used to monitor the activity of the A domain in the presence of different amino acids. As shown in **Figure 9B**, AIB is clearly the preferred substrate for adenylation by the A-domain of *apo* TqaB; while D-Ala, L-Ala, L-α-aminobutyric acid (AABA), and D-AABA also promoted exchange above background level (37, 27, 21, and 9% the level observed for AIB, respectively). Therefore, it is evident that compared to the functionally analogous Afl2050 which does not activate AIB and only weakly activates D-Ala [111], the A domain of TqaB has a clearly different substrate spectrum. The preference towards AIB hence strongly suggests that the *gem*-dimethyl in **1** and **18** is the result of AIB activation by TqaB.
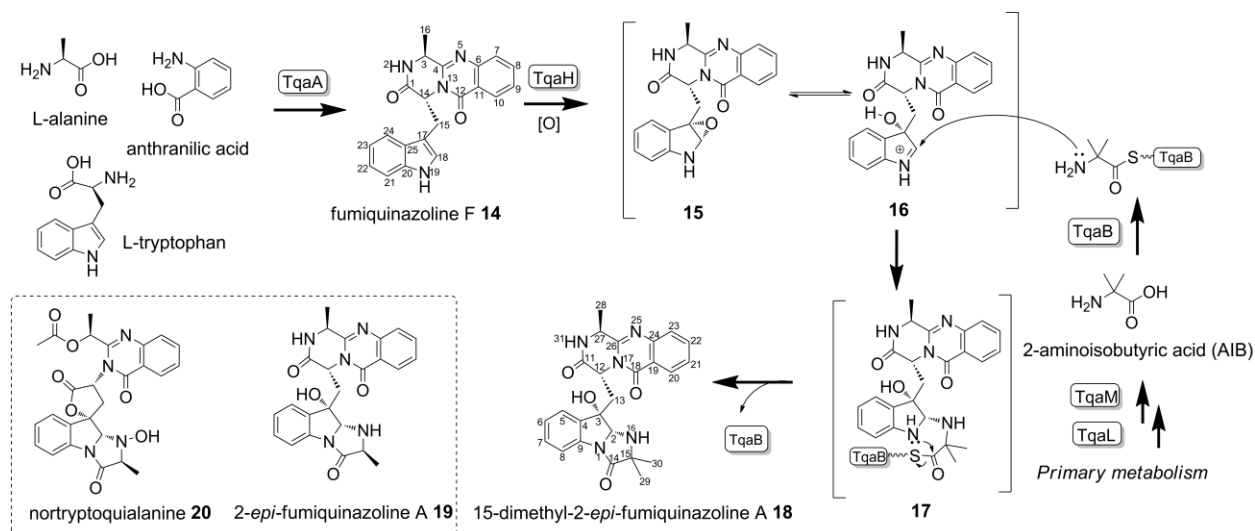
Scheme 7. Enzymes involved in the synthesis of fumiquinazoline intermediates **14** and **18**

To gain insight into the functional difference between the A domains of TqaB and Afl2050, we aligned the 10-residue substrate specificity-determining sequence (10AA code) of

the A domains, along with known L-Ala specific fungal A-domains (e.g. Af012050) and proposed AIB activating A domains of **2** (ACLA_017900) and that of peptaibol synthetases (Tex1 from *Trichoderma virens*) [118](**Figure 9C**). The amino acid sequence of TqaB was submitted to the web-based NRPSpredictor and the 10AA code was extracted as DLFMMCGCIK [119]. ACLA_017900 shares the exactly same 10AA code as TqaB, which indicates that ACLA_017900 is likely to activate AIB as well. The 10AA code of Af12050 is highly similar between TqaB and ACLA_017900, but are different at position 3, 4 and 5. On the other hand, the 10AA code of TqaB A domain bears little similarity to AIB-activating domains in Tex1. Additional details with regard to how the 10AA code residues of TqaB and Af12050 may dictate their respective substrate specificities are provided in the discussion.

Comparison of the 10 amino acids responsible for Adenylation domains substrate specificity										
Name	Pos1[a] (235)	Pos2 (236)	Pos3 (239)	Pos4 (278)	Pos5 (299)	Pos6 (301)	Pos7 (322)	Pos8 (330)	Pos9 (331)	Pos10 (551)
TqaB (U[b])	D	L	F	M	M	C	G	C	I	K
ACLA_017900 (U)	D	L	F	M	M	C	G	C	I	K
AFUA6g012050 (A)	D	L	F	V	L	A	G	C	I	K
Tex1-A1 (U)	D	L	G	Y	L	A	G	V	F	K
Tex1-A5 (U/A)	D	V	F	W	L	C	G	V	F	K
Tex1-A9 (U)	D	L	G	Y	L	A	G	-	C	K
Tex1-A12 (U)	D	L	G	Y	L	A	G	V	F	K
Tex1-A15 (U)	D	L	G	F	L	A	G	V	F	K
Tex1-A16 (U)	D	L	G	F	L	A	G	-	C	K

[a] Residue positions (Pos) are numbered according to GrsA-PheA. [b] Abbreviations: U,2-aminoisobutyric acid; A, alanine.

Figure 9. Activation of aminoisobutyric acid (AIB) by TqaB. (A) Expression and purification of TqaB from BAP1; (B)ATP-[^{32}P]PP$_i$ exchange assay using purified TqaB (100% relative activity corresponds to 55,000 cpm); (C) Alignment of the specificity-determining residues in TqaB with other related AIB (U)/L-alanine (A) activating domains.

Having established that AIB is a likely building block of **1**, we next investigated the possible *tqa* enzymes that are involved in the synthesis of AIB. As with many NRPS clusters, the enzymes that are required for the synthesis of nonproteinogenic amino acids used by NRPS are typically encoded in the respective gene clusters [22].

These dedicated enzymes are therefore expressed only during the production of the nonribosomal peptides, which minimizes the interference of the products with ribosomal translational machinery. Since no AIB biosynthetic pathway is known to date and because none of the remaining *tqa* enzymes standout as potential candidates, we decided to generate single gene knockout strains of all remaining genes starting with the Δ*gsfA* strain. All of the *bar*-selected clones were verified by PCR to confirm the deletions; and were then cultured and extracted for metabolite analysis. The results of these knockout experiments are shown in **Table 1**.

From these studies, two genes with unknown functions were identified as likely to be involved in the biosynthesis of AIB. Knocking out of either *tqaM* or *tqaL* led to the production of the same shunt product **20** (*m/z* = 504), which matches the mass of, and is confirmed by NMR to be nortryptoquialanine (or tryptoquialanine B) (**Figure 10A** trace ii and trace iv, **Table S3, Figure S11**). Therefore, it appears these mutants are blocked in the synthesis of AIB, and TqaB instead selected L-
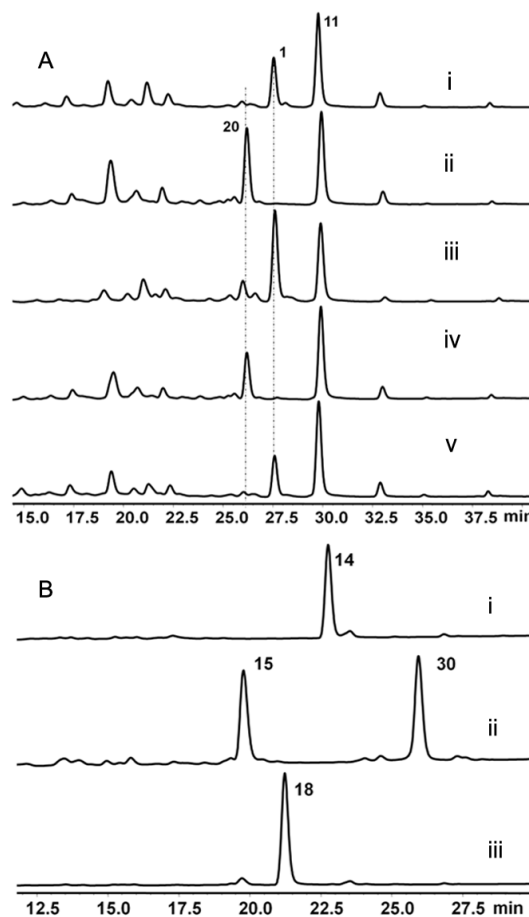
Figure 10. Identification of *tqa* genes that are likely involved in the biosynthesis of AIB. (A) Extract from the following strains that produced nortryptoquialaine **20** are shown here. Trace i: Δ*gsfA;* trace ii: Δ*gsfA*/Δ*tqaL*; trace iii: Δ*gsfA*/Δ*tqaL* supplied with 1.5 mM AIB; trace iv: Δ*gsfA*/Δ*tqaM*; trace v: Δ*gsfA*/Δ*tqaM* supplied with 1.5 mM AIB demonstrating restoration of biosynthesis of **1**. (B) In vivo feeding of 120 μM **14** to trace i: BAP1 strain (no enzyme overexpression control); trace ii: BAP1 expressing TqaH; and trace iii: BAP1 expressing TqaB and TqaH and supplemented with 1.5 mM AIB.

33

Ala leading to the synthesis of **20**. Hence, the downstream enzymes that convert **18** to **1** are nonspecific towards the C15 *gem*-dimethyl. To prove that the Δ*tqaL* and Δ*tqaM* mutants are indeed blocked in AIB synthesis, we supplemented the mutant cultures with 1.5 mM AIB. As expected, production of **1** was restored to wild type levels in both mutants, thereby establishing TqaL and TqaM are essential for the *de novo* AIB synthesis in *P. aethiopicum* (**Figure 10A** trace iii and trace v).

To further prove AIB is involved in the biosynthesis of **1**, we aimed to reconstitute the conversion of **14** to **18** using an *E. coli* strain overexpressing TqaB and TqaH, and supplemented with AIB. The *tqaH* and *tqaB* cDNA were amplified with reverse-transcription (RT)-PCR and both genes were cloned into pCDFDuet™-1 vector for expression in BAP1 [120]. After induction with IPTG and culturing overnight at 16°C, **14** was added to a final concentration of 120 μM and the culture was extracted with ethyl acetate after 2 hour. When TqaH was expressed alone, *in vivo* conversion of **14** to an oxidized product, likely **15**, was observed; along with a new compound that has mass consistent with a possible crosslinked dimer **30** (**Figure 10B, Figure S19**). Formation of the dimer was previously observed in the *in vitro* reaction containing **14** and Afl2060, and our result here further verifies the function of TqaH as analogous to Afl2060 (**Scheme 7**) [111]. When both TqaB and TqaH were overexpressed in *E. coli,* along with supplementation with 1.5 mM AIB and **14**, complete oxidation followed by near complete acylation with AIB to yield **18** was observed. Excluding AIB led to the accumulation of **19**, further confirming the origin of the *gem*-dimethyl in **1.**

Tailoring Enzymatic Reactions Leading to Synthesis of 1

Analysis of the extract from the single gene knockout strains (in the background of Δ*gsfA*) shown in **Table 1** also allowed us to assign functions to the remaining enzymes in the gene cluster, of which most are suggested to be involved in the conversion of **18** to **1** (**Scheme 8**).
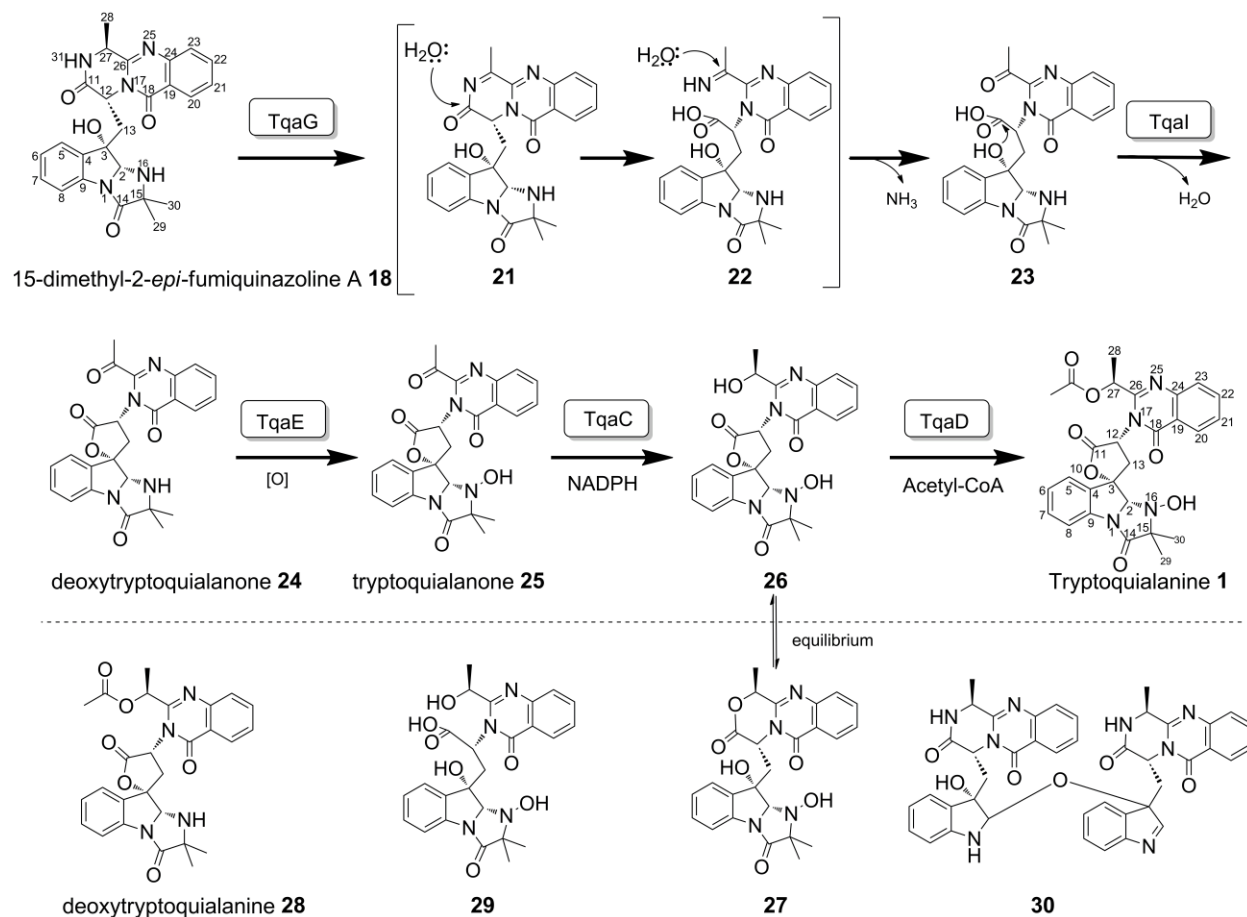
Scheme 8. Proposed enzymatic steps that convert **18** to **1**

The accumulation of **18** in the Δ*tqaG* knockout mutant suggests that TqaG is immediately involved in transforming **18** *en route* towards **1**. A BLAST search identified that TqaG has a FAD-binding site and belongs to the berberine bridge enzyme (BBE) superfamily. The BBE is proposed to initiate the oxidative cyclization of the N-methyl moiety of (S)-reticuline via the formation of a methylene iminium ion that undergoes subsequent ring closure

to form the berberine bridge carbon, C-8, of (*S*)-scoulerine [121]. Therefore, we propose that TqaG might play a possible role in the 2-electron oxidation of the pyrazinone ring of **18** to yield the α-imine intermediate **21**. Hydrolysis of **21** yields the imino acid **22**, which can be rapidly converted to the ketone **23** upon nucleophilic attack by water at C27. Alternatively, the pyrazinone ring of **18** may be hydrolyzed to yield the C27 free amine, which can then be transaminated by a pyridoxal-5'-phosphate (PLP)-dependent enzyme to afford **23**. Several lines of reasoning however, make the second pathway unlikely: 1) no enzyme bearing resemblance to a possible transaminase is observed in the gene cluster; 2) hydrolysis of the highly stable pyrazinone ring without oxidation is difficult and should result in rapid recyclization to afford the starting compound **18**; and 3) the homolog of TqaG in the pathway of **7**, Af12070, has been proposed to initiate the intramolecular cyclization of **7** towards fumiquinazoline C and D via oxidation of the same carbon of the pyrazinone ring [122].

TqaI is similar to trypsin-like serine proteases present in insects (e.g. >30% identity to the homologs in the dust mite *Dermatophagoides farinae*) [123]. A BLAST search using TqaI matched to only three fungal homologs (in *A. clavatus, A. terreus* and *Gibberella zeae*). The *tqaI* homolog in *A. clavatus* (ACLA_017930) is clustered together with the other *tqa* homologs in the genome, but absent in the gene cluster of **7** (**Figure S3**). Thus, it is likely that TqaI and ACLA 017930 play a common role in biosynthesis of **1** and **2** following formation of the pyrazinoquinazoline scaffold. From the Δ*gsfA*/Δ*tqaI* knockout strain, a compound (*m/z* = 476) that is most likely to be **23** was isolated (**Figure S12**). Upon purification of **23** by reverse-phase HPLC, the compound rapidly dehydrated to form the keto-lactone **24** (*m/z* =458). NMR characterization of **24** revealed the appearance of signals that correspond to an aliphatic ketone at

δ=195.2 ppm (**Table S6, Figure S14**). Thorough 2D NMR confirmed the structure of **24** to be that of deoxynortryptoqialanone, in which the bridging spirolactone is installed. These evidences therefore suggest that TqaI is likely an accessory enzyme in the enzymatic lactonization of **23** to produce **24**, a reaction that may also proceed spontaneously. Indeed, the Δ*gsf*/Δ*AtqaI* strain continued to produce **1** as shown in **Figure 10**; and the combined level of **23** and **1** in this strain is near the titer of **1** in the wild type strain.

TqaE belongs to class A flavoprotein monooxygenases and shares moderate similarity to the characterized TqaH (39% identity) and Afl2060 (35% identity). Recently, a pair of TqaE/TqaH homologs (NotB/NotI) were identified in the notoamide gene cluster (both shared 45% identity to TqaE), which were proposed to catalyze a 2,3-epoxidation and a *N*-hydroxylation of the

Figure 11. The remaining steps of 1 biosynthesis as elucidated from the single gene knockout studies. Trace i: Δ*gsfA;* trace ii: Δ*gsfA*/Δ*tqaC*; trace iii: Δ*gsfA*/Δ*tqaD;* trace iv: Δ*gsfA*/Δ*tqaE;* and trace v: Δ*gsfA*/Δ*tqaI;*

tryptophan-derived indole ring to form the final notoamide A [100]. Indeed, deoxynortryptoqialanone **24** was also isolated from the Δ*gsfA*/Δ*tqaE* strain together with a tryptoqialanine-like compound **28** (*m/z* = 502). Compared to that of **1,** the NMR signals of **28** are nearly identical but with the loss of the *N*-hydroxyl signal at δ = 7.95 ppm and appearance of a new NH signal at δ = 3.18 ppm (**Table S7, Figure S17**). Based on the NMR information, **28** is

assigned to be deoxytryptoquialanine as shown in **Scheme 8**. The isolation of **24** and **28** are consistent with the predicted *N*-hydroxylation function of TqaE. While the exact timing of the hydroxylamine formation is not known, isolation of the ketone **24** suggests that *N*-oxidation of **24** to **25** may take place immediately following spirolactone formation. The high titer of **28** also indicates that the remaining tailoring steps in the *tqa* pathway can function in the absence of *N*-hydroxylation.

Formation of **1** from **25** requires the stereospecific reduction and acetylation of the C27 ketone. The most likely enzyme candidate in the *tqa* gene cluster for the ketoreduction is TqaC, which is homologous to putative NADPH-dependent short chain dehydrogenases. Inactivation of TqaC should therefore lead to accumulation of **25** in the culture extract. As expected, the Δ*gsfA*/Δ*tqaC* strain produced a single shunt product with mass (*m/z* = 474) and NMR data consistent with that of tryptoquialanone **25** (**Figure 11, Table S6**). Interestingly, 27-*epi*-isomers of **2** and **4** have been isolated from *Corynascus setosus* [124, 125]. Based on the deduced biosynthetic pathway of **1** and the role of TqaC, the stereochemical difference at position C27 between **2**/**4**, and the corresponding epimers can be attributable to the different stereospecificity of the ketoreductase *tqaC* homologs in *A. clavatus* (ACLA_061530) and in *C. setosus*.

Finally, in the Δ*gsfA*/Δ*tqaD* strain in which putative acetyltransferase TqaD is inactivated, two metabolites **26** and **27** that have masses *(m/z* = 476) consistent with the TqaC-catalyzed ketoreduction of **25** were observed (**Figures S15 and S16**). **26** and **27** existed in equilibrium during extraction and purification, which prevented NMR characterization of individual compounds. However, this equilibrium is expected for a C27-reduced and unacetylated intermediate, as the interconversion between the spirolactone **26** (γ-lactone) and the

oxazinoquinazoline **27** (δ-lactone) should take place readily under aqueous conditions. To examine the acetyltransfer reaction in more detail, we overexpressed and purified the hexahistidine tagged TqaD from BL21 (DE3). When incubated with a mixture of **26** and **27** purified from Δ*gsfA*/Δ*tqaD* and acetyl-CoA, formation of **1** was readily observed (Figure S4). Similarly when incubated with **1** and assayed for the reverse hydrolysis reaction with 20 μM TqaD, we were able to detect the formation of both **26** and **27.** In both assays, we also observed the formation of a new compound **29** that has the mass (m/z = 494) corresponding to the ring opened form of **26** and **27** (**Figure S18**). When extracted under strong acid conditions (5% TFA), **29** can be nearly completely lactonized into **26** and **27.** Acetylation of **26** by TqaD to yield **1** is therefore the last step in the *tqa* pathway and is critical to prevent opening of the connecting spirolactone ring. The only remaining gene that has not been assigned a putative function is *tqaF*, which encodes an enzyme belonging to the haloacid dehalogenase superfamily. The Δ*gsfA*/Δ*tqaF* strain continued to synthesize **1** at the same level as the wild type, which indicates that this enzyme may not be essential in the proposed pathway.

P. aethiopicum is closely related to the penicillin-producing *P. chrysogenum*, whose genome has been sequenced [8], but both species produce distinct secondary metabolites [106]. As one of the demonstrations, we previously showed that comparative genomics can be a useful tool to narrow down the biosynthetic genes responsible for production of a particular metabolite by exclusion of orthologous genes [116]. The structural similarities between **1** and **2** suggest that homologous genes are likely involved in their biosynthesis. Using a similar strategy coupled with a genome-wide search of common NRPSs in *P. aethiopicum* and *A. clavatus* [34], which produces **2**, we were able to pinpoint the trimodular NRPS TqaA and single module NRPS TqaB,

and subsequently confirm their involvement in biosynthesis of **1** by targeted gene deletion. The observation of conserved syntenic regions flanking the *tqa* gene cluster when compared to the corresponding genetic locus in *P. chrysogenum,* is akin to the *vrt* and *gsf* loci in the previous study [116].

The corresponding *tqa* homologs in the *A. clavatus* genome that are predicted to be involved in the biosynthesis of **2** were identified via BLAST search (**Table 1, Figure S3**). As in *tqa* cluster, the corresponding homologs for *tqaA, tqaB, tqaE, tqaG, tqaH,* and *tqaI* are clustered in the *A. clavatus* genome. Interestingly, there are several genes in the putative *tqv* cluster for **2** that are not clustered together with the NRPS genes but fall on a separate genomic scaffold. Specifically, the homologs for the ketoreductase (*tqvC*) and acetyltransferase (*tqvD*) are adjacent to each other and fall on the genomic scaffold 1099423829796. The corresponding *A. clavatus* homologs for *tqaL* and *tqaM* are also located next to each other on the same genomic scaffold as *tqvC* and *tqvD,* but the two pairs are located 440 kbp apart. Similar fragmentation of secondary metabolic gene clusters has also been observed in the pathway for dothistromin, a mycotoxin that is structurally similar to the aflatoxin intermediate versicolorin A [126]. The presence of repeating sequences in the *tqa* gene cluster may suggest recent recombination or horizontal gene transfer events, which brought the genes in the *tqa* pathway into proximity. The clustering of *tqa* genes in the *P. aethiopicum* genome therefore presents an excellent opportunity to study the function of individual genes in the pathway.

Initial examination of the peptide linkages in **1** suggested that the amino acids may be assembled in the order of alanine or pyruvic acid, anthranilic acid and tryptophan, followed by the lactonization and release of the tripeptide from a trimodule NRPS. *N*-acylation of the indole

ring with alanine/AIB could follow thereafter. However, the identification of TqaA as a trimodular NRPS with shared domain architecture and sequence similarity to Afl2080 that synthesizes **14** (**Figure S3**), strongly indicates that **14** could be a common intermediate for both pathways. The knockout of *tqaH* confirmed that **14** is indeed the common intermediate and the formation of the spirolactone in **1** therefore requires opening of the pyrazinone ring, which partially masked the biosynthetic origin of **1**. Furthermore, identification of **14** as the authentic intermediate demonstrated that the TqaA trimodular NRPS utilizes L-Ala instead of pyruvic acid. Other homologous genes shared by the two gene clusters are the *tqaB, tqaH* and *tqaG*. From the corresponding knockout studies, the roles of TqaB and TqaH are indeed consistent with those corresponding homologs involved in the synthesis and tailoring of **7**. The isolation of **18** from Δ*gsfA*/Δ*tqaG* mutant suggests that TqaG is the immediate oxidative tailoring enzyme in the pathway. The intriguing stereochemical difference between **18** and **7** across the C2 and C3 positions of the indole ring may be attributed to the functional difference between TqaB and Afl2050. Whereas **7** contains the *anti* configuration that would be expected from epoxide opening by the free amine group of alanine, the *syn* addition in **18** points to a mechanism in which the 3-hydroxyiminium cation **16** is the true intermediate for nucleophilic attack of the TqaB-activated α-amino group on the iminium ion to yield **17** (**Scheme 7**). The nucleophilic nitrogen on the dearomatized indole ring presumably then attacks the aminoacyl-TqaB thioester to form the 6-5-5 imidazoindolone scaffold.

By a combination of genetic and biochemical means, we determined the *gem*-dimethyl moiety in **1** is incorporated via the activation of the unnatural amino acid AIB by the monomodular NRPS TqaB. Gene deletion of *tqaM* and *tqaL* abolished the production of AIB,

and TqaB instead activated L-Ala to produce **20**. We therefore propose that TqaM and TqaL are responsible for the production of AIB in *tqa* gene cluster. Although AIB is a commonly found amino acid constituent of *many fungal secondary metabolites*, the enzymatic basis for its biosynthesis is not known. BLAST search of the GenBank database using the amino acid sequences of TqaM and TqaL showed that homologs of these two enzymes can be found in other fungal genomes, and in most cases, adjacent to each other. These include ACLA_063360 and ACLA_063370 in *A. clavatus*, NCU01071 and NCU01072 in *Neurospora crassa* OR74A and SMAC_03146 and SMAC_03147 in *Sordaria macrospora*. AIB is most well-known for its abundant incorporation into a class of linear antimicrobial peptides named peptaibols (peptaibiotics), characterized prominently by high proportion of α,α-dialkylated amino acids [118, 127]. The membrane-modifying properties of peptaibols and their ability to form transmembrane voltage-dependent channels have attracted much interest [128]. Since the peptaibol synthetase Tex1 homolog has been found in the sequenced *Trichoderma reesei* genome [51, 129], we searched the JGI *T. reesei* v2.0 database for TqaM/L homologs. Indeed, e_gw1.17.140.1 and e_gw1.19.93.1 were identified as homologs for TqaM and TqaL respectively. However unlike in the other fungal genomes, the two homologs in *T. reesei* fall on different genomic scaffolds (**Figure S3**). The presence of TqaM and TqaL homologs in other fungal genomes maybe indicative of their undiscovered capability to produce AIB, and thus can be a useful tool for genome mining of the antimicrobial peptaibols and other AIB-containing secondary metabolites.

TqaM was predicted to have a conserved Class II aldolase domain with a Zn^{2+} binding site. The closest homolog of TqaM is NCU01071, which shares homology to NovR/CloR (40%

and 39% identity, respectively) from the novobiocin/clorobiocin biosynthesis pathway [130]. CloR has been verified to be a bifunctional non-heme iron oxygenase [131]. Although a conserved DUF2257 domain was found among the TqaL and similar proteins, the function of this conserved domain is not known. Raap *et al.* [132] reported that 2,2-dialkylglycine decarboxylase (DGD) is capable of converting AIB to acetone, and hence the PLP-dependent enzyme was also proposed to catalyze the reverse reaction, where AIB is synthesizes from acetone and CO_2 [129]. Another possible biosynthetic mechanism might be modification of alanine with a PLP-dependent enzyme to generate the α-carbanionic species to attack the electrophilic methyl group of *S*-adenosylmethionine. However, neither TqaM nor TqaL contain the required PLP or SAM (*S*-adenosylmethionine) binding domains. Therefore, the functions of TqaM and TqaL cannot be predicted at this point and is the subject of further investigations. It is also to be determined if TqaM/L homologs are involved in biosynthesis of other α, α-dialkylated amino acids, such as isovaline (IVA).

Homology modeling and analysis of the putative substrate binding pockets of the TqaB and Af12050 A-domains provides a means to rationalize the observed differences in substrate specificities (**Figure S5**). The key change appears to be at position 4 (Pos4) of the 10 AA code, in which the bulkier methionine in TqaB is modeled to favorably contact both the Pos2 Leu and the *pro-R* methyl of 2-AIB. In Af12050 the Pos4 residue is changed to valine. The shorter side chain length of Val compared to Met allows for an alternate conformation of the Pos2 Leu, in this conformation the side chain would make favorable contacts with L-Ala but clash with the *pro-R* methyl group of 2-AIB, therefore resulting in the preferential binding and activation of L-Ala in the biosynthesis of **7**. However, the 10AA code of TqaB-A domain shares low similarities

43

to the proposed AIB activation A domains in Tex1, which suggests that the AIB-activating A domains in TqaB/TqvB and those in other peptaibol synthetases, such as ampullosporin synthetase [133], and alamethicin synthetase [134], may have evolved separately.

The conversion of the tricyclic fumiquinazoline F **14** scaffold to the bicyclic framework in the tremorgens **1** and **2** with installation of the γ-spirolactone is an intriguing set of chemical transformations. The scaffold of **14** is first elaborated to an epimer of **7** in an annulation of the indole ring derived from tryptophan. The annulation can use L-alanyl thioester linked to the pantetheinyl arm of the NRPS protein TqaB to produce **19**, but TqaB prefers the unusual AIB yielding **18**. This imidazolindolone-containing intermediate is then subjected to a series of steps which take apart the pyrazinone ring of the tricyclic quinazoline framework. It appears that the process starts with oxidation of the secondary amine to the imine and that the C27-N10 bond is then fragmented in an unusual manner to yield formally the imine and acid components. The imine can hydrolyze to the ketone, observed as an intermediate **24**. *N*-hydroxylation requires opening of the pyrazinone ring since no *N*16-hydroxylated pyrozinoquinazoline intermediate was obtained. Reduction of the ketone **25** to alcohol **26** then sets up the possibility of an equilibrium between the dihydroxy acid, the spiro-γ-lactone, and the δ-lactone, all of which are detectable in specific knockout mutants. Regioselective acetylation of the C27-OH by the acetyltransferase TqaD fixes the final product **1** with the γ-lactone ring. Whether the ring-opening of the oxidized pyrazine ring in **21** from TqaG action is by net hydrolysis or involves intramolecular capture of **21** by the OH at C3 of the imidazoindolone moiety to yield **24** directly is not yet known but would provide a driving force for fragmentation of **21**. Note that the nucleophilic-OH for spirolactone formation in **25** was introduced by oxidation/annulation of the indole side chain that

happened at the stage of annulation of **14**. Conversion of **18** to **26** with a dramatically rearranged molecular architecture occurs by cryptic redox processes: regiospecific oxidation of the secondary amine in the quinazoline framework of **18** and then rereduction of the carbonyl after imine hydrolysis and spirolactone formation.

Given the remarkable morphing of the fumiquinazoline F scaffold **14** to the rearranged framework of the tryptoquialanine scaffold **1** with the above annulation and also γ-spirolactone formation, the pathway is remarkably short and efficient. Four redox enzymes are called into play: three of them (TqaH,G,E, acting in that order) contain FAD, the other (TqaC) utilizes NADPH in a conventional ketone to alcohol reduction of **25** to **26**. The flavoenzymes were proposed to carry out epoxidation of the indole side chain in 14 (TqaH), oxidation of the secondary amine linkage to cyclic imine in the pyrazinone ring of the pyrazinoquinazoline(TqaG) as the initiating step in fragmentation, and N-hydroxylation of the imidazoindolone (TqaE: 24 to 25), respectively. These transformations underscore the versatility of the FAD coenzyme for a wide chemical range of redox transformations by these biosynthetic enzymes. The detailed mechanisms of these novel enzymatic reactions are currently under investigation.

2.1.3 Conclusions

Tremorgenic mycotoxins are a group of indole alkaloids which include the quinazoline-containing tryptoquivaline **2** that are capable of eliciting intermittent or sustained tremors in vertebrate animals. The biosynthesis of this group of bioactive compounds, which are characterized by an acetylated quinazoline ring connected to a 6-5-5 imidazoindolone ring system via a 5-membered spirolactone, has remained uncharacterized. Here, we report the identification of a gene cluster (*tqa*) from *P. aethiopicum* that is involved in the biosynthesis of

tryptoquialanine **1**, which is structurally similar to **2**. The pathway has been confirmed to go through an intermediate common to the fumiquinazoline pathway, fumiquinazoline F, which originates from a fungal trimodular nonribosomal peptide synthetase (NRPS). By systematically inactivating every biosynthetic gene in the cluster, followed by isolation and characterization of the intermediates, we were able to establish the biosynthetic sequence of the pathway. An unusual oxidative opening of the pyrazinone ring by an FAD-dependent berberine bridge enzyme-like oxidoreductase has been proposed based on genetic knockout studies. Notably, a 2-aminoisobutyric acid (AIB)-utilizing NRPS module has been identified and reconstituted *in vitro*, along with two putative enzymes of unknown functions that are involved in the synthesis of the unnatural amino acid by genetic analysis. This work provides new genetic and biochemical insights into the biosynthesis of this group of fungal alkaloids, including the tremorgens related to **2**.

2.1.4   Materials and Methods

Materials. *P. aethiopicum*, IBT 5753, was obtained from the IBT culture collection (Kgs. Lyngby, Denmark). All other chemicals and solvents were purchased from either Sigma-Aldrich or Fisher Scientific unless otherwise noted.

Spectroscopic Analysis. The NMR identification of compounds was performed on a Bruker ARX500 at the University of California Los Angeles Department of Chemistry and Biochemistry NMR facility. LC/MS spectra were obtained on a Shimadzu 2010 EV liquid chromatography mass spectrometer using positive and negative electrospray ionization and a Phenomenex Luna 5 μm, 2.0 mm × 100 mm C18 reverse-phase column. Samples were separated on a linear gradient

of 5 to 95% CH$_3$CN in water (0.1% formic acid) for 30 min at a flow rate of 0.1 mL/min followed by isocratic 95% CH$_3$CN in water (0.1% formic acid) for another 15 min.

Bioinformatic Analysis. The 454-generated partial genomic sequencing data of *P. aethiopicum* is the same version as previously published [116], and was formatted into a local database for BLAST searches. Gene predictions were performed using the FGENESH online server (Softberry) and manually checked by comparing with homologous gene/proteins in the GenBank database. Functional domains in the translated protein sequences were predicted using Conserved Domain Search (NCBI). The amino acid sequences of TqaB and other NRPS adenylation domains were submitted to NRPSpredictor for automated extraction of specificity-defining residues as the 10AA code [119]. The sequence of the *tqa* gene cluster has been submitted to GenBank with the accession number HQ591508.

X-ray Crystallographic Analysis. 1 was crystallized from a mixture of methylene chloride/heptanes and **18** was obtained from the acetone/methanol mixture. X-ray diffraction was performed by the University of California Department of Chemistry and Biochemistry crystallography facility. Additional details can be found in the supplementary methods. The crystal structures of **1** and **18** are deposited at the Cambridge Crystallographic Data Centre and allocated the deposition numbers CCDC 800378 and 800379, respectively.

Construction and Screening of Fosmid Library. Fosmid library of *P. aethiopicum* was constructed using the CopyControl™ Fosmid Library Production Kit (Epicentre Biotechnologies, Madison, WI) following the manufacturer's instruction. Screening by direct colony PCR was carried out using GoTaq polymerase (Promega, Madison, WI). Initial screening was performed using pools of ~200 colonies and narrowed down to single colonies. Positive clones identified

were sent for fosmid end-sequencing and primer walking. Adjacent contigs were identified from the *P. aethiopicum* genomic database using local BLAST search of the partial fosmid sequences and assembled by pairwise alignment.

Fungal Transformation and Gene Disruption in *P. aethiopicum*. Polyethylene glycol-mediated transformation of *P. aethiopicum* was performed essentially as described previously [116]. The homologous regions flanking the resistant marker were increased to ~2 kb and other steps for construction of fusion PCR knockout cassettes containing the *bar* or *zeocin* gene were performed as described elsewhere [135]. Fusion PCR products were sequenced before using for transformation. ~7 μg DNA was gel-purified for each transformation. The *bar* gene with the *trpC* promoter was amplified from the plasmid pBARKS1[136, 137], which was obtained from the Fungal Genetics Stock Center (FGSC). The zeocin gene with the *gpda* promoter was amplified from the plasmid pAN8-1 (a gift from P. J. Punt, Institute Biology Leiden) [138]. Glufosinate used for the selection of *bar* transformants was prepared as described previously [116]. Miniprep genomic DNA from *P. aethiopicum* transformants was used for PCR screening of gene deletants and was prepared as described elsewhere for *A. nidulans* [139]. Primers used for amplification of fusion PCR products and screening of transformants, are listed in **Table S2**. Approximately 50 glufosinate-/zeocin- resistant transformants were picked and screened with PCR using a bar/zeocin gene primer and primers outside of the deletion cassette.

Chemical Analysis and Compound Isolation. For small-scale analysis, the *P. aethiopicum* wild-type and transformants were grown in stationary YMEG liquid culture (4 g/L yeast extract, 10 g/L malt extract, 5 g/L glucose and 16 g/L agar) for 4 days at 25°C. The cultures were extracted with equal volume of ethyl acetate and evaporated to dryness. The dried extracts were

48

dissolved in methanol for LC-MS analysis. For large-scale analysis, the ethyl acetate (EA) extract from a two liters stationary liquid culture of each mutant was evaporated to dryness and partition between EA/H_2O twice. After evaporation of the organic phase, the crude extracts were separated by silica chromatography. The purity of each compound was checked by LC-MS and the structure was confirmed by NMR.

Expression and Purification of Recombinant Enzymes. TqaB and TqaD cDNA were cloned into a pET28 vector with an *N*-terminal hexahistidine tag and expressed in *E. coli* BL21(DE3) strain. The transformant was cultured in 500 mL LB medium containing 35 mg/L kanamycin at 37°C to optical density (OD600) value of 0.4~0.6. Protein expression was induced with 0.1 mM IPTG and the subsequent expression was performed at 16°C overnight. Cells were collected by centrifugation (2000 g, 4°C, 15 min), resuspended in 30 ml Buffer A (50 mM Tris–HCl, pH 8.0, 2 mM DTT, 2 mM EDTA), and lysed by sonication. Cell debris and insoluble proteins were removed by centrifugation (20,000 g, 4°C, 1 hr). To the cleared cell lysate, excess amount (0.5 ml) of Ni-NTA resin (QIAGEN, Valencia, CA) was added to each sample. The TqaB protein was then purified using a step gradient of Buffer A with increasing concentration of imidazole (10 and 20 mM) and was eluted with 5 mL Buffer A containing 250 mM imidazole. Protein purity was qualitatively assessed by SDS-PAGE and concentration was quantitatively determined by the Bradford protein assay using bovine serum albumin as the standard.

ATP-[^{32}P]PP$_i$ exchange assay for TqaB. Reactions (100 μL) contained 2 mM ATP, 2 mM MgCl$_2$, 3 mM Na$_4$[^{32}P]PPi (0.2 μCi), 2 μM TqaB, and 2 mM amino acid substrate in buffer (50 mM Tris-HCl (pH7.5), 100 mM NaCl, 5mM TCEP, and 5% glycerol). Enzyme was added last to initiate the reactions and, following a 60 min incubation at 25 °C, the reactions were stopped by

adding 400 µL of a quench solution (1.6% (w/v) activated charcoal, 100 mM sodium pyrophosphate, and 3.5% perchloric acid in water). The charcoal was collected by centrifugation, washed twice with quench solution minus charcoal, and the absorbed radioactivity detected by liquid scintillation counting.

***E. coli* Mediated Biotransformation of 14 to 18.** TqaH cDNA was amplified by RT-PCR using M-MLV reverse transcriptase (Promega) and AccuPrime™ Pfx DNA polymerase (Invitrogen, Carlsbad, CA); and was cloned into pCDFDuet™-1 (EMD Chemicals, Gibbstown, NJ) and transformed into *E. coli* BAP1 strain [120]. The transformant with the TqaB and TqaH dual-expression plasmid was grown in LB medium at 37 °C to an OD600 of 0.4–0.6, at which time the cultures were cooled to 16 °C, and then induced with 0.1 mM IPTG at 250 rpm and grown at 16°C for overnight. To increase cell density, the *E. coli* cells were concentrated 10-fold before addition of substrates. A 10 mL aliquot of each culture was collected by centrifugation (4 °C, 2000 g, 10 min). The cell pellet was gently resuspended in 1 mL of medium supernatant, followed by addition of 6 µL **14** (20 mM stock) to a final concentration of 120 µM. The small cultures were then shaken at 300 rpm at 25 °C for 2 hrs. For product detection, 100 µL of cell culture was collected and extracted with 500 µL of ethyl acetate. The organic phase was separated, evaporated to dryness, redissolved in methanol, and then subjected to LC-MS analysis.

2.2 Cyclization of Fungal Nonribosomal Peptides by a Terminal C_T Domain

A version of this section was published as: Gao, X., Haynes, S. W., Ames, B. D., Wang, P., Vien, L. P., Walsh, C. T., Tang, Y. "Cyclization of fungal nonribosomal peptides by a terminal condensation-Like domain." *Nat. Chem. Biol.,* 2012, 8, 823-830. Copyright © 2012, Rights Managed by Nature Publishing Group.

2.2.1 Introduction
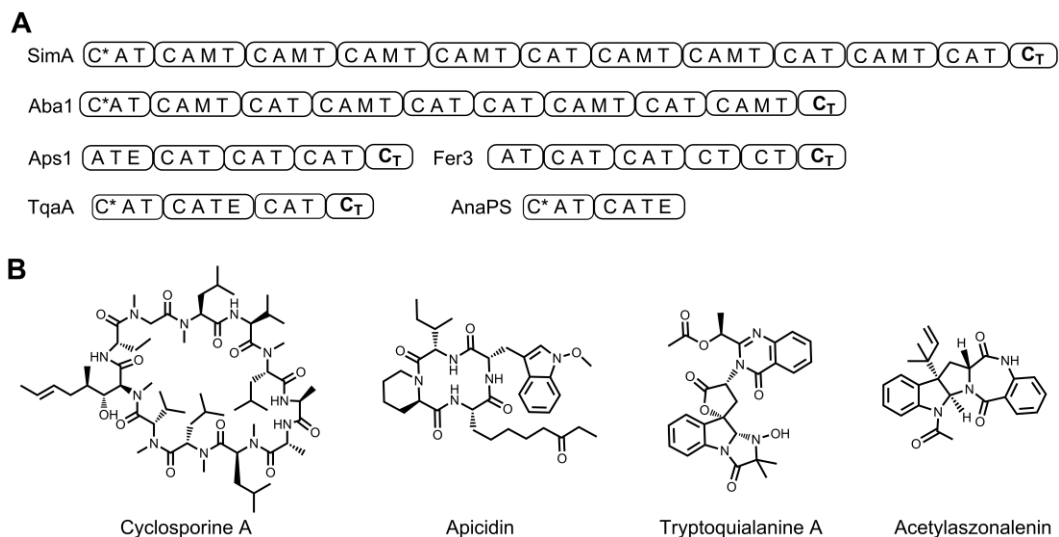
Figure 12. Fungal nonribosomal peptide synthetases (NRPSs) that are terminated with a C-like domain. (A) Representative NRPSs: cyclosporine NRPS SimA, aureobasidin A NRPS Aba1, apicidin NRPS Aps1, ferrichrome A NRPS Fer3, tryptoquialanine NRPS TqaA, acetylaszonalenin NRPS AnaPS. Domain abbreviations: C: condensation domain; A: adenylation domain; T: thiolation domain; E: epimerization domain; C*: truncated and presumably inactive C domain; C_T: Terminal C-like domain. (B) Representatives structures of final products synthesized by the clusters in part A.

In macrocyclic fungal NRPs such as cyclosporine A, aureobasidin A, apicidin and ferrichrome A, each corresponding NRPS terminates with a condensation-like domain (C_T) (**Figure 12**) [25, 140-142]. Similarly, in smaller polycyclic NRP alkaloids such as tryptoquialanine, the NRP core is assembled by a NRPS (TqaA) that terminates with a C_T domain [143](**Figure 12**).

51
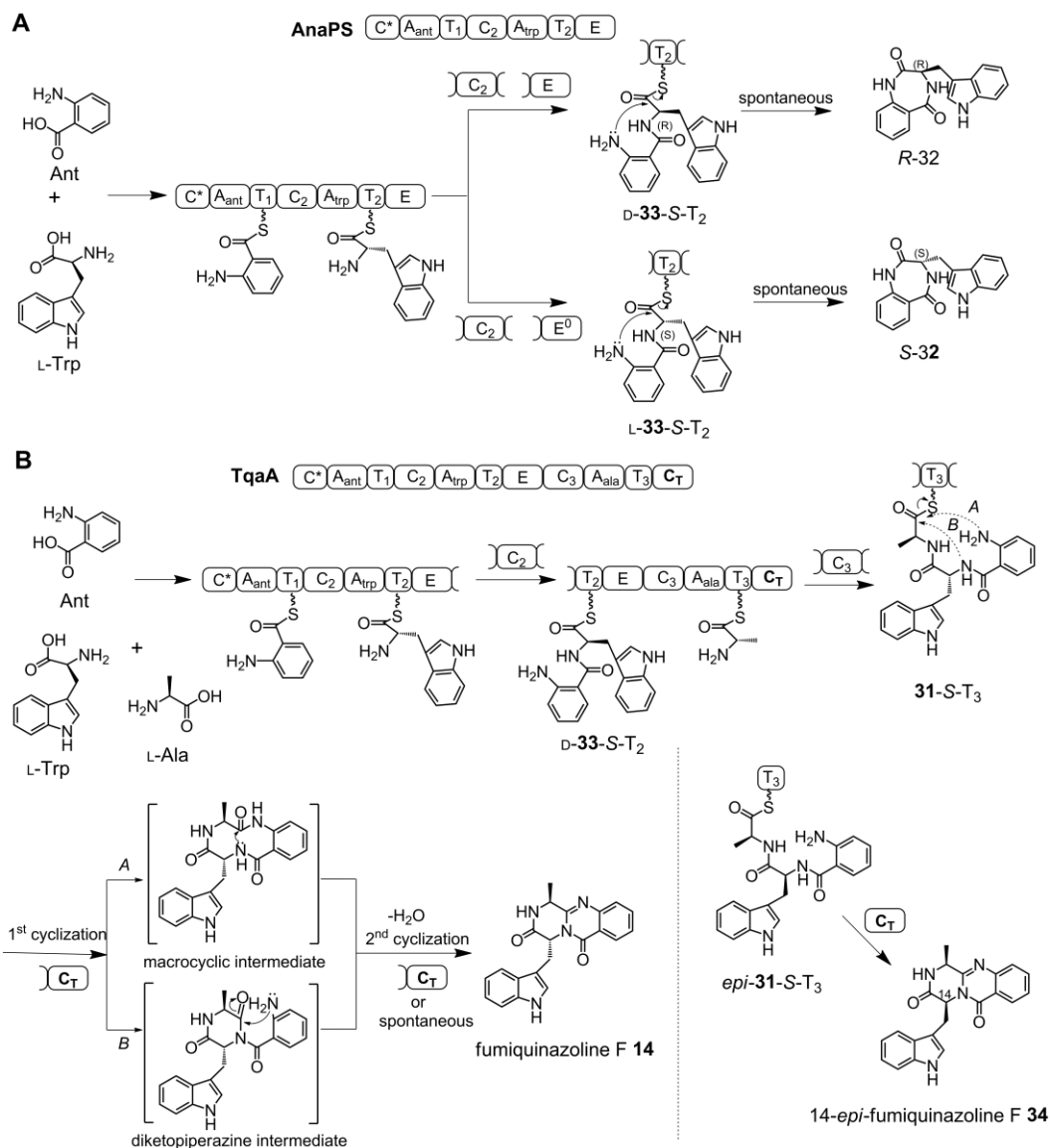
Figure 13. Proposed AnaPS (A) and TqaA (B) cyclization mechanisms. (Domain abbreviations: C: condensation domain; A: adenylation domain; T: thiolation domain; E: epimerization domain; C*: truncated and presumably inactive C domain; C_T: Terminal C-like domain).

In the NRPS paradigm, C domains are canonically categorized to catalyze the formation of a peptide bond between the growing peptidyl-S-T_n and the activated aminoacyl-S-T_{n+1} using an active site histidine as the general base [144, 145]. The active site serves as a scaffold to bring together the donor and acceptor aminoacyl thiolation domains [146]. Therefore, it is surprising that the C_T domain (a subset of C domains) can be recruited by fungal NRPS to perform the

equivalent reaction of a TE domain, which relies on a serine residue for nucleophilic catalysis [49]. Understanding the role of the C_T domain in fungal NRPSs will have broad implications, as many of the cryptic NRPSs discovered from fungal genome sequences terminate with a C_T domain [50, 147].

To investigate the role of the C_T domain and the cyclization of fungal NRPs, we focused on the trimodular TqaA, which has been shown genetically to be involved in the biosynthesis of the tricyclic peptidyl alkaloid fumiquinazoline F **14** (**Figure 13**) in *Penicillium aethiopicum* [143]. TqaA is functionally equivalent to Afl2080 from the fumiquinazoline A biosynthetic pathway in *Aspergillus fumigatus* [38]. The three modules (C^*AT_1-CAT_2E-CAT_3-C_T) of TqaA are proposed to activate anthranilate (Ant), L-tryptophan and L-alanine to yield the tripeptide Ant-D-Trp-L-Ala **31** attached to the last T domain (**31**-S-T_3) (**Figure 13**). Formation of the pyrazinoquinazoline ring and release of **14** require two cyclization steps, hypothesized to be catalyzed by the C_T domain (**Figure 13**). Interestingly, the domain organization and substrate selectivity of the first two modules of TqaA are identical to those of AnaPS (C^*AT_1-CAT_2E), which is involved in the synthesis of the 3-indolylmethyl-3,4-dihydrobenzo-1,4-diazepine-2,5-dione precursor *R*-**32** of acetylaszonalenin in *Neosartorya fischeri* [26]. Intriguingly, AnaPS terminates with an epimerization (E) domain rather than a C_T domain, although it is known that E and C domains share sequence and structural similarities [145].

Here we present the complete *in vitro* reconstitution of the trimodular TqaA expressed from *Saccharomyces cerevisiae*, allowing biochemical investigation of the cyclization steps. The functions of AnaPS are studied in parallel to compare and contrast the cyclization steps

53

between the two NRPSs. We show that the C_T domain of TqaA is indispensable for the cyclization of **14** and requires protein-protein interactions with the T_3 domain for catalysis.

2.2.2 Results and Discussions

***In vitro* Reconstitution of TqaA and AnaPS Activities**

To pinpoint the domains that are involved in the cyclization of linear peptides, we first attempted to reconstitute the activities of the NRPSs *in vitro*. This is particularly challenging for the megasynthetases such as TqaA (450 kDa) as no NRPS with more than two modules has been successfully expressed in an active form to date. Towards this end, *Saccharomyces cerevisiae* 2-*μ*m episomal vectors encoding *tqaA* (12 kb) and *anaPS* (7 kb) under the control of the ADH$_2$ promoter were constructed using a modified yeast-based homologous recombination method [148] (**Figure S20**). Both *N*-FLAG-tagged enzymes were expressed using the engineered *S. cerevisiae* BJ5464-NpgA, which contained multiple vacuolar protease knockouts and a chromosomally integrated copy of the *Aspergillus nidulans npgA* gene that encoded a phosphopantetheinyl transferase [149]. The co-expression of NpgA in BJ5464 is required for the conversion of *apo* thiolation domains into the phophopantetheinylated *holo* forms [150]. This strain had been successfully used as a host for heterologous reconstitution of large fungal polyketide synthases [149]. To verify functional expression of the NRPSs, small molecule metabolites in the three-day growth media were extracted with ethyl acetate and analyzed using LC-MS. Compared to the untransformed yeast strain, BJ5464-NpgA expressing TqaA and AnaPS produced **14** and *R*-**32** at ~2 mg/L, respectively (**Figure 14A**). Following this confirmation of NRPS activities, full length TqaA and AnaPS (264 kDa) were purified to near

homogeneity by using anti-FLAG antibody affinity chromatography followed by gel filtration to

final yields of ~ 1 mg/L (**Figure S21** and **S22**).
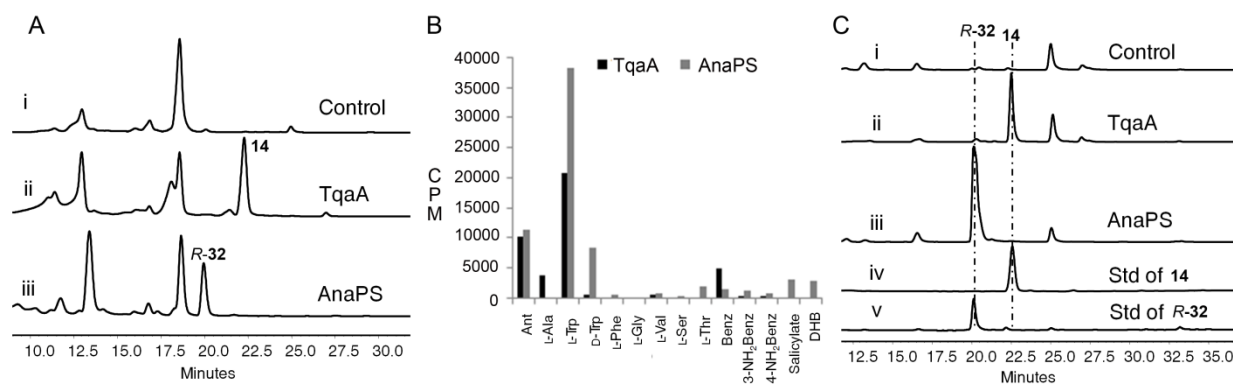
Figure 14. Characterization of TqaA and AnaPS. (A) *in vivo* reconstitution of TqaA and AnaPS. Shown are HPLC analysis (λ = 272 nm) of metabolites extracted from 3-day cultures of i) untransformed BJ5464-NpgA; ii) BJ5464-NpgA expressing TqaA; iii) BJ5464-NpgA expressing AnaPS. (B) amino acid-dependent ATP-PP$_i$ exchange assays with AnaPS and TqaA. Abbreviations: benzoate (Benz); 1, 2-dihydroxybenzoate (DHB). (C) *in vitro* reconstitution of TqaA and AnaPS. Shown are HPLC analysis (λ = 272 nm) of compounds from extraction of reaction mixtures containing 2 mM of the amino acid building blocks and i) no enzyme; ii) 10 µM TqaA; iii) 10 µM AnaPS; iv) standard of **14**; and v) standard of *R*-**32**.

After the affinity purification step, activities of the A domains of NRPSs were assessed

using amino acid-dependent [γ-^{32}P]-ATP-PP$_i$ exchange assay (**Supplementary Methods** and

Figure 14B) [151]. TqaA specifically activated the three building blocks Ant, L-Trp and L-Ala

over the respective D-amino acids or aryl acids. Similarly, AnaPS A domains activated Ant and

L-Trp. Both NRPSs also activated aryl substrates such as benzoate (by TqaA), salicylate (by

AnaPS) and 1, 2-dihydroxylbenzoate (DHB, by AnaPS), pointing to the relaxed substrate

specificities of the Ant-specific A domains. The preference displayed by both NRPSs toward L-

Trp over D-Trp strongly points to an active epimerization (E) domain in the second module of

both NRPSs.
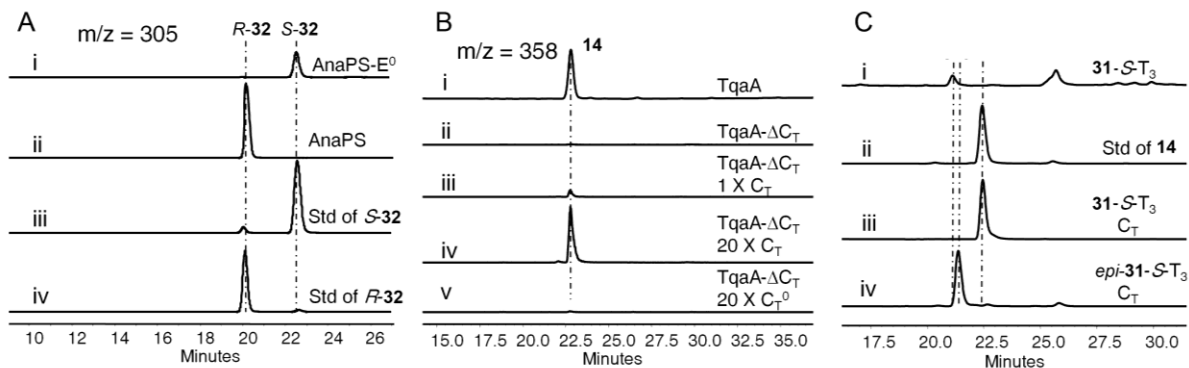
Figure 15. Probing the cyclization steps of TqaA and AnaPS. (A) chiral HPLC followed by selected mass ion monitoring (*m/z* 305) of *R*-**32** or *S*-**32** synthesized from AnaPS (trace ii) and AnaPS-E⁰ (trace i), in which the E domain has been inactivated. Standards of *R*-**32** (trace iv) and *S*-**32** (trace iii) are shown for comparison. AnaPS only produced *R*-**32**, while the AnaPS-E⁰ produced *S*-**32**. The intensities of peaks are not drawn to scale. In trace iii, the peak of *S*-**32** is magnified. (B) confirmation of TqaA C_T domain functions. All assays contained 10 μM of the NRPS (TqaA or TqaA-ΔC_T) and 2 mM of each amino acid. The traces shown are selected mass ion monitoring (*m/z* 358) of compounds extracted from reactions containing i) wild type TqaA; ii) 10 μM TqaA-ΔC_T; iii) 10 μM TqaA-ΔC_T with 10 μM of standalone C_T added *in trans*; iv) 10 μM TqaA-ΔC_T with 200 μM of standalone C_T added *in trans*; and v) 10 μM TqaA-ΔC_T with 200 μM of standalone C_T^0 (C_T domain with H¹⁷³A mutation) added *in trans*. (C) Confirmation of C_T cyclization activity using preloaded peptidyl thiolation domain. **31**-*S*-T_3 or *epi*-**31**-*S*-T_3 (200 μM) was used as the substrate to assay the production of **14** or **34**, respectively. Hydrolysis of **31**-*S*-T_3 yields **31** free acid. Traces shown are HPLC analyses (λ = 272 nm) of i) **31**-*S*-T_3 without C_T domain; ii) standard of **14**; iii) **31**-*S*-T_3 with 20 μM standalone C_T; iv) *epi*-**31**-*S*-T_3 with 20 μM standalone C_T.

Complete reconstitution of TqaA and AnaPS activities was then performed by adding the amino acid building blocks (2 mM each) and ATP (6 mM) to each of the purified enzymes (10 μM). The reaction mixtures were incubated at room temperature for 12 hours, extracted with ethyl acetate and analyzed by LC-MS (**Figure 14C**). **14** (obtained from the $\Delta tqaH$ strain of *P. aethiopicum* [143]) and chemically synthesized *R*-**32** were used as authentic standards (**Supplementary Methods**). In the TqaA assay, **14** was synthesized as the single peptidyl product; no dipeptidyl or other tripeptidyl products could be detected through LC and selective ion monitoring by MS. A HPLC-based time course study revealed that the reaction proceeds with initial product formation rate of 0.40 ± 0.014 min⁻¹ (**Figure S23a**). Similarly in the AnaPS assay, the only product detected was the 6, 7-cyclic dipeptide *R*-**32** with an initial product formation rate of 1.47 ± 0.047 min⁻¹ (**Figure S23c**).

Reconstitution of the synthesis of **14** and *R*-**32** using purified TqaA and AnaPS, respectively, showed that each NRPS was sufficient to construct and cyclize the linear peptide into the multicyclic products. The proposed cyclization mechanisms are shown in **Figure 11**. In the simpler case of AnaPS, attack of the Ant free amine on the thioester carbonyl of Ant-D-Trp-*S*-T_2 (D-**33**-*S*-T_2) yields *R*-**32**. However in TqaA, this reaction is apparently suppressed in favor of chain elongation by the third module to yield **31**-*S*-T_3. Two cyclization routes to **14** are possible starting from **31**-*S*-T_3 (**Figure 11B**). In route A, attack of the Ant amine on the thioester carbonyl yields a 10-membered macrocyclic intermediate, which can undergo a bridging amide bond cyclization and dehydration to yield **14**; in route B, the D-Trp amide serves as the initial nucleophile in attacking the thioester **31**-*S*-T_3 to yield the diketopiperazine (DKP) intermediate. Subsequent intramolecular attack by the Ant amine and dehydration completes the pyrazinoquinazoline scaffold.

Formation of *R*-32 Can Be Spontaneous

To investigate the cyclization step of AnaPS, we chemically synthesized Ant-D-Trp-*S*-*N*-acetylcysteamine (D-**33**-SNAC). Under enzymatic assay conditions, D-**33**-SNAC was cyclized to yield *R*-**32** with an initial product formation rate of 0.1 min^{-1} (**Figure S24a**). The significant rate of spontaneous cyclization suggested that product release from AnaPS could be uncatalyzed. The spontaneous cyclization rate, however, was at least fourteen-fold slower than that of D-**33**-*S*-T_2, hinting that domains in AnaPS may promote a conformation of the dipeptide that was more favorable for formation of the seven-membered ring. Adding AnaPS to D-**33**-SNAC, however, did not increase the cyclization rate, suggesting that the proposed templating effect may only take place with an enzyme-tethered dipeptide (**Figure S24b**).

57
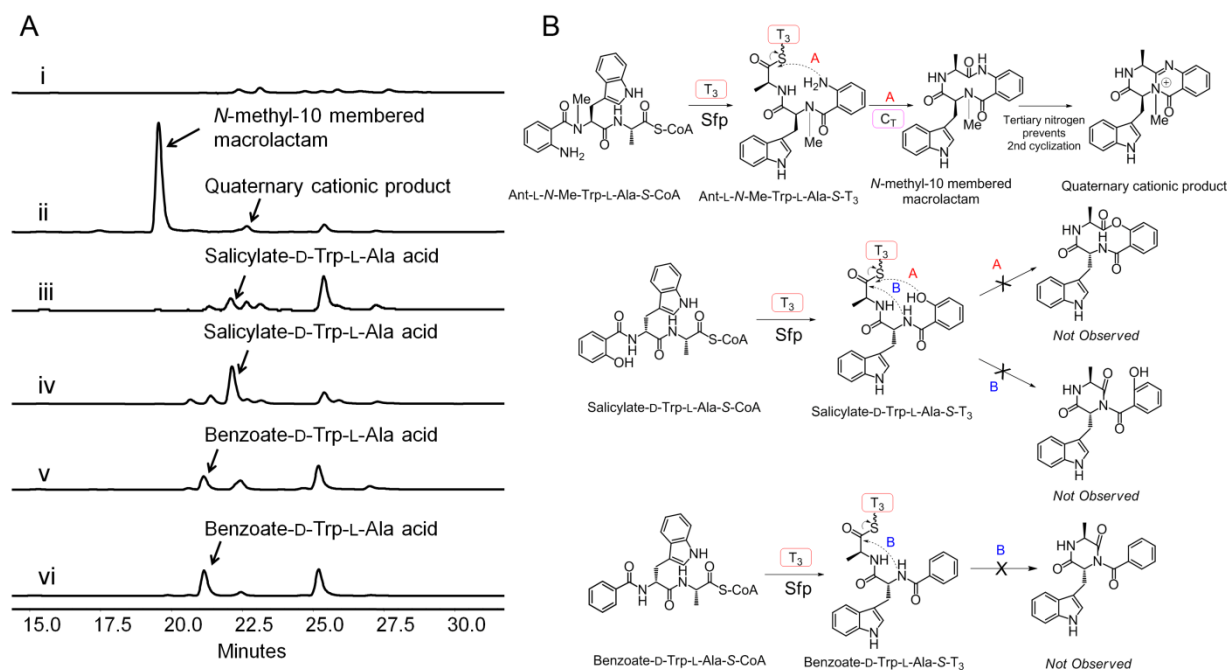
Figure 16. Probing the cyclization mechanism of TqaA C_T domain. (A) Analogues of the natural tripeptide **31** were used to probe the cyclization route of C_T domain. Ant-L-N-Me-Trp-L-Ala-S-T_3, salicylate-D-Trp-L-Ala-S-T_3 and benzoate-D-Trp-L-Ala-S-T_3 (200 μM) were used as substrates to react with C_T domain (20 μM). Traces shown are HPLC analyses (λ = 272) of i) Ant-L-N-Me-Trp-L-Ala-S-T_3 without C_T domain; ii) Ant-L-N-Me-Trp-L-Ala-S-T_3 with C_T domain. Formation of a new compound consistent with the mass of a 10-membered macrolactam was observed; iii) salicylate-D-Trp-L-Ala-S-T_3 without C_T domain; iv) salicylate-D-Trp-L-Ala-S-T_3 with C_T domain; v) benzoate-D-Trp-L-Ala-S-T_3 without C_T domain; vi) benzoate-D-Trp-L-Ala-S-T_3 with C_T domain. (B) structures discussed in text and proposed TqaA C_T domain cyclization mechanism.

The enantiomer Ant-L-Trp-SNAC (L-**33**-SNAC) was also prepared synthetically, and an identical rate of cyclization to S-**32** was observed (**Supplementary Methods** and **Fiture S24c**). The availability of both R-**32** and S-**32** as standards allowed us to examine the enantiopurity of the product released by AnaPS using chiral HPLC (**Figure 15A**). Under assay conditions, only enantiomerically pure R-**32** was detected in the organic extract, suggesting AnaPS exerts strict product stereochemical control. To probe the importance of the epimerization step catalyzed by the E domain on R-**32** formation, we generated a point mutant of AnaPS (AnaPS-E^0) in which the second His in the $H^{2059}HxxxDx$ catalytic motif was mutated to Ala (**Figure S22**) [152]. As expected, AnaPS-E^0 no longer produces R-**32** when Ant and L-Trp are added. Instead, we

observed the appearance of *S*-**32** at only ~1% of the level of *R*-**32** produced by the wild type AnaPS (**Figure 15A**, note the sizes of *R*-**32** and *S*-**32** peaks in traces i and ii are not drawn to scale). Based on the very low level of *S*-**32** production by AnaPS-E^0, two possible control mechanisms can be proposed. In the first model, AnaPS could suppress the rapid cyclization of L-**33**-*S*-T_2 until epimerization to D-**33**-*S*-T_2 takes place. The chiral environment in the E domain active site may position the Ant amine away from the thioester carbonyl. In this model, epimerization of the Trp residue took place on the dipeptidyl-*S*-T_2, which was proposed to be favored over the epimerization of the free aminoacyl thioester due to stabilization of the carbanion intermediate through acylation of the nitrogen in the peptide bond [153]. In the second model, formation of L-**33**-*S*-T_2 by the C_2 domain was strongly disfavored in the absence of epimerization, therefore requiring the E domain to function on L-Trp-*S*-T_2 prior to condensation with Ant-*S*-T_1. Intriguingly, D-Trp could be activated and transferred to the T_2 domain of AnaPS (**Figure S25**). However, AnaPS did not produce any product in the presence of Ant and D-Trp, suggesting that L-Trp was required for condensation to take place and thus providing evidence against the second model. Further attempts to distinguish between the two models through excision of the E domain from AnaPS were not successful because no soluble protein was recovered.

Given that the D-**33**-*S*-T_2 intermediate can rapidly cyclize spontaneously, it is intriguing how formation of *R*-**32** can be effectively suppressed by TqaA following formation of this dipeptidyl intermediate, especially given that the first two modules of TqaA and AnaPS appear to be functionally equivalent. Efforts to express modules 1 and 2 of TqaA as a standalone enzyme, or replace them with AnaPS resulted in insoluble proteins, despite the high sequence

59

similarity. An initial hypothesis was that the subsequent condensation with L-Ala catalyzed by C_3 of TqaA was sufficiently rapid to sequester the reactive dipeptide. However, mutation of the C_3 active site general base His[2658] to Ala not only abolished the synthesis of **14**, but also failed to yield any trace of *R*-**2**. Similarly, performing the TqaA *in vitro* assay with only Ant and L-Trp also did not produce *R*-**2** (**Figure S26**). Hence, it was reasonable to propose that the C_3 active site, in particular through the donor peptidyl entrance site, can structurally protect the D-**4** moiety and can prevent cyclization and offloading of the intermediate as *R*-**2**.

Formation of Fumiquinazoline F requires the C_T domain

Having probed the formation of *R*-**2** by AnaPS, we next examined the cyclization of **3**-*S*-T_3 into **14** by TqaA. This was a particularly intriguing transformation as a simple linear tripeptide was morphed into the 6, 6, 6 pyrazinoquinazoline tricyclic ring system. First, to understand the stability of the tripeptidyl thioester, we synthesized **31**-SNAC starting from the free acid **31** (**Supplementary Methods**). In contrast to **33**-SNAC, **31**-SNAC was remarkably stable under assay conditions, and levels of the hydrolysis product **3** only reached a noticeable level after approximately 8 hours. No cyclized products such as **14** can be observed after more than 24 hours of incubation (**Figure S24d**). Therefore, whereas formation of *R*-**32** can be spontaneous, formation of **14** definitely required enzyme catalysis. The C_T domain in TqaA was the most likely domain that could catalyze this reaction. Phylogenetic analysis showed that fungal C_T domains group separately from the canonical chain-extending C domains (**Figure S27 and S28**). Based on bioinformatic analysis, the percentage of fungal NRPSs incorporating a C_T domain varied from 60 to 90%, depending on the genome analyzed (**Supplementary Table 9)**.

Multiple sequence alignment of the active sites of different sequenced C_T domains revealed variations in the first His residue of the highly conserved **HHxxxDxxS** motif of C domains ($\underline{\textbf{S}}^{3765}$**H**SQW**D**GV**S**^{3773} in TqaA) [144] (**Figure S29**). The His proposed to be acting as the general base is conserved in all C_T domains. To probe the role of the C_T domain in pyrazinoquinazoline formation, we constructed the following mutants of TqaA: $S^{3765}A$ and $S^{3765}H$ to probe the role of the nonconserved serine, and $H^{3766}A$ to probe the role of the putative active site base. Among them, TqaA $H^{3766}A$ was expressed and purified from BJ5464-NpgA for *in vitro* assay (**Figure S22**). Whereas the S^{3765} mutants maintained production of **14** at levels comparable to that of the wild type in *S. cerevisiae*, the $H^{3766}A$ mutant failed to yield any detectable peptidyl product both in *S. cerevisiae* and *in vitro* using the purified enzyme (**Figure S30**). To exclude the possibility that an inactivated C_T domain may structurally prevent the cyclization of **31**-*S*-T$_3$, a truncated version of TqaA without the C_T domain (TqaA-ΔC_T, 400 kDa) was constructed, expressed and assayed. In parallel, the standalone C_T domain was expressed and purified from *Escherichia coli* (**Figure S22**). When assayed *in vitro*, whereas TqaA-ΔC_T failed to produce any detectable amount of **14**, complementation of TqaA-ΔC_T with equimolar amounts of C_T *in trans* restored the synthesis of **14,** albeit to a lesser extent compared to full length TqaA (**Figure 15B**). The lower level of **14** may be a result of inefficient *in trans* protein-protein interactions. Indeed, addition of twenty-fold molar excess of C_T to the reaction mixture dramatically increased the production of **14**, resulting in levels comparable to that observed when using intact TqaA. In contrast, *in trans* addition of a C_T^0 mutant in which the catalytic His was mutated to Ala failed to restore the synthesis of **14**. With all of the inactivated mutants described above we did not detect any hydrolyzed **31** free acid in the reaction mixture, suggesting that the

61

NRPS-tethered tripeptide was remarkably resistant to hydrolysis compared to **31**-SNAC (**Figure S24d**). Collectively, these results unequivocally proved the essential role of the terminal C_T domain in formation of the final product **14**.

C_T Cyclization Activity Requires T Domain

Having confirmed the catalytic nature of the C_T domain, we next assayed the standalone C_T in the cyclization reaction. Cyclizing TE domains from bacterial NRPSs can cyclize small molecule thioester mimics of the linear precursors [39, 41]. However, when **31**-SNAC or **31**-*S*-CoA was used as a substrate for TqaA C_T, no cyclized **14** was detected after prolonged incubation (**Supplementary Methods** and **Figure S31**). To prepare the authentic **31**-*S*-T_3 substrate for C_T, the *apo* form of standalone T_3 was expressed from *E. coli* (**Figure S22)** and modified *in vitro* in the presence of **31**-*S*-CoA and the broadly specific phosphopantetheinyl transferase Sfp [154]. MALDI-TOF mass analysis confirmed the near quantitative conversion from the *apo* form (molecular weight 11197) to **31**-*S*-T_3 (molecular weight 11913) (**Figure S32**). In the absence of the C_T domain, slow hydrolysis of this thioester to **31** was observed. When the wild type C_T was combined with **31**-*S*-T_3, a new peak was observed during HPLC that corresponds to **14,** and no hydrolysis was detected (**Figure 15C**), confirming the importance of the T domain in facilitating C_T functions. Adding the mutant C_T^0 did not yield any cyclized product, but partially suppressed the hydrolysis of **31**-*S*-T_3 possibly due to the templating effect of the C_T active site. The kinetic parameters of the cyclizing reaction were measured as shown in **Figure S33.** The k_{cat} of the reaction was determined to be 0.82 ± 0.053 min^{-1} and K_m of C_T for **3**-*S*-T_3 was 1.1 ± 0.12 mM. **14** can also be formed in a one-pot reaction containing *apo*-T_3, **31**-*S*-CoA, Sfp and C_T, in which the phosphopantetheinyl modification took place *in situ*. When Ant-

L-Trp-L-Ala-S-CoA (*epi*-**31**-S-CoA) was used to prepare *epi*-**31**-S-T$_3$, the C$_T$ domain also efficiently cyclized the diastereomeric peptide into the corresponding 14-*epi*-fumiquinazoline F **34**, indicating relaxed substrate specificity of the cyclization reaction towards the tripeptide (**Figure 13** and **Figure 15C**).

To examine the T domain substrate specificity of C$_T$, we also prepared noncognate T domains **31**-S-T$_{ApdA}$ and **31**-S-T$_{LovF}$, which correspond to the ACP domains of the fungal PKS-NPRS ApdA and PKS LovF, respectively [85]. Both T$_{ApdA}$ and T$_{LovF}$ share low sequence identity/similarity (~20%/60%) with TqaA T$_3$. T$_{ApdA}$ naturally interacts with a downstream C domain in ApdA and the tethered polyketide intermediate is condensed with a tyrosine residue, while T$_{lovF}$ does not interact with any C-like domain. TqaA C$_T$ was able to recognize **31**-S-T$_{ApdA}$ to produce **14**, albeit with <10% efficiency compared to **31**-S-T$_3$ (**Figure S34**) and was accompanied by hydrolysis of **31**-S-T$_{ApdA}$. Only background hydrolysis of **31**-S-T$_{ApdA}$ was observed when T$_{LovF}$ was used as the peptidyl carrier. Therefore, the protein-protein interactions between C$_T$ and the T domains are selective, and may rely on T domain sequence elements that are unique for recognition by C domains. Noncognate T domains such as T$_{LovF}$ that do not interact with downstream C-like domains likely do not possess these elements and are not substrates of the C$_T$ domain.

Cyclization Mechanism of C$_T$

In an attempt to differentiate between the two alternate mechanisms of cyclization to the observed 6, 6, 6 tricyclic compound **14** (**Figure 13B**), we examined the propensity of analogues of the natural Ant-Trp-Ala system to undergo C$_T$ mediated cyclization (**Figure 15D** and **Figure S35**). Using the standalone T$_3$ domain of TqaA and Sfp as loading catalyst we generated three

tripeptidyl-T_3 analogues specifically designed to disfavor either the macrocyclization (pathway A in **Figure 13B**) or DKP formation (pathway B in **Figure 13B**). From Ant-L-N-Me-Trp-L-Ala-S-T_3, a released product with mass $[M+H]^+$ of 391 (exact mass = 391.1765) consistent with the proposed N-methyl-10 membered macrolactam (**Figure 16B** and **Supplementary Fig. 17**) was observed. Further cyclization of this intermediate was strongly disfavored (both sterically and by the formation of a quaternary cationic product) by the tertiary nitrogen which would be required to act as nucleophile. Nonetheless, a mass consistent with this cationic product resulting from the transannular attack of the tertiary amine actually was observed as a very minor product. The formation of this product despite the steric and electronic barriers to its formation (as stated above) lends support to the geometry of this proposed macrocycle favoring the required orientation for transannular cyclization. It is worth noting that the same mass would also be consistent with the formation of a cation N-Me-Trp derived DKP, but such a product would likely be very unstable and be subject to rapid hydrolysis to form Ant-N-Me-Trp-Ala free acid, which was not detected. Conversely salicylate-D-Trp-L-Ala-S-T_3 or benzoate-D-Trp-L-Ala-S-T_3, which would be expected to disfavor/prevent macrocyclization (but not interfere with DKP formation), gave no products corresponding to the mass of the dehydrated tripeptide (corresponding to either DKP or macrocycle isomers) (**Figure 16B**). Although the mechanism of C_T mediated cyclization remains difficult to elucidate due to the reactivity of proposed intermediates, taken together these data support a route where macrolactamization precedes DKP formation in the formation of these C_T mediated 6, 6, 6 tricyclic scaffolds.

In this work, we have reconstituted the activities of two fungal NRPSs, the trimodular TqaA and the bimodular AnaPS that make tricyclic and bicyclic heterocyclic scaffolds,

64

respectively. The choice of the two NRPSs was based on the observation that while the first two modules of TqaA are identical to AnaPS, addition of the third module dramatically altered the structures of the resulting polycyclic NRP alkaloid products. The difference in the cyclization pattern is likely due to the precise control of the reactive dipeptide D-**33** intermediate, and the use of a C-like cyclization domain that regioselectively directs the intramolecular attack of the Ant amine nucleophile. While bimodular NRPSs from both bacteria and fungi have been reconstituted previously [155-157], the sheer size of NRPSs such as TqaA complicated soluble expression and functional reconstitution. The use of *S. cerevisiae* BJ5464-NpgA allowed high-yield heterologous expression and full manifestation of TqaA activities. The successful use of this strain in affording intact NRPSs should open the door to the functional analysis of the expanding collection of fungal NRPSs available through different sequencing efforts.

The use of the cyclic aromatic β-amino acid anthranilate as the chain initiation building block by AnaPS and TaqA is key for the generation of the 6, 7 fused bicyclic *R*-**32** (AnaPS) and the fused 6, 6, 6 tricyclic **14** (TqaA) scaffolds. The planar 1, 3 disposition of the amino group and the carbonyl in the anthranilyl unit efficiently sets up the D-**33**-*S*-T_2 intermediate for cyclization on the second module of AnaPS. This facile cyclization, demonstrable in the nonenzymatic thioester model, sets the unusual 6, 7 bicyclic ring system that is subsequently elaborated by prenylation and acetylation in the acetylaszonalenin pathway [26]. Notably this same 6, 7-scaffold would be an unwanted derailment product in the trimodular TqaA. Such premature release is seemingly suppressed by the action of the C-domain that results in the generation of a **31**-*S*-T_3 species. Chain release at this point mediated by C_T instead creates the tricyclic scaffold of **14**. The NH_2 of the anthranilyl unit is again acting as an internal nucleophile on a thioester but

this does not happen appreciably nonenzymatically. Clearly, the use of anthranilate as a NRPS starter unit enables construction of distinct patterns of nitrogen-containing bicyclic and tricyclic scaffolds at the core of related families of fungal peptidyl alkaloids. In the case of the fumiquinazolines, these can be further elaborated into highly constrained architectures with up to seven fused rings in the mature scaffold [122]. Thus, understanding how anthranilate gets activated, elongated, and cyclized on fungal NRPS assembly lines offers prospects for diversification of peptidyl alkaloid scaffolds.

However, based on recent genome sequencing efforts, a majority (60-90%) of the NRPSs found in filamentous fungi terminate with a C_T domain (**Table S9**). Previous sequencing of individual gene clusters has also revealed that C_T domains are associated with macrocyclic and/or polycyclic NRPs. In this work, using the pyrazinoquinazoline **14** as an example, we demonstrated that the C_T domain in the TqaA NRPS was responsible for the cyclization of the highly stable tripeptide **31-*S*-T$_3$**. The cyclization function of C_T is likely to be general for other fungal NRPSs such as those shown in **Figure 12**, including cyclosporine synthetase SimA [25]. Macrocyclization C_T domains may also play a role in the biosynthesis of cyclohexadepsipeptide destruxins in the insect pathogenic fungi *Metarhizium robertsii* [158] and in NRPSs from filamentous cyanobacteria, as illustrated in the apratoxin biosynthetic gene cluster from *Lyngbya bouilloni* [159]. C_T domains have also been proposed to be involved in the cyclization via ester bond formation during the biosynthesis of bacterial polyketide-peptidyl hybrid FK520 and rapamycin [160, 161]. However, those C_T domains fall into a different group by phylogenetic analysis (**Figure S28**).

The use of a C_T domain to heterocyclize or macrocyclize peptidyl products is a different strategy from the canonical TE domains found in bacterial NRPSs. The C_T domains and TE domains have completely different folds and utilize different mechanisms to catalyze the same reaction. In the TE domain where a catalytic triad is present, the peptidyl intermediate is first transferred to the active site serine, followed by intramolecular nucleophilic attack on the oxyester to complete cyclization [49]. In contrast, as seen for canonical C domains that catalyze peptide bond formation, no covalent adduct with the enzyme is expected to be formed during C_T catalyzed cyclization [40]. Instead, the catalytic histidine deprotonates the amine nucleophile, thus facilitating its attack on the thioester carbonyl to cyclize and release the product in one step. Another significant mechanistic difference is the requirement for a specific T domain partner in the C_T domain catalyzed cyclization reaction. We showed here that only peptide substrates tethered to the natural T domain partner of the C_T domain are efficiently recognized, while peptidyl-SNAC or peptidyl-CoA substrates are not recognized. This contrasts with most TE domains where small-molecule mimics of the peptidyl-thioester can be used to probe the cyclization functions [41, 162]. One exception is the fengycin TE that involves direct interactions with the upstream T domains to mediate the cyclization [163]. The interactions between T and C_T domains likely evolved from the interactions between upstream T and downstream C domains during peptide elongation, which facilitate the unidirectional delivery of the peptidyl substrate in a productive conformation to the donor site. This specific interaction may be advantageous to prevent spontaneous cyclization of the peptide from backbone amides or side chain amines. C_T domains also appear to advantageously eliminate the competing hydrolysis reaction, which is observed *in vitro* with bacterial TE domains where hydrolysis of the oxyester

intermediate can take place in competition with the cyclization reaction [49]. This is especially prevalent under *in vitro* conditions where a standalone TE is used. In contrast, the templating nature of the C_T highly disfavors hydrolysis as illustrated in the reconstitution assay using **31**-*S*-T_3.

2.2.3 Conclusions

Cyclization of linear peptidyl precursors produced by nonribosomal peptide synthetases (NRPSs) is an important step in the biosynthesis of bioactive cyclic peptides. Whereas bacterial NRPSs use thioesterase (TE) domains to perform the cyclization, fungal NRPSs have apparently evolved to use a different enzymatic route. In verified fungal NRPSs that produce macrocyclic peptides, each megasynthetase terminates with a condensation-like (C_T) domain that may perform the macrocyclization reaction. To probe the role of such a C_T domain, we reconstituted the activities of the *P. aethiopicum* trimodular NPRS TqaA in *S. cerevisiae* and *in vitro*. Together with a reconstituted bimodular NRPS AnaPS, we dissected the cyclization steps of TqaA in transforming the linear anthranilate-D-tryptophan-L-alanyl tripeptide into fumiquinazoline F. Extensive biochemical and mutational studies confirmed the essential role of the C_T domain in catalyzing cyclization in a thiolation domain-dependent fashion. Our work provided evidence of a likely universal macrocyclization strategy employed by fungal NRPSs.

2.2.4 Materials and Methods

Cloning, Expression and Purification of TqaA and AnaPS. Cloning of the intact gene into the expression vector was performed using the modified yeast-based homologous recombination method [148]. The *tqaA* gene was divided into 4 pieces (P1 ~ P4) with a maximum size of ~4.5 kb. The only intron (577 - 661 bp) of the *tqaA* gene within P1 was removed by RT-PCR. P2 to

P4 were PCR amplified from the genomic DNA of *P. aethiopicum*. Each successive piece was designed to overlap (35 - 40 bp) with the two flanking ones, and the 5′-end of P1 and 3′-end of P4 overlap with the expression vector. The vector fragment was generated by digesting pXW55 (an expression vector with an engineered FLAG tag on *N*-terminus and 6×His-tag on C-terminus) with *Spe*I and *Pml*I. The individual *tqaA* pieces (1 µg per piece) as well as the vector (~ 0.5 µg) were co-transformed using the EasyComp™ Transformation Kit (Invitrogen) into *S. cerevisiae* strain BJ5464-NpgA to allow assembly of the entire gene *in vivo*. The assembled plasmid was miniprepped from *S. cerevisiae* using yeast plasmid miniprep II kit (Zymo) and further verified by PCR and restriction enzyme digestion. The cloning procedure for the *anaPS* gene was the same as described for *tqaA*. The verified plasmid containing intact *tqaA* or *anaPS* gene was transformed into *S. cerevisiae* strain BJ5464-NpgA for protein expression. For 1 L of yeast culture, the cells were grown at 28 ºC in YPD media with 1% dextrose for 72 hours. For detailed protein purification procedure, refer to **Supplementary Information**. FLAG-tagged TqaA and AnaPS proteins were purified by using ANTI-FLAG® M1 Agarose Affinity Gel (Sigma-Aldrich). The cleared cell lysate was directly applied onto a column packed with ANTI-FLAG Agarose Affinity Gel. TqaA/AnaPS protein was eluted with buffer containing 0.5 mg/ml FLAG peptide. For the cloning and expression procedures for other proteins mentioned in this paper, see **Supplementary Information**.

Synthesis of Peptidyl-SNAC, Peptidyl-*S*-CoA and Peptidyl-*S*-T Substrates. Syntheses of peptidyl-SNACs and peptidyl-*S*-CoAs are described in the **Supplementary Information.** Loading reactions of *apo*-T_3 or other ACP domains were carried out in 50 mM MES buffer (pH 6.8) and initiated by addition of 5 µM Sfp. After 2 hours, **31**-*S*-T_3 was buffer-exchanged into

MilliQ water using an Amicon Ultra 10 MWCO 10,000 Daltons, followed by identification with MALDI-TOF MS and used for kinetics assays.

***In Vitro* Assays.** To test the *in vitro* activities of TqaA, AnaPS or other mutant proteins, 2 mM of each amino acid building block, 6 mM ATP, 4 mM MgSO$_4$ and 10 µM of purified proteins were combined in 50 mM Tris-HCl buffer (pH 7.5) to a total volume of 100 µl. After 12 hours incubation, reactions were quenched and extracted with 1 ml of ethyl acetate. The organic layer was dried and re-dissolved in methanol for LC-MS analyses. For kinetic characterization of the C$_T$ domain towards **31**-*S*-T$_3$, the concentration of the C$_T$ domain was kept constant at 2 µM. Initial velocities were determined at various **31**-*S*-T$_3$ concentrations ranging from 20 to 1053 µM. Samples were analyzed on a Shimadzu 2010 EV liquid chromatography mass spectrometer using positive and negative electrospray ionization and a Phenomenex Luna 5 µm, 2.0 mm × 100 mm C18 reverse-phase column. Samples were separated on a linear gradient of 5 to 95% CH$_3$CN in water (0.1% formic acid) for 30 min at a flow rate of 0.1 ml/min followed by isocratic 95% CH$_3$CN in water (0.1% formic acid) for 15 min.

2.3   Elucidation of Downstream Tailoring Pathway of Tryptoquialanine

A version of this section was published as: Haynes, S. W., Ames, B. D., Gao, X., Tang, Y., Walsh, C. T. "Unraveling terminal C-domain-mediated condensation in fungal biosynthesis of imidazoindolone metabolites." *Biochemistry,* 2011, 50, 5668-5679. Copyright © 2011 American Chemical Society. The data in section "**TqaM, L and F were involved in the biosynthesis of 2-aminoisobutyric acid**" is unpublished.

2.3.1   Introduction

Having established the biosynthetic pathyway of tryptoqialanine and studied the product macrocyclization and releasing mechanism of the tri-modular NRPS TqaA, we next turn our attention to other tailoring enzymes involved in the biosynthesis of tryptoquialanine. Recently, a monomodular NRPS, Af12050 (of domain structure A-T-C) from the fumiquinazoline pathway in *A. fumigatus* was reported to catalyze the formation of fumiquinazoline A (FQA) from an activated oxidation product of fumiquinazoline F (FQF) [111]. Comparison of FQA biosynthesis to the tryptoquialanine (TQA) pathway of *P. aethiopicum* reveals closely related enzymatic logic for the generation of early pathway intermediates [143]. Both pathways are proposed to use a trimodular NRPS to condense and cyclize Anthranilate (Ant), Trp, and Ala to form FQF, and an FAD-dependent monooxygenase and monomodular NRPS for the oxidative-acylation of FQF to form a tricyclic imidazoindolone framework. In FQA biosynthesis, oxidation across the C2′ and C3′ of the pendant indole of FQF by the monooxygenase Af12060 is followed by alkylation with a unit derived from L-alanine catalyzed by the monomodular NRPS Af12050. In TQA biosynthesis, the monooxygenase TqaH (56/76% identity/similarity to Af12060) and monomodular NRPS TqaB (53/75% identity/similarity to Af12050) are postulated to catalyze an analogous process but yield a product with the opposite stereochemistry at C2′ and a *gem*-dimethyl at C11′ (the dimethyl group results from the preferential coupling of 2-AIB by TqaB). A small amount of the C2′ epimer of FQA (2′-*epi*-FQA) has also been observed in *P. aethiopicum* extracts, arising from TqaB-mediated coupling of L-Ala [143] (**Figure 17**).

The basis for the difference in stereochemistry at C2′ is likely to be dictated by FQF oxidation (by Af12060 or TqaH) or amino acid coupling (by Af12050 or TqaB). Because the stereochemistry of the C3′ hydroxyl group of FQA is identical to that found in 2′-*epi*-FQA, we

hypothesize that Af12060 and TqaH catalyze the same stereospecific C3′ hydroxylation of FQF. Consequently, we believe that the monomodular NRPS Af12050 or TqaB, and specifically the carboxy terminal C-domain, directs a specific stereochemical outcome at C2′ during amino acid coupling to the oxidized FQF scaffold (**Figure 17**). Although the consequences of the structural differences between FQA and 11′-dimethyl-2′-*epi*-FQA with regard to biological activity and effect on downstream enzymatic events are currently unknown, this differentiation provides an interesting example of divergence of secondary metabolite populations and divergent evolution of the corresponding biosynthetic enzymes.
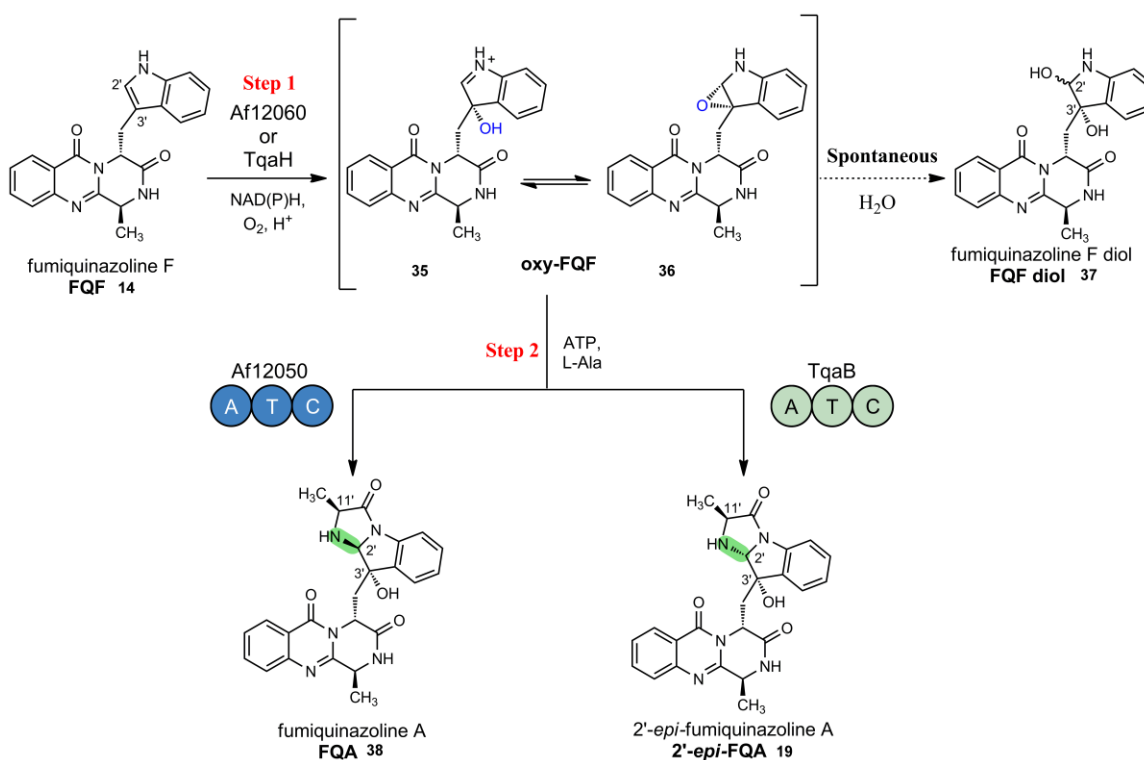
Figure 17. Stereochemical divergence of biosynthetic intermediates mediated by the action of a monomodular NRPS, Af12050 or TqaB, and illustration of spontaneous formation of diol shunt metabolite from the oxidized intermediate **35** or **36**.

72

The similarities between FQA from *A. fumigatus* and 2'-*epi*-FQA from *P. aethiopicum*, both in terms of structure and anticipated biosynthetic routes provide an opportunity to study the role of the terminal C-domain of the monomodular NRPSs in catalyzing regio- and stereospecific nucleophilic addition. In this work we reconstitute the activity of both the intact proteins and excised C domains from Af12050 and TqaB of the fumiquinazoline and tryptoquialanine pathways, respectively, in order to explore the function of the terminal C-domains in the enantioselective annulation of amino acids to indole in fungal alkaloid biosythesis. Our biochemical and bioinformatic analysis presents the first clear evidence of the catalytic activity and biological context of C-terminal condensation domains of fungal monomodular NRPSs.

2.3.2   Results and Discussions

Examination of the Effect of Monomodular NRPS (TqaB) and Monooxygenase (TqaH)

To gain initial insight into the enzymatic requirements for directing the stereochemical course of dual N-C bond formation for alanine coupling to oxy-FQF, we mixed-and-matched combinations of NPRS (Af12050 and TqaB) and monooxygenase (Af12060 and TqaH) and analyzed their respective products. The expression and characterization of Af12050 and Af12060 has been previously described [111]; for comparative analysis we overexpressed TqaH and TqaB as the corresponding His$_6$ tagged constructs from *E. coli* and purified them via Ni-NTA affinity chromatography. Interestingly, though TqaH and Af12060 share 76% similarity at the amino acid level, their expression profiles, purity, and flavin content vary significantly. Incubation of FQF with TqaH generated a peak at a retention time of 9 min that is identical to that seen upon incubation of FQF with Af12060 (**Figure 18**). Isolation and characterization of this compound was hindered by rapid decomposition to a complex mixture of apparent dimerization products

73

upon concentration and lyopholization (post preparative HPLC). However, removal of the dimerization products by SPE (Sep-Pak® C18) allowed us to identify this product by 1D and 2D NMR experiments as a diol shunt product (FQF diol, Figure **17**). The FQF diol arises from the quenching of the predicted unstable oxy-FQF epoxide/hydroxy imine product of Af12060 (or TqaH) with water. Co-incubation of TqaH with the monomodular NRPS from *A. fumigatus* Af12050 resulted in the production of a compound matching the retention time and UV pattern of FQA as previously characterized from an *in vitro* reconstitution of the natural Af12060/Af12050 system (**Figure 18**). By comparison when the monomodular NRPS TqaB is co-incubated with either monooxygenase (Af12060 or TqaH) a new peak is produced with a retention time of 9.4 minutes (**Figure 18**). This compound was isolated in sufficient quantity for investigation by NMR spectroscopy and structurally characterized as the C2′ epimer of FQA (2′-*epi*-FQA).
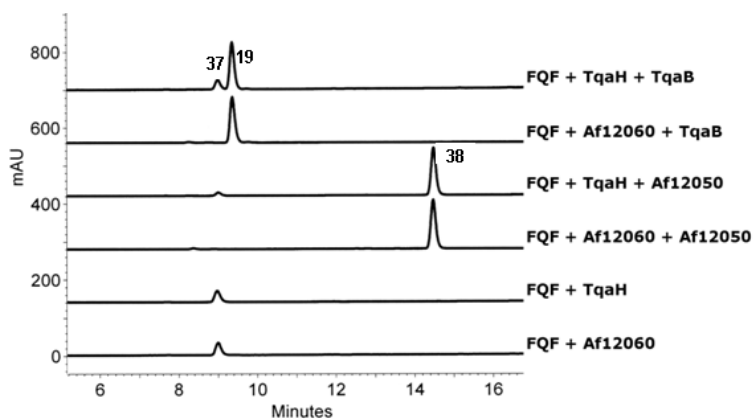
Figure 18. TqaH and TqaB functions. HPLC analysis detected at 254 nm of reactions containing 2.5 μM monoxygenase, 2 mM NADH, and 250 μM FQF in NaP$_i$ buffer, in addition to 5 μM holo-NRPS, 1 mM ATP, 2 mM MgCl$_2$ and 1 mM L-Ala when both monoxygenase and monomodular NRPS are included. **37**, FQF diol; **19**, 2′-*epi*-FQA; **38**, FQA.

Construction of C-domain swapped chimeric proteins

With the knowledge that the monomodular NRPSs Af12050 and TqaB (and not the monooxygenases) direct the stereochemical outcome at C2' of FQA and *epi*-FQA, respectively, we then wanted to test the hypothesis that the divergent NRPS-based activities could be specifically attributed to the carboxy terminal C-domains. In order to investigate the effect of upstream A and T domains on C-domain functionality we constructed two hybrid (or chimeric) proteins. The C-domains of TqaB and Af12050 were swapped, resulting in one protein containing the A and T domains from Af12050 and the C-domain from TqaB (Chimera 1) and another protein with the A and T domains from TqaB and
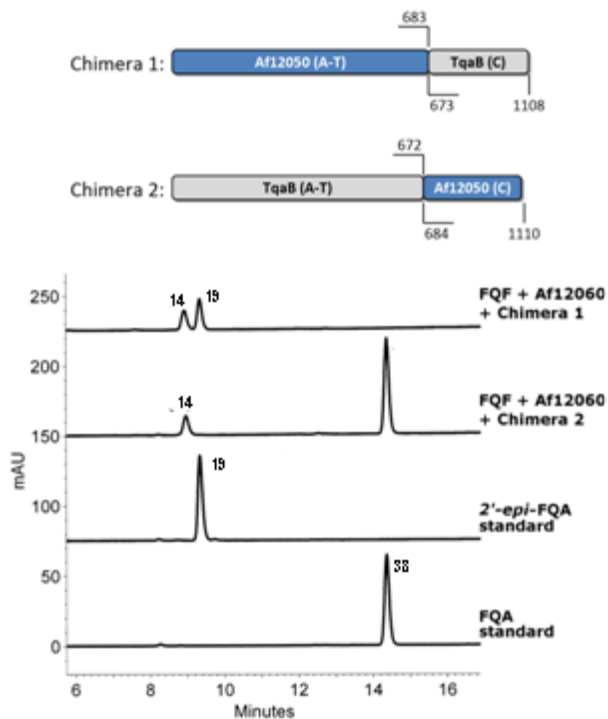
Figure 19. C domains of TqaB and Af12050 control the stereochemistry. (A) Cartoon representation of the structure of chimeras 1 and 2, numbers indicate the amino acid numbers from the native protein included in each chimera. (B) HPLC assay indicating the chimera mediated formation of FQA or 2'-epi-FQA with detection at 254 nm.

the C-domain from Af12050 (Chimera 2) (**Figure 19**). The enzymatic activity of Chimera 1 was demonstrated *in vitro* by incubation with FQF, monoxygenase (Af12060), and the required cofactors and shown to produce two peaks, one peak of retention time and characteristic UV pattern matching that of the FQF diol and the other matching that of 2'-*epi*-FQA. Incubation of Chimera 2 produced two peaks, one corresponding to FQF diol and the other matching the retention time and UV pattern of FQA (**Figure 19**). The ability of the chimeric constructs to activate and load a variety of amino acids was investigated via *in vitro* reconstitution of their

activity by including L-Ala, D-Ala or 2-aminoisobutyric acid (2-AIB). Chimera 1 efficiently activates, loads and condenses L-Ala with the FQF-derived oxidized species to give 2′-*epi*-FQA. 11′-*epi*-2′-*epi*-FQA is also produced by Chimera 1 when using D-Ala but at a level much below that seen for 2′-*epi*-FQA. Fuminquinazoline A (FQA) produced by *A. fumigatus* contains an interesting pendant indole moiety derived from a tryptophan unit that has been modified by the stereospecific addition of L-alanine to form a new 5-membered ring resulting in a tricyclic 5-5-6 imidazoindolone framework. FQA has been shown to be the direct precursor to two further previously identified metabolites fuminquinazoline C (FQC) and fuminquinazoline D (FQD) [122]. Characterization of intermediates from the tryptoquialanine pathway in *P. aethiopicum* has revealed a close similarity in the early biosynthetic steps to intermediates of the FQA pathway in *A. fumigatus*. Intermediates on pathway to tryptoquialanine which share a remarkable resemblance to FQA have been observed in mutants of *P. aethiopicum* and appear to be substrates for downstream enzymes which mediate their transformation to tryptoquialanine. Notably, 2′-*epi*-FQA identified in *P. aethiopicum* only differs from FQA in the stereochemistry at C2′ of the imidazoindolone moiety. The stereochemical difference between FQA and 2′-*epi*-FQA may influence the downstream fate of these two molecules to ultimately give FQC/FQD and (the structurally rearranged) tryptoquialanine, respectively. FQA formation in *A. fumigatus* is controlled by the action of a monomodular NRPS from the fumiquinazoline pathway (Af12050). The biosynthesis of 2′-*epi*-FQA in *P. aethiopicum* has been proposed to be controlled by the action of TqaB, based on its close homology to Af12050. The opposite stereochemistry at C2′ of the metabolites observed in *A. fumigatus* and *P. aethiopicum* poses the question of how the cyclization of an amino acid derived unit to generate the observed 5-membered ring is controlled:

76

specifically how the stereochemistry of addition of the amine nucleophile to the C2′ of the indole is differentially directed in FQA and 2′-*epi*-FQA biosynthesis.

Enzymatic assays of both TqaH and Af12060 revealed the product of each to be identical, whereas the activity of the A-T-C tridomain-containing Af12050 and TqaB resulted in two different products, FQA and 2′-*epi*-FQA, respectively. These data confirm the crucial role that the monomodular NRPS plays in directing the stereospecific formation of the C2′ carbon-nitrogen bond.

To determine whether the C-domain of the NRPS was in fact responsible for the facial specific addition as we proposed required the study of C-domain activity in a non-native environment. Characterization of chimeric constructs in which the C-domains of both Af12050 and TqaB have been swapped confirmed the enantiospecificity of FQA scaffold formation is consistent with that of the parent protein from which the C-domain of each chimeric construct was taken. The A-domain of Af12050 has previously been shown to preferentially activate L-Ala over D-Ala, which is in line with the observed amino acid coupling to oxy-FQF to form FQA in *A. fumigatus*. On the other hand, the A-domain of TqaB exhibits much broader substrate specificity, activating not only 2-aminoisobutyric acid as observed in tryptoquialanine but also both L-Ala and D-Ala.

To our knowledge, this is the first reported example of C-domain swapping yielding functional multidomain NRPS assembly lines. Previously, the N-terminal C-domain of the enterobactin synthetase EntF was excised and activity reconstituted *in vitro* [164], adenylation domain swapping has been performed [165, 166], and whole modules and domains of the tyrocidine synthetase have been fused to generated hybrid enzymes [167]. In our case we

recognize that the swapped domains are the carboxy terminal domains of monomodular proteins that are highly homologous, which facilitates the process of generating functionally intact hybrid

The condensation of a T-domain-tethered amino acid to the oxy-FQF scaffold by the action of the terminal condensation domains of Afl2050 and TqaB could proceed through two distinct pathways depending on the order of N-C bond formation. Both of these pathways differ significantly from standard C-domain chemistry. Classically a C-domain would catalyze the formation of an amide bond between a T-domain tethered upstream acyl-thioester intermediate with a nucleophile tethered to a downstream T-domain [109]. The biosynthesis of FQA (or 2'-*epi*-FQA) from FQF could be envisaged to proceed via one of two routes, both of which presumably proceed by an initial C-domain mediated addition, followed by a spontaneous intramolecular cyclization. Initial thioester cleavage by a C-domain catalyzed addition of the indole derived secondary amine followed by spontaneous ring closure by attack of the L-alanine derived primary amine aligns well with the conventional function of C-domain mediation of directing nucleophilic attack on a T-domain bound thioester. Indeed the observation of a C-domain catalyzing condensation with a free soluble nucleophile has been previously reported, for example in the biosynthesis of pseudomonine where condensation of a T-domain bound intermediate with a free amine monomer has been shown [168], or vibriobactin biosynthesis where a soluble small-molecule derived from norspermidine acts as the incoming nucleophile [169]. Alternatively, attack of the L-alanine primary amine may precede a subsequent ring closure by intramolecular spontaneous amide bond formation. This would require the C-domain to catalyze addition of a thioester bound nucleophile to a free soluble electrophile, via a process currently unprecedented in the chemistry of C-domain catalysis.

TqaM, L and F were involved in the biosynthesis of 2-aminoisobutyric acid (AIB) (unpublished data)

As we discussed in Section 2.1, genetic study has shown that TqaM and L are directly involved in the biosynthesis of the nonproteinogenetic amino aicd, 2-aminoisobutyric acid (AIB). Deletion of *tqaF* didn't totally abolish the production of tryptoquialanine, however, the yield of metabolites produced by the Δ*tqaF* strain is indeed decreased. In order to study the biosynthesis of AIB, firstly we cloned the TqaM, L and F into

Figure 20. TqaM, L and F are sufficient to produce AIB in *E.coli*. (A) SDS_PAGE gel of TqaM, L and F purified from *E.coli*; (B) OPA derivatization of AIB produced in *E.coli*; (C) isotope labeled amino acids feeding to *E.coli*. After 2,3 C13 L-Ala was added to the TqaM, L and F expressing *E.coli* culture, ~45% enrichment was observed in the OPA-AIB; however, no enrichment in the methyl-D3- L-methionine feeding culture.

pACYCDuet™-1, pET28a and pCDFDuet™-1, respectively. As shown in **Figure 20A**, all three proteins are solubly expressed when purified from *E.coli* BL21. We then co-trasformed the *tqaM, tqaL, tqaF* encoding-plasmids into *E.coli* strain. After induction with IPTG for overnight, we detect the fluoraldehyde (OPA) reagent derivatized AIB from the TqaM, L and F expressing strain. In the control assay, without IPTG induction or taking out any one or two of TqaM, L and F proteins, no OPA-AIB can be detected (**Figure 20B**). This indicates that TqaM, L and F are sufficient and necessary to making AIB from the primary metabolism in *E.coli*.

Figure 21. Crystallization of TqaM. (A) A picture of the TqaM crystal. (B) Overall structure of TaM; (C) Active site of TqaM. His126, His128 and His191 were well positioned with the Ni2$^+$.

Having reconsitituted the AIB biosynthesis in *E.coli*, we next want to exam how AIB derived from primary metabolism. L-Ala is the natural amino acid structurally most close to AIB, so 2, 3 C13 isotope labeled L-Ala was fed into the TqaMLF expressing *E.coli* culture. ~45% of the n+2 ([M+H]$^+$ = 310) mass peak was identified which L-Ala indicates might be the precursor for AIB. *S*-adenosyl methionine was commonly involved in the methyl transfer reactions which might be derived from methionine [170]. Deuterium labeled methyl-D3-L-methionine was also fed to the TqaMLF culture; however, no enrichment was found (**Figure 20C**).

In order to investigate the function of TqaM, a class II aldolase, we performed the X-ray crystallization experiments. TqaM was crystallized and the resolution is ~2.1 Å (**Figure 21A and Table 2**). 4 subunits were found in one crystal packing and each subunit has a single domain comprised of a central β-sheet flanked on either side by layers of α-helices (**Figure 22B**). One of the characteristics for the class II aldolase is the metal-dependent catalysis. We indeed observed the Ni^{2+} ion is well positioned with three histines (126, 128 and 191) which might be involved in the catalysis.

Here we reconstituted the AIB biosynthesis in the heterologous *E.coli* system, demonstrated that TqaMLF are sufficient and necessary to make AIB and L-Ala might be the precursor enoute to synthesize AIB. However, the detailed function and mechanism of each TqaM, L and F genes, as well as the further analysis of the metal content of TqaM under endogenerous condition, are under investigation.

Table 2. Data collection and refinement statistics for TqaM

	TqaM-native
Data collection	
Space group	P222$_1$
Cell dimensions	
a, b, c (Å)	140.2, 95.7, 97.3
α, β, γ (°)	90.0, 90.0, 90.0
protomers/asymmetric unit	4
Resolution (Å)	2.10
R_{merge}	0.081 (0.452)
$I/\sigma I$	18.8 (2.0)
Completeness (%)	97.4(79.3)
Redundancy	5.4 (2.6)
Refinement	
Resolution (Å)	2.10
No. reflections	74,556
R_{work}/ R_{free}	0.181/0.202
No. atoms	
Protein	8232
Ligand/ion	28
Water	381
B-factors (Å^2)	
Protein	41.3
Ligand/ion	42.6
Water	47.1
R.m.s deviations	
Bond lengths (Å)	0.010
Bond angles (º)	1.0

2.3.3 Conclusions

The fungal peptidyl alkaloids of the tryptoquialanine and fumiquinazoline families are nonribosomally assembled by annulation of the indole side chain of fumiquinazoloine F (FQF) with an alaninyl or aminoisobutyryl unit by monomodular NRPS enzymes containing adenylation, thiolation and condensation (A-T-C) domains. The enzyme pair Af12060 and Af12050 from *Aspergillus fumigatus* thereby convert FQF to FQA while the homologous enzyme pair TqaH and TqaB from *Penicillium aethiopicum* make the 2′-epi diastereomer of FQA, differing only in the stereochemistry of one of the C-N bonds formed in the annulation with L-Ala. To evaluate the basis for this stereochemical control we have mixed and matched the flavoprotein oxygenases Af12060 and TqaH with the A-T-C modular enzymes Af12050 and TqaB to show

that the NRPS enzymes control the stereochemical outcome. The terminal 50 kDa condensation domains of Af12050 and TqaB are solely responsible for the stereochemical control as shown both by making chimeric (e.g. A-T-C* and A*-T*-C) forms of these monomodular NRPS enzymes and also by expression, purification and assay of the excised C domains. The Af12050 and TqaB condensation domains are thus a paired set of enantioselective annulation catalysts on the fumiquinazoline F scaffold.

2.3.4 Materials and Methods

Cloning of TqaH, standalone C- and T-domains, and chimeric constructs. TqaH cDNA was amplified from *P. aethiopicum g*DNA using reverse transcription (RT)-PCR and cloned into the NdeI and EcoRI restriction sites of the pET28a vector for protein expression with an N-terminal His$_6$-tag (483 residues [53.2 kDa]). Standalone C- and T-domain constructs were cloned from plasmid DNA (encoding Af12050 [111] and TqaB [143]) into pET30 Xa-LIC and pET46 Ek-LIC vectors, respectively. Domain boundaries were determined by homology detection and structure prediction using the online server HHpred [171], and primers were designed to PCR amplify a nucleotide region corresponding to amino acids: 684-end, Af12050 C-domain (464 residues [53.4 kDa], N-terminal His$_6$-S-tag); 673-end, TqaB C-domain (483 residues [53.4 kDa], N-terminal His$_6$-S-tag); 591-678, Af12050 T-domain (103 residues [11.2 kDa], N-terminal His$_6$-tag); 581-668, TqaB T-domain (103 residues [11.2 kDa], N-terminal His$_6$-tag). The boundaries for the standalone T- and C-domain constructs are annotated on the Af12050/TqaB sequence alignment.

The domain boundaries determined above for the C-domains of Af12050 and TqaB were used for the generation of chimeric constructs by splice overlap extension PCR. Chimera 1

consists of the A- and T-domains of Afl2050 (amino acids 1-683) and the C-domain of TqaB (673-end), while Chimera 2 consists of the A- and T-domains of TqaB (1-672) and the C-domain of Afl2050 (684-end) (**Figure 19**). The chimeric constructs were cloned into pET52b 3C-LIC to generate proteins containing an N-terminal StrepII-tag and a C-terminal His$_6$-tag (Chimera 1, 1161 residues [127.9 kDa]; Chimera 2, 1159 residues [128.8 kDa]).

Crystalization condition for TqaM. 20 mg/ml TqaM was used to set up the crystallization trays. Crystals of TqaM were grown at RT by a hanging drop vapor diffusion method using a 2:1 protein to reservoir solution ratio for a total drop size of 1.5 μl. Diffraction quality crystals were obtained in 3~4 days when using 0.1 M Sodium acetate trihydrate pH 4.6, 12% w/v Polyethylene glycol 4,000, 10mM DTT as a reservoir solution. In preparation for data collection crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen.

2.4   Investigation of the Biosynthetic Pathway of Asperlicin

A version of part of this section was published as: Haynes, S. W., Gao, X., Tang, Y., Walsh, C. T. "Assembly of asperlicin peptidyl alkaloids from anthranilate and tryptophan: a two enzyme pathway generates heptacyclic scaffold complexity in asperlicin E." *J. Am. Chem. Soc.,* 2012, 134, 17444-17447. Copyright © 2012 American Chemical Society.

A version of part of this section was published as: Gao, X., Jiang, W, Jiménez-Osés, G, Choi S. M., Houk K.N., Yi Tang, Walsh C.T., "An Iterative, Bimodular Nonribosomal Peptide Synthetase that Converts Anthranilate and Tryptophan into Tetracyclic Asperlicins." *Chem. Biol., accepted in principle.*

2.4.1   Introduction

The *aspergilli* are prolific producers of polyketide and nonribosomal peptide natural products, some of which appear to be constitutive while others are conditional metabolites [3, 7]. The peptidyl alkaloids represent a prevalent class, typified by fused ring scaffolds with rigidified frameworks and extended molecular architectures [107, 143, 172-174]. These are assembled efficiently by short nonribosomal peptide synthetase (NRPS) assembly lines [22, 109] and then further modified by post-assembly line tailoring enzymes that can add electrophilic acyl, prenyl, and oxygen functionalities in scaffold maturation events [26, 111, 175]. We have recently examined the chemical logic and enzymatic machinery for assembly of *Aspergillus* signature metabolites such as fumiquinazolines from *Aspergillus fumigatus* [38, 122], ardeemins from *Aspergillus fischeri* [176] and asperlicins from *Aspergillus alliaceus* [177], which are notable for utilizing the nonproteinogenic β-amino acid anthranilate (Ant) as a chain-initiating building block [31, 178]. The tricyclic scaffold of fumiquinazoline F is assembled from Ant, L-tryptophan (L-Trp), and L-alanine (L-Ala) by trimodular NRPS enzymatic action [179].

In analogy, two molecules of Ant and one L-Trp building block are processed to the tetracyclic isomers asperlicin C and asperlicin D [177]. Surprisingly, the AspA NRPS presumed to assemble these asperlicin isomeric frameworks is predicted to be bimodular not trimodular by bioinformatic identification and subsequent genetic deletion analysis in the producer [177]. This suggests either an unusual iterative action in Ant activation or a possible third module encoded elsewhere in the *A. alliaceus* genome acting *in trans* with AspA. In this work we have expressed the 276 kDa AspA protein in *Saccharomyces cerevisiae*, purified it in soluble active form, and shown that it is necessary and sufficient to produce both asperlicin C and D in a constant ratio,

84

indicating they derive from a common intermediate. These tetracyclic product regioisomers represent a dramatic morphing of the linear tripeptide backbone of Ant-Ant-Trp.

2.4.2   Results and Discussions

Genetic Deletion of *aspA*, *aspB* and *aspC* validates their roles in asperlicin biosynthesis.

To validate the proposed function of this gene cluster (**Figure 22**), a double crossover deletion of the bimodular orf *aspA* was constructed [135, 180] in the producing *A. alliaceus* strain, resulting in loss of any detectable asperlicin metabolites, including asperlicins C, D, E and asperlicin itself (**Figure S38b**). Feeding of synthetic asperlicin C to the AspA mutant restored production of asperlicin and asperlicin E, whereas feeding of asperlicin D to this mutant did not result in the restoration of any downstream products (**Figure S38c** and **S38d**). When *aspB* is deleted via double crossover (via the same methodology to used to knock out *aspA*) production of asperilicin and asperlicin E is abolished and levels of asperlicin C are considerably increased (**Figure S38e**).
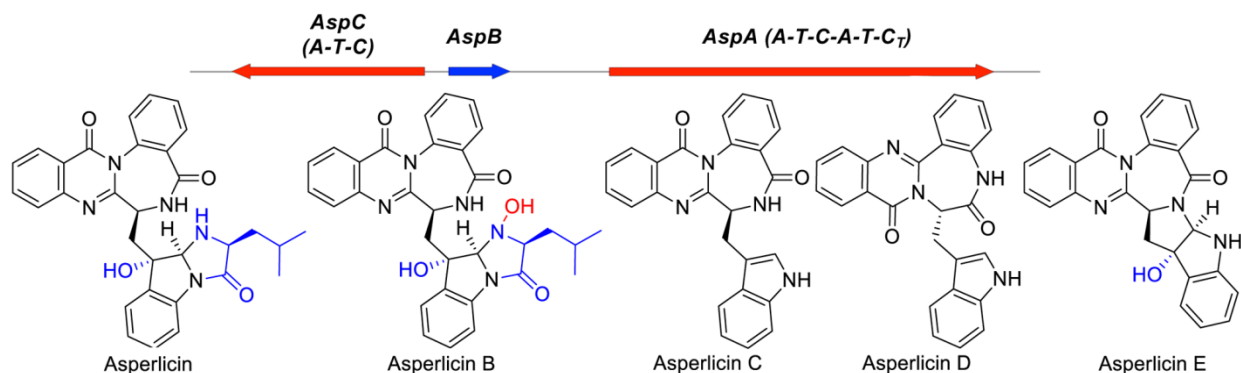
Figure 22. Schematic of *asp* cluster and structures of asperlicins

The third enzyme encoded in the gene cluster, Asp C, is proposed to be a valine actvating monomodular NRPS, akin to the alanine activating NRPS that captures the epoxygenated FQF species and annulates it to fumiquinazoline A [111]. To generate the asperlicin AspC presumably

provides a tethered valyl thioester to compete with the intramolecular capture of the epoxyindole that would yield asperlicin E as shown in **Figure 23**. Instead the NH_2 of the valyl moiety tethered to AspC could capture the epoxy indole, thus annulating the indole moiety to yield asperlicin, a distinct scaffold where the seven rings systems are in separate tetracyclic and tricyclic substructures (**Figure S39**). Analysis of that competition will require purified active AspC in subsequent studies. Meanwhile, the *aspC* knockout mutant gives asperlicin E as the major product (**Figure S39**), consistent with shutting down the conversion of asperlicin C to asperlicin.
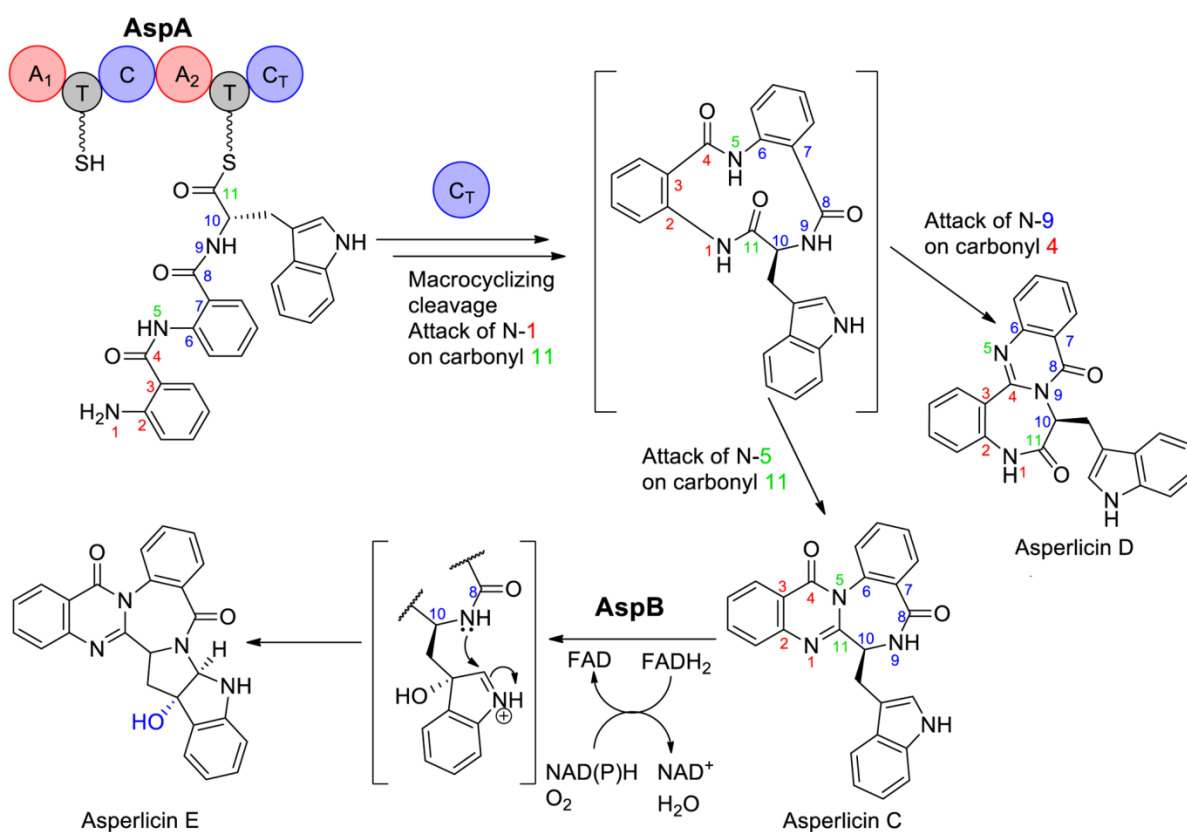
Figure 23. Asperlicin biosynthetic pathway en-route to heptacyclic asperlicin E, mediated by the action of AspA and AspB.

Expression and Purification of *A. alliaceus* AspA

86
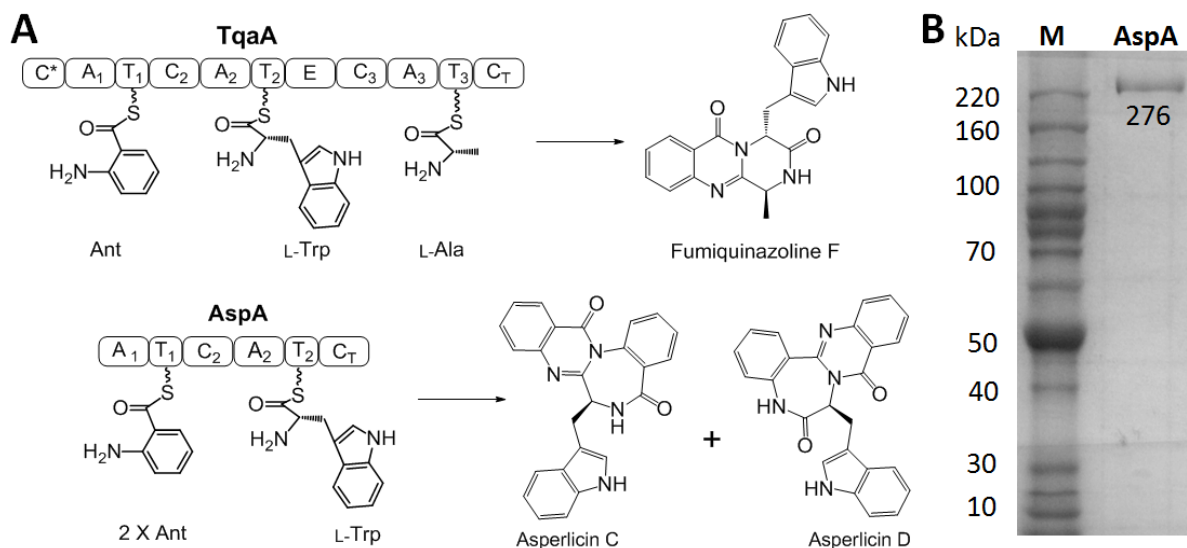
Figure 24. Heterologous expression of AspA in *Saccharomyces cerevisiae*. (A) Amino acid building blocks for assembly of fumiquinazoline F (FQF), and asperlicins C and D. AspA is a 276 kDa two-module NRPS enzymes with the indicated Adenylation (A), Thiolation (T) and Condensation (C) domains. The terminal condensation domains (C_T) of TqaA and AspA act as cyclization/release catalysts for Ant-D-Trp-L-Ala and Ant-Ant-L-Trp, respectively, presented as thioesters on the pantetheinyl arms of immediate upstream T domains. Intramolecular capture of the thioester carbonyl by the NH_2 of Ant_1 is the proposed common release mechanism. The subsequent transannular cyclizations and aromatizing dehydrations yield 6,6,6-tricyclic quinazolinedione (fumiquinazoline F) or tetracyclic 6,6,7,6 asperlicin C/D regioisomeric scaffolds, affected by the presence or absence of Ant_2 in the tripepitidyl-*S*-thiolation domain intermediates. (B) SDS-PAGE gel of AspA expressed and purified from *S. cerevisiae* BJ5464-NgpA.

To obtain the 276 kDa hexa-domain bimodular (A_1-T_1-C_2-A_2-T_2-C_T: where A = Adenylation, T = Thiolation and C = Condensation domains; C_T represents a terminal condensation domain at the end of the NRPS assembly line) NRPS AspA (**Figure 24A**) from *A. alliaceus* as a purified protein for catalytic characterization, we utilized *S. cerevisiae* strain BJ5464-NpgA that has vacuolar proteases deleted and contains a phosphopantetheinyl transferase (NpgA) that primes the *apo* forms of NRPS carrier domains [149]. We have previously found that this yeast strain will provide the TqaA trimodular NRPS in soluble, active form [179]. The DNA encoding the *aspA* gene was cloned by PCR with reverse transcription (RT-PCR) in six sections and one intron of 63 base pairs (bp) was removed from fragment 1 before ligation to yield a 7329 bp coding sequence, as described in **Material and Methods.** The

full length *aspA* gene with a C-terminal His$_6$ tag was then moved into the *S. cerevisiae* strain for expression and affinity purification.

As shown in **Figure 24B** the tagged AspA protein could be purified to apparent homogeneity in soluble form with a yield of 9 mg/L. No proteolytic fragments were detected, suggesting the six domain protein is likely well folded.

Assay of the Adenylation Domains of AspA for Amino Acid Activation
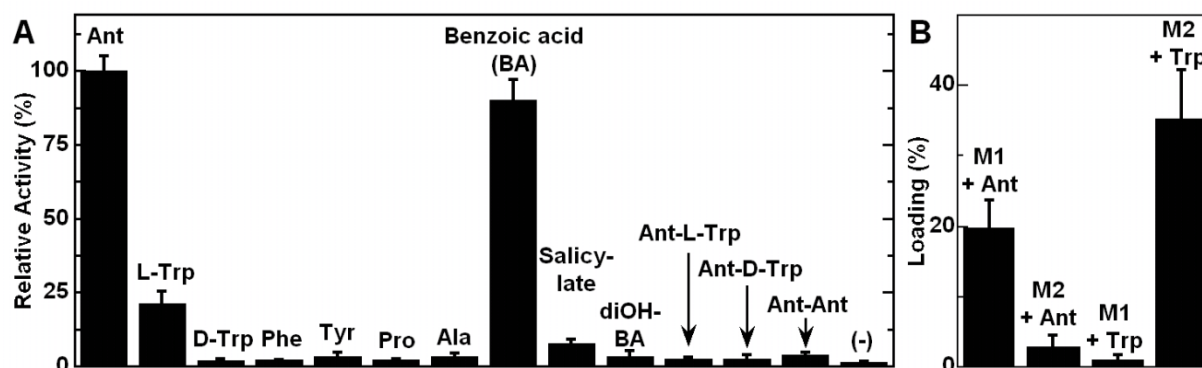
Figure 25. Anthranilate and L-tryptophan are substrates of AspA. (A) ATP-[^{32}P]PP$_i$ exchange assay of full length AspA; (B) Loading assay of A$_1$-T$_1$-C$_2$ (M1) and C$_2$-A$_2$-T$_2$-C$_T$ (M2) with [^{14}C]anthranilate or [^{14}C]tryptophan. M1 activates Ant while M2 activates L-Trp. Assay components are labeled in the figure. Data represent mean values ± s.d.

The scaffolds of the suite of known asperlicin metabolites suggest that two molecules of Ant and one tryptophan are utilized as building blocks. The purified AspA protein was assayed for the ability to form the aminoacyl-AMPs from Ant and L-Trp via amino acid-dependent exchange of radioactivity from ^{32}PP$_i$ into ATP, a classical assay for reversible formation of acyl-adenylates (Linne and Marahiel, 2004). As shown in **Figure 25A**, Ant supports robust exchange activity with k_{cat} = 1.6 sec^{-1}, while L-Trp is also active with lower rates (k_{cat} = 1.5 min^{-1}, **Figure S40**). No significant activity was detected with other proteinogenic amino acids. Benzoate and salicylate as Ant analogs also gave evidence of reversible formation of the acyl-AMPs (**Figure**

S40). Ant-L-Trp, Ant-D-Trp and Ant-Ant dipeptides didn't support exchange, suggesting they are not utilized as free intermediates (**Figure 25A**).

We also successfully cloned and expressed protein fragments spanning the two AspA modules. The A_1-T_1 construct was not solubly expressed but the A_1-T_1-C_2 tridomain (150 kDa) representing Module 1 (M1) was soluble at a yield of 15 mg L^{-1}). Likewise A_2-T_2-C_T was not soluble, but the four domain C_2-A_2-T_2-C_T (176 kDa) representing Module 2 (M2) version was soluble at a yield of 10 mg L^{-1} (**Figure S41**). As shown in **Figure 25A,** these protein constructs allowed unambiguous assignment that the A_1 domain in M1 activates Ant; and A_2 in M2 activates L-Trp. This result therefore strongly indicates the iterative use of Ant by M1 in synthesis of the tripeptide. In addition, it was possible to express and purify in soluble form of the terminal C_T (52 kDa, yield 4 mg L^{-1}) and the didomain T_2-C_T (61 kDa, yield 16 mg L^{-1}) fragment from M2 (**Figure S41**), which proved useful in studies with di- and tripeptidyl-*S-N*-acetylcysteamine (SNAC) surrogate substrates noted later.

Holo-AspA is Catalytically Competent to Generate Both Asperlicin C and D and Another Product Isomer

Having isolated the intact AspA protein and verified its A domains are catalytically active, we next examined if the bimodular NPRS can produce the tripeptidyl alkaloids. When 10 μM pure AspA was mixed with 1 mM Ant, 1 mM L-Trp, 3 mM ATP, 5 mM $MgCl_2$ in 50 mM Tris-HCl at pH 7.5 for 16 hours; and small molecules analyzed by HPLC, **Figure 26A** shows new major and minor peaks which chromatograph with the elution times of standard asperlicin C and asperlicin D, respectively. Analysis of the peaks by mass spectrometry confirmed that both product peaks had *m/z* = 407 $[M + H]^+$, equal to that of both asperlicin C and D isomers (**Figure**

S42). There is one additional enzyme-dependent minor product peak detectable in the HPLC traces, product **39**, and it too has the same $m/z = 407$ $[M + H]^+$ mass.
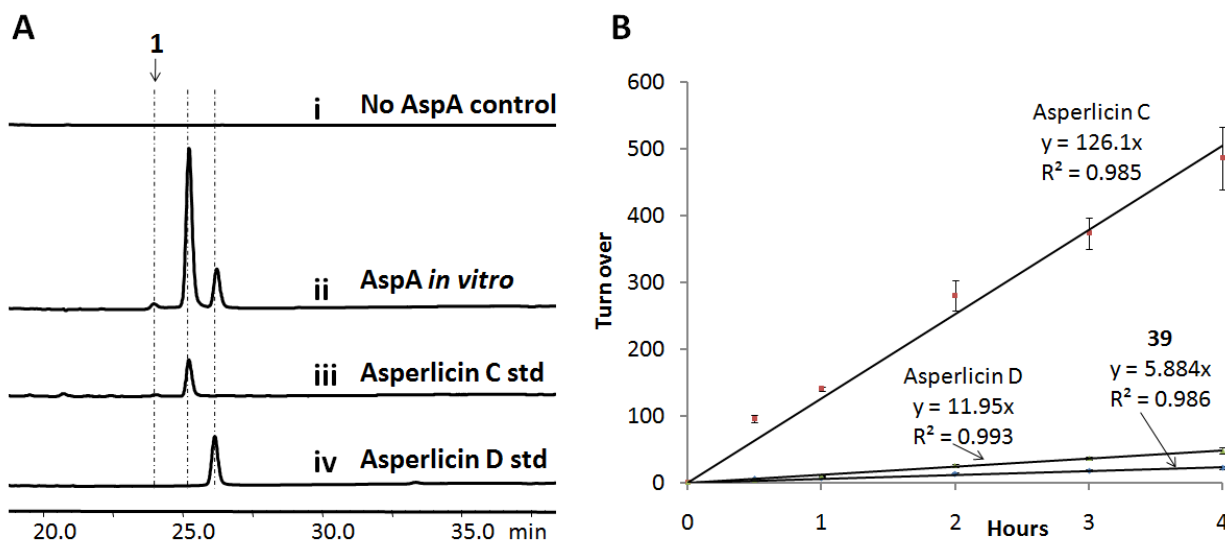
Figure 26. *In vitro* reconstitution of AspA. (A) LC/MS analysis (280 nm) shows purified AspA generates asperlicin C/D and a third product **39** of identical mass from Ant, L-Trp, and ATP; (B) Catalytic turnover number and product partition ratios for **39**, asperlicin C and D production. Both asperlicin C and D standards were synthesized from our previous study.

Rate assays under those conditions indicate linear formation rates for asperlicin C/D and the unknown product **39** (**Figure 26B**). Given the known extinction coefficients (almost the same at 280 nm) of the asperlicin C and D synthetic standards, the apparent turnover number for asperlicin C was calculated to be 126.1 ± 7.6 h^{-1} and asperlicin D is 12.0 ± 0.8 h^{-1}. That product ratio of ~10:1 does not change during the course of the incubation, suggesting the minor product asperlicin D is not forming from asperlicin C in a post reaction transformation, but is in fact an initially formed alternative product of AspA. Until the structure of the third isomeric product **39** is determined, we cannot confidently assign a turnover number to it. However, assuming this is a novel asperlicin isomer and might have a comparable A280 extinction coefficient, we would estimate an apparent turnover number of ~5.9 ± 0.3 h^{-1} which represents 5% of the product flux

that is also constant over the incubation time frame. Hence, all three products are likely derived from the same precursor, which we proposed to be an 11-membered macrolactam formed via macrocyclization of the tripeptide.

Assay of Truncated Forms of AspA for Product Formation: Activity of the Two Modules
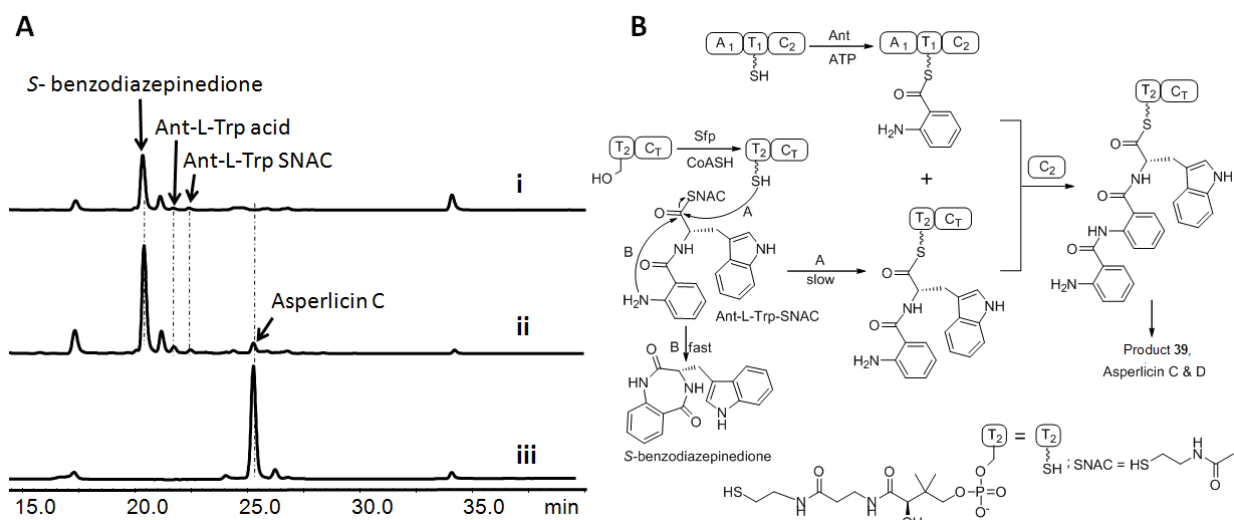
Figure 27. The first module of AspA iteratively utilizes two molecules of Ant. (A) The dipeptidyl Ant-L-Trp-SNAC is a surrogate substrate for M2 of AspA: the *holo* form of M1 (50 μM A_1-T_1-C_2) was preloaded with Ant (1 mM Ant, 3 mM ATP) for 1 hour while in parallel the T_2-C_T didomain was converted to the *holo* (HS-pantetheinyl) form using 20 μM Sfp and 1mM CoASH. The two solutions were mixed with addition of 400 μM Ant-L-Trp-SNAC and incubated overnight before aliquots were analyzed by LC/MS (280 nm). Trace i shows the spontaneous products (*S*-benzodiazepinedione and Ant-L-Trp acid) formed from an assay without T_2C_T di-domain; trace ii shows products formed from an assay including *holo* T_2C_T. In addition to the spontaneously formed products, formation of asperlicin C can be detected. Trace iii shows the profile of products formed from full length AspA starting from Ant, L-Trp and ATP. (B) Reaction scheme: M1 loaded covalently with Ant on the pantetheinyl arm of T_1 reacts with Ant-L-Trp-*S*-pantetheinyl-T_2-C_T to yield the tripeptidyl Ant-Ant-L-Trp-*S*-T_2-C_T form of the bi-domain T_2-C_T protein fragment. That can be acted on by C_T to yield asperlicin C, D and product **39**. The added Ant-L-Trp-SNAC is proposed to undergo acyl exchange onto the HS-pantetheinyl arm of T_2-C_T in competition with hydrolysis to Ant-L-Trp and intramolecular cyclization to *S*-benzodiazepinedione.

To examine the iterative features of AspA, we dissected the AspA NRPS into M1 and M2. Mixing of the purified M1 and M2 with Ant, L-Trp and ATP gave reconstitution of asperlicin production with asperlicin C and D in the same ratio as full length AspA and also the third minor product **39** (**Figure S43**). This validates the dissected modules can work *in trans* and the tethered, activated anthranilyl moiety can be transferred productively to M2. It seemed likely that the

second condensation domain C_2 would work twice in a catalytic cycle of AspA, first in the canonical mode to condense T_1-tethered Ant onto Trp-S-T_2 to yield Ant-L-Trp-S-T_2 and concomitantly free the thiol in T_1-SH (where -S-$T_\#$ designates the S-pantetheinyl thiolation domain in module #). If that Ant-L-Trp-S-enzyme had a sufficiently long lifetime, T_1 could re-load with another Ant; a second round of transfer catalyzed by C_2 would generate Ant-Ant-L-Trp-S-T_2. We sought to test these two predictions by synthesis of dipeptidyl Ant-L-Trp-SNAC, tripeptidyl Ant-Ant-L-Trp-SNAC and adding them as surrogate substrates to particular purified fragments of AspA. The expectation was that HS-pantetheinyl-T_2 could engage in thioester exchange with the di- and tripeptidyl-SNACs to generate the corresponding Ant-L-Trp-S-T_2 and Ant-Ant-L-Trp-S-T_2 forms, respectively, to allow evaluation of their catalytic competence.

To evaluate whether Ant-L-Trp-S-T_2 was on pathway, the purified terminal didomain T_2C_T (50 μM) was employed and converted to *holo* form by prior phosphopantetheinylation via the phosphopantetheinyltransferase Sfp and CoA-SH [154]. In parallel, *holo* M1 (50 μM) was incubated with Ant and ATP to produce the covalent Ant-S-T_1 form. Then the two proteins were combined, 400 μM synthetic Ant-L-Trp-SNAC was added, and product formation assayed after 1 hour. As shown in F**igure 27a** (trace ii), products with the diagnostic UV of asperlicins were detected; the anticipated pattern of asperlicin C/D and **39** were detected by LC/MS analysis with focus on extracted ion $m/z = 407$ [M + H]$^+$ (**Figure S44**). Nonenzymatic cyclization of Ant-L-Trp-SNAC to the bicyclic benzodiazepinedione and hydrolysis of the thioester to Ant-L-Trp acid compete with the enzymatic consumption as we have reported previously [179]. In the absence of the AspA fragments, these spontaneous products were the only compounds detected (**Figure 27A**, trace i). **Panel B** of **Figure 27** shows the proposed acyl exchange between SNAC and HS-

pantetheinyl-T_2 to yield the Ant-L-Trp-S-T_2 which is competent to receive the Ant moiety on T_1 and proceed through the catalytic cycle.

T_2C_T Bidomain Converts Ant-Ant-L-Trp-SNAC to the Suite of Asperlicin Products

The above results suggest that Ant-Ant-L-Trp-S-T_2 will be formed from Ant-S-T_1 and Ant-L-Trp-S-T_2 during the AspA catalytic cycle (**Figure 27B**). To evaluate that directly we prepared the corresponding Ant-Ant-L-Trp-SNAC and added it to either the purified C_T domain or to the purified T_2C_T protein construct. As shown in **Figure 28A** trace iv, no asperlicins are generated in the absence of AspA fragments, demonstrating the tripeptidyl thioester (unlike the Ant-Trp-SNAC) is stable in solution and any cyclization requires the action of an enzyme, most likely the C_T. However, the isolated C_T domain was incompetent (trace iii) but the T_2C_T (trace i) was able to generate the characteristic trio of asperlicin products at $m/z = 407$ $[M + H]^+$ in the presence of the tripeptide-SNAC as assessed in an overnight incubation (**Figure S45**). Thus, the T_2 domain appears essential in recognition/presentation of the tripeptide by the C_T domain. In a prior study by us [179], the same requirement held for presentation of Ant-D-Trp-L-Ala to the C_T domain of TqaA as a thioester bound to the immediately upstream T in that fumiquinazoline F-generating trimodular NRPS. In the absence of CoA and Sfp, assay containing T_2C_T still produced ~80% yield of the same product mixture as shown in trace ii. Taken together, we propose asperlicin formation in trace i and ii was initiated by the nonenzymatic transfer of the Ant-Ant-L-Trp moiety from SNAC to the HS-pantetheinyl arm of T_2 (**Figure 28B)** through thioester interchange. At that juncture the downstream C_T domain can catalyze the intramolecular amide bond formation(s) that constitutes release of the nascent cyclic product.
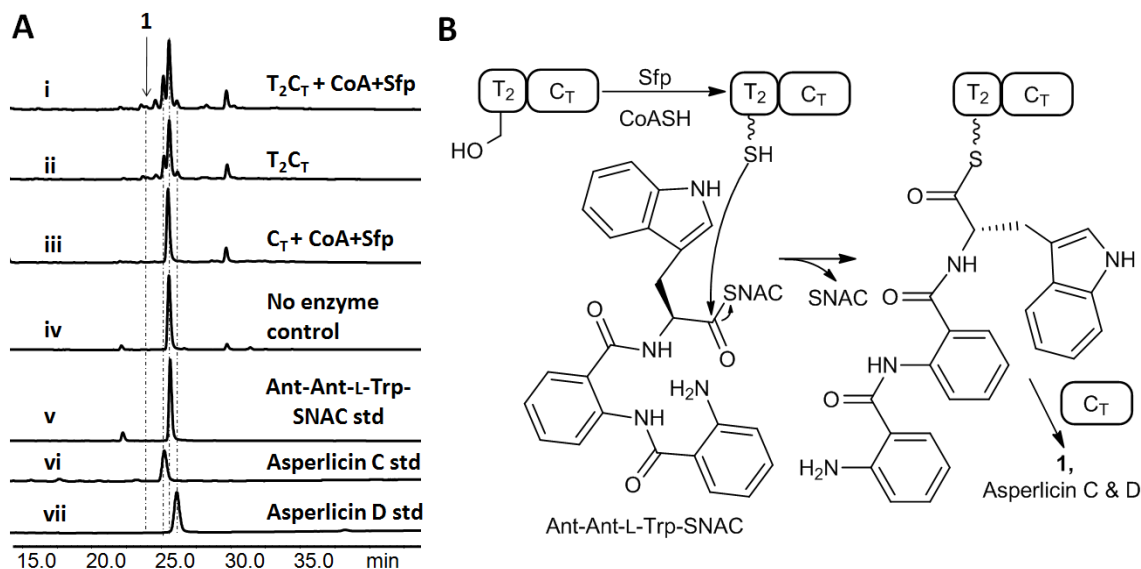
Figure 28. The T_2C_T di-domain fragment of AspA generates asperlicin C and D from exogenous Ant-Ant-L-Trp-SNAC. (A) Panel A shows UV-vis analyses of HPLC traces (280 nm) from the indicated incubations. Reactions contained 50mM Tris-HCl buffer, pH 7.5, in 100 μL. Reactions represented by traces i to iv contained 100 μM Ant-Ant-L-Trp-SNAC. 50 μM T_2C_T, 20 μM Sfp and 2 mM CoASH were added to reaction I; but only 50 μM T_2C_T was added in reaction ii; 50 μM C_T, 20 μM Sfp and 2 mM CoASH were added in reaction iii; No enzyme was added in reaction iv. Traces (v to vii) are standards. The suite of asperlicin C, D and **39**($[M+H]^+ = 407$) are found in traces i and ii but not iii and iv. (B) Reaction scheme featuring nonenzymatic acyl exchange from the –SNAC to the T_2 domain pantetheinyl thiol arm to yield the tripeptidyl-S-T_2 covalent enzyme as substrate for cyclizing release by the downstream C_T domain.

Proposed Cyclization Mechanisms of Asperlicin C and D and Computation Prediction of Product Distributions

Insight into that common released precursor from AspA comes from inspection of the differences between the isomeric asperlicin C and D and their molecular connectivity (**Figure 29**). The mechanistic assumption, which is supported by the reconstitution studies with intact and dissected AspA, is that the linear Ant-Ant-Trp-S-T_2, tripeptidyl-S-enzyme is a late intermediate. The terminal condensation domain C_T would then act to release the tripeptidyl chain by cyclization, analogous to that demonstrated role by the C_T domain of TqaA as it releases tricyclic fumiquinazoline F. While an alternative macrocyclization mechanism may be considered, the aniline NH_2 of Ant_1 will likely be the most competent nucleophile (**Figure 29A**). That release

94

step (attack of N_8 on (C_7=O)) would generate a tricyclic product 6,11,6-macrocycle **I**, fully analogous to the 6,10-bicyclic product (vide infra) we have proposed in TqaA action to generate fumiquinazoline F (FQF) (**Figure 24**). As shown in **Figure 29A**, the 11-membered macrocycle **I** could be subject to transannular attack with three different regiochemistries, by attack of each of the three amide nitrogen (N_1, N_5, N_8) on one of the three carbonyls [(C_4=O), (C_7=O), (C_{11}=O)].

Attack of N_1 on (C_7=O) across the ring generates a new 6-membered and a 7-membered ring, converting the nascent macrocycle **I** into a tetracyclic framework. Loss of water from the initial addition product yields the major product (~85% of the final product flux) asperlicin C. Alternatively, amide N_5 could capture (C_{11}=O). This also builds new 6- and 7-membered rings, generating a regioisomeric tetracyclic scaffold, a dehydration step away from the minor product asperlicin D (~10% of the flux). A third potential route would be transannular attack of N_8 on (C_4=O). That would generate a third regiosiomer, at the same mass of 406, but now with an angular 6,8,6,5 tetracyclic scaffold. The very minor product **39** with the identical mass could be this third isomer, but it was not isolable in sufficient quantity to prove structure.

In order to gain insight into the energetics of the three different modes of cyclization, we performed quantum mechanical calculations on the free energy changes ($\Delta\Delta G$) associated with each of the products shown in **Figure 29** (See Computational Details in **Material and Methods**). Interestingly, and as shown in **Figure 30**, asperlicin D is the most stable product among the three, while the structure proposed for product **39** in **Figures 29** and **30** is much higher in energy. However, examination of other possible cyclization outcomes of macrocycle **I** suggests that this is the only reasonable option. These results suggest that not the thermodynamic stability of the
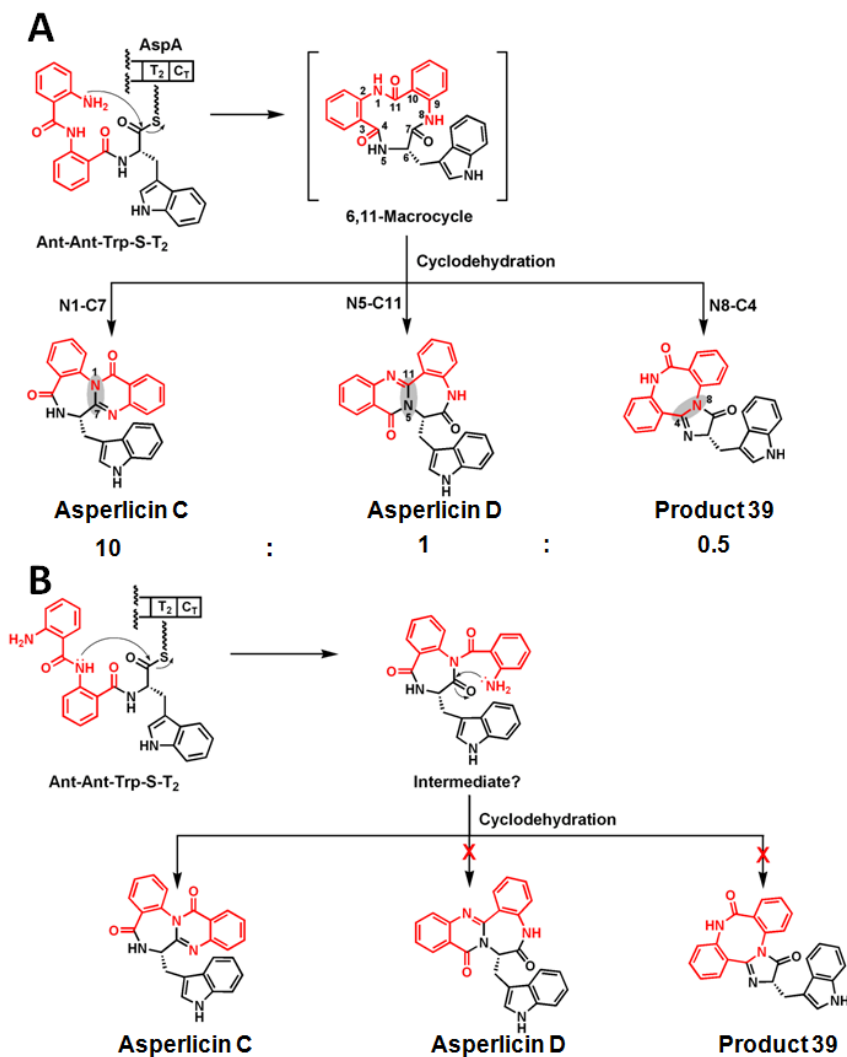
Figure 29. Mechanistic proposal for generation of a constant ratio of asperlicin C (major) and asperlicin D (minor) from macrocyclizing release of a linear Ant-Ant-L-Trp-S-T_2 thioester by the C_T domain. (A) The attack of the free NH_2 of the Ant_1 residue generates a transient 6,11,6-tricyclic product macrocycle I. Subsequent transannular attack and dehydration can proceed with distinct regiochemistries of intramolecular amide capture of carbonyl groups: N_1 on C_7=O gives major product asperlicin C, N_5 on C_{11}=O gives asperlicin D, while N_8 on C_4=O could give a third asperlicin regioisomer, putative product **39**. Whether these steps happen in one of the active sites of AspA or in solution have not yet been determined. (B) Alternative mode of tripeptidyl chain release by initial attack of the amide NH of Ant_2 residue on the thioester carbonyl would yield a diketopiperazine as nascent product. That can go on to asperlicin C as shown but has the wrong connectivity to generate the observed minor product asperlicin D and **39** and so is ruled out.

three possible adducts, but kinetic factors determine the regiochemical outcome of the reaction.

As also shown in **Figure 30**, starting from the 6,11,6-macrocycle **I**, different *trans* to *cis*

isomerizations of the peptide bonds must take place in order to reach conformations suitable for

cyclization. These are shown as macrocycles **II**, **III**, **IV** in **Figure 30** and are appropriate for formation of asperlicin C, D, and product **39**, respectively. The idea here is similar to Bruice's NAC, or "near-attack conformation" [66, 181-183]. Macrocycle **II**, the corresponding transition state (**TS2**) leading to asperlicin C entail the most favorable pathway, while the precursors to asperlicin D are much higher in energy, consistent with the observed 10:1 ratios of asperlicin C/D (**Figure 30**). In view of these results, it can be hypothesized that C_T domain is responsible for recognizing and stabilizing the adequate conformation of the 6,11,6-macrocycle by promoting the isomerization of the N_5–C_4=O peptide bond and directing the formation of asperlicin C. To what extent such nascent products are released and cyclized nonenzymatically or are cyclodehydrated on the way out of the active site of the C_T domain of AspA is a subject for subsequent evaluation.

We note that an alternate route of intramolecular release of an Ant-Ant-L-Trp-S-T_2 could be imagined, via formation of an anthranilyl-diketopiperazine. This would involve amide N_5 attack (from the Ant_2 residue) on the tripeptidyl thioester carbonyl (C_{11}=O) as the chain release step (**Figure 29B**). Whether that amide would be a kinetically competent nucleophile compared to the Ant_1 free NH_2 groups seems less likely. Moreover, while that putative diketopiperazine (DKP) could then undergo cyclization and dehydration to asperlicin C, it has the wrong connectivity to get to the observed asperlicin D framework, as noted in **Figure 29B**. Thus, the DKP release route is ruled out and formation of the 6,11,6-tricyclic macrocycle **I** seems the likely mechanism of AspA.
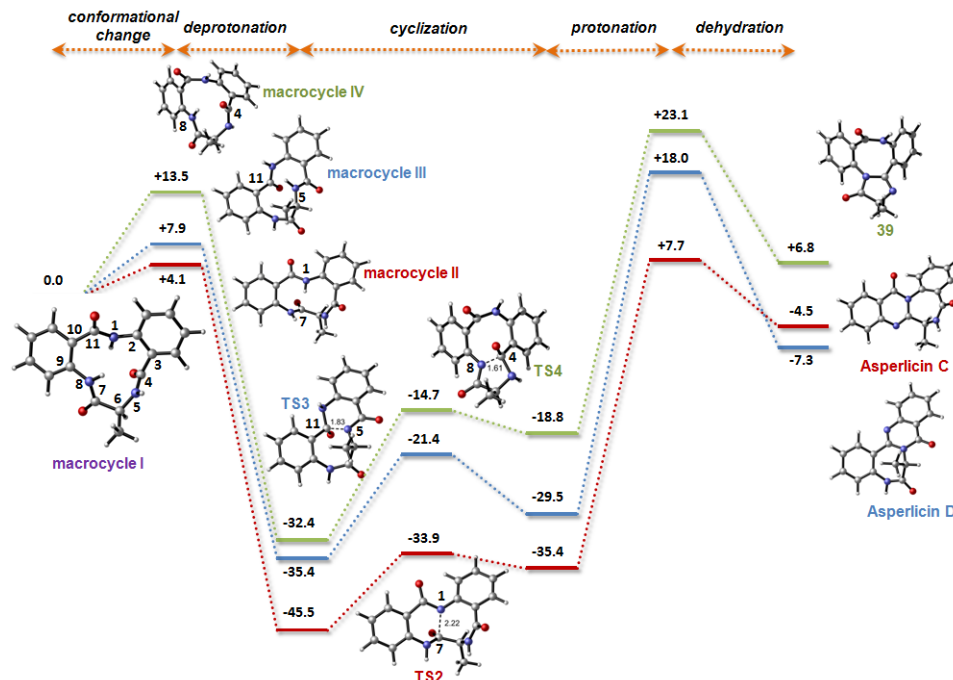
Figure 30. Structures and relative energies calculated for the different conformations of the initial macrocycle, cyclization transition states and final dehydrated products relevant to the biosynthesis of asperlicin C, D and **39**. The calculations were performed at the PCM(water)/B3LYP/6-31G(d) level using reduced models in which the indole ring from Trp side chain has been replaced by a methyl group. Amide bonds subjected to isomerization from *trans* to *cis* orientation prior to cyclization are marked in blue. Relative free energies are in kcal/mol and distances in Angstroms.

Our initial identification of the *aspABC* gene cluster by sequencing of the *A. alliaceus* genome relied on gene knockouts of *aspA* and *aspB*, the NRPS- and epoxygenase-encoding genes respectively, and the accumulation of different asperlicins in extracts from those mutants. We had synthesized authentic samples of asperlicin C and asperlicin D both as standards and to evaluate their capacity to serve as substrates for purified AspB, the indole epoxygenase [177]. Given that pure AspB acted selectively on asperlicin C (to yield asperlicin E) and not asperlicin D, we concluded that while asperlicin D was a natural *Aspergillus* metabolite [174], it was a dead end for subsequent processing to either asperlicin E or asperlicin itself. The question remained how and why asperlicin D formed as a co-product with asperlicin C.

In this study we have turned to characterization of the NRPS AspA enzyme. Bioinformatic analysis of AspA from the genome sequence predicted it would have only two modules (A_1-T_1-C_2-A_2-T_2-C_T), whereas asperlicin C and D are comprised of three amino acid units, two molecules of Ant and one Trp. Thus, a question at the start of this study was whether there was an additional, as yet missing, monomodular NRPS to produce a third module, on which a tripeptidyl (Ant-Ant-L-Trp)-thioester could be tethered. Alternatively, it was possible that bimodular AspA could act twice on Ant and generate a linear tripeptidyl-S-T_2 intermediate to be released by cyclizing action of the C_T domain.

To address whether AspA is sufficient to generate and release both asperlicin C and asperlicin D, we constructed a C-terminal His$_6$-tagged version of the 276 kDa AspA for expression in the vacuole protease-deficient *S. cerevisiae* strain BJ5464-NpgA. This strain has been of use to us for cloning the comparably sized bimodular AnaPS as well as the trimodular TqaA in aszonalenin and tryptoquialanine pathways, respectively. Soluble AspA with the pantetheinyl-SH prosthetic groups installed posttranslationally on T_1 and T_2 was thereby obtained in a yield approaching 9 mg/L, suitable for biochemical characterization.

First, we assayed the capacity of *holo* AspA to activate Ant and L-Trp by reversible formation of the aminoacyl-AMPs and validated that each A domain was active. Because both A_1 and A_2 are contained in the same protein, one could not determine from such radioactive exchange assays if A_1 was activating Ant and A_2 was activating Trp, or A_2 activated both Ant and Trp. To that end, the expression and separate purification of C-terminally His$_6$-tagged M1 (A_1-T_1-C_2) and M2(C_2-A_2-T_2-C_T) allowed that distinction. Inclusion of C_2 in both constructs was necessary to achieve soluble expression of A domain containing fragments. The A_1-containing

99

protein activated Ant and the A_2-containing module activated L-Trp as Trp-AMP but did not act on Ant. The two module AspA assembly line thus tethers Ant to T_1 and L-Trp to T_2. The Ant-Ant and Ant-Trp dipeptides were not substrates for either A_1 or A_2 (**Figure 25A**). If A_1 made an Ant-Ant-tethered thioester before transferring to Trp-S-T_2, one might have expected detection of the cyclic Ant-Ant dimer. No such dimer was detected from A_1-T_1-C_2 incubations with Ant and ATP (**Figure S46**). Mixing the M1 and M2 reconstituted asperlicin C and D formation as efficiently as the intact AspA, proving the separate modules could function *in trans*, and that the second copy of the C_2 domain was not a problem for reconstitution (**Figure S43**). M1 acts iteratively, first to generate a canonical Ant-Trp-S-T_2 that is then elongated by a reloaded Ant-S-T_1 to give the Ant-Ant-Trp-S-T_2 as full length intermediate that undergoes release by intramolecular cyclization.

Full length, purified AspA is clearly sufficient to make the tetracyclic scaffold of asperlicin C from 2 molecules of Ant, 1 L-Trp. It also generates asperlicin D, a known *A. alliaceus* minor metabolite. In this study a 10:1 constant ratio of asperlicin C:D is observed by kinetic analysis as asperlicin D represents ~10% of the product flux. The constant ratio indicates both products are formed from a common intermediate or nascent product. From the failure of asperlicin D to be carried forward by AspB, the epoxygenation enzyme that takes asperlicin C on to asperlicin E, we have reasoned that asperlicin D is a dead end metabolite and so represents off pathway partitioning of a common precursor. The mixture of asperlicin C and D (and product **39**) are also produced in the two module reconstitution experiment as well as in the incubations with Ant-Trp-SNAC and Ant-Ant-Trp-SNAC as surrogate substrates to probe the nature of later stage intermediates during AspA catalysis. Computational analysis clearly indicated that while

asperlicin D is more stable, asperlicin C is the kinetically favored product, thereby explaining the observed preference in the formation of asperlicin C from AspA. The extent to which C_T domain influences the near attack conformation of the macrolactam in its active site is unclear. However, it is evident that the formation of asperlicin D cannot be suppressed under experimental conditions (both *in vivo* and *in vitro*), thereby giving support to a mechanism of transannular cyclization in which the outcomes are dictated by free energy barriers.

The AspA NRPS illustrates the unusual cyclopeptide ring sizes attainable in the asperlicin system and is distinct from the related fumiquinazoline and ardeemin systems, where comparable 6,10-bicyclic nascent products gave rise to regioisomeric tricyclic quinazolinedione scaffolds (**Figure 24**). In both cases we argue that the terminal C_T domains are the chain release catalysts by intramolecular capture of the thioester carbonyl by the Ant_1-NH_2 group, yielding a 6, 10 (TqaA) or 6,11,6 (AspA) tricyclic nascent product. Subsequent closure to the very different 6,6,6-tricyclic versus 6,6,7,6-tetracyclic scaffold reflects the use of an additional Ant unit in the second position of the tripeptide framework, uniquely in the asperlicin assembly line. One planar β-amino acid unit (Ant) at the amino terminus of a tripeptidyl thioester directs cyclization to the tricyclic quinazolinedione framework. The second Ant β-aminoacyl building block adds another carbon and incorporates the fourth ring into the final asperlicin scaffolds.

While we proposed the aniline amine on the first Ant residue is responsible for initiating the macrolactam formation, there are alternative nucleophiles that may be considered to initiate the cascade of reactions. As shown in **Figure S47**, in addition to the possibility that the aniline nitrogen serves as the product releasing nucleophile (mechanism 3), two alternative mechanisms can be envisioned. In mechanism 1, the (C_4=O) carbonyl oxygen in imide form could be

sufficiently nucleophilic to first attack the thioester to generate an oxazol-5(4H)-one ring, followed by attack of the (C_{11}=O) carbonyl oxygen and undergo ring expansion to yield a 9-membered ring, which can then be opened to the 6,11,6-macrocylce **I**. Alternatively in mechanism 2, the amide nitrogen bridging the two Ant residues can initiate attack on the thioester to form a 7-membered ring, followed by ring expansion to the 6,11,6-macrocycle **I**.

To assess the possibilities of these routes, we computed the relative free energies of each transition state and product among the three possible mechanisms (See Computational details in **Matherial and Methods**). In order to locate and characterize transition structures for these reactions, fully deprotonated nucleophiles (both amides and anilines) were used despite the low acidities of aniline and amide groups (**Figure S48**) for attempts to locate neutral TS including explicit solvation and a general base. Such reactions require general base catalysis, and this is assumed to take place in the enzyme active site. The inclusion of a general base and solvent would be necessary to compute the barriers to various reactions occurring in water or the enzyme. Our calculations therefore only test the factors aside from deprotonation energetics that are required in order to distort the different intermediates along the reaction coordinate into various transition state geometries. The energetics of deprotonation of different nucleophiles by bases would be contributors to the actual free energies of reaction. As shown in **Figure S47**, while formation of the oxazolone product in mechanism 1 is kinetically possible ($\Delta\Delta G^{\ddagger}$=14 kcal mol^{-1}), it represents a thermodynamic dead-end that is unlikely to undergo additional modifications. In fact, subsequent ring expansion reactions towards either 9-membered or 7-membered rings are kinetically unfeasible ($\Delta\Delta G^{\ddagger}$ > 35 kcal mol^{-1}). In mechanism 2, calculations showed that the attack of the amidate nitrogen to form the 7-membered ring is highly unfavored ($\Delta\Delta G^{\ddagger}$=34 kcal

mol^{-1}) and is therefore unlikely to take place. In contrast, the direct attack of the aniline in mechanism 3 is the favored pathway, since the activation barrier is low ($\Delta\Delta G^{\ddagger}$=13 kcal mol^{-1}). These results are in good agreement with the observed nonenzymatic reactivity of dipeptide Ant-L-Trp-SNAC, which undergoes fast cyclization exclusively through the terminal aniline.

In conclusion, we have focused on the mechanism of AspA-catalyzed formation of asperlicin C and D. In a prior study we showed that the monooxygenase AspB will take asperlicin C and do an oxygenative cyclization to give asperlicin E, now with a fused heptacyclic framework. This is remarkable complexity generation from a two enzyme pathway with economical strategy and catalytic execution.

2.4.3 Conclusions

Fungal peptidyl alkaloids represent a group of compounds with diverse chemical structures and important biological activities. The planar nonproteinogenic amino acid, anthranilate (Ant), is an important building block in many bioactive fungal peptidyl alkaloids. Here we heterologously expressed the bimodular NRPS AspA in *Saccharomyces cerevisiae* and successfully reconstituted AspA to produce the regioisomers asperlicin C and D *in vitro*. Significantly, differing from the canonical, colinear programming rule of NRPSs, in which every module activates and appends one amino acid to the growing peptide, we showed the first module of AspA iteratively uses two molecules of Ant to build the Ant-Ant-Trp tripeptide precursor. The C-terminal condensation domain (C$_{T}$) was demonstrated to cyclize the linear tripeptide and to produce a macrocycle that can undergo different intramolecular cyclization fates. Experimental and computational studies were performed to examine the regioselectivity

and energetics of this step, which showed the kinetically most favored product asperlicin C dominates over the thermodynamically more stable product asperlicin D.

2.4.4   Materials and Methods

Cloning of Intact *aspA* Ggene from *A. alliaceus.* The AspA encoding gene was assembled from six pieces (P1-P6, each ~1 to 1.5 kb) by using modified yeast-based homologous recombination methods. The only intron (493–555 base pairs (bp)) in *aspA* gene was found by PCR with reverse transcription (RT–PCR) and no other introns were found in other regions. The assembled AspA expression plasmid was recovered from *S. cerevisiae* using a yeast plasmid miniprep II kit (Zymo Research) and was verified by restriction digestion and PCR. Protein expression and purification procedures are described in **Supplemental Methods**.

ATP-[^{32}P]PP$_i$ Exchange Assay for AspA. A typical reaction mixture (500 μL) contained 1.0 μM AspA, 1 mM substrate (unless specified), 5 mM ATP, 10 mM MgCl$_2$, 5 mM Na[^{32}P]-pyrophosphate (PP$_i$) (~1.8 x 10^6 cpm mL^{-1}), and 100 mM Tris-HCl (pH 8). Mixtures were incubated at ambient temperature for regular time intervals (e.g., 5 min), and 150 μL aliquots were removed and quenched with 500 μL of a charcoal suspension (100 mM NaPPi, 350 mM HClO$_4$, and 16 g L^{-1} charcoal). The mixtures were vortexed and centrifuged at 13000 rpm for 3 min. Pellets were washed twice with 500 μL of wash solution (100 mM NaPPi and 350 mM HClO$_4$). Each pellet was resuspended in 500 μL wash solution and added to 10 mL Ultima Gold scintillation fluid. Charcoal-bound radioactivity was measured using a Beckman LS 6500 scintillation counter.

Loading of [^{14}C]-substrate onto NRPS. A 50 μL assay mixture containing 100 mM HEPES (pH 7), 10 mM MgCl$_2$, 2 mM ATP, 1mM TCEP, 10 μM AspA-M1/M2, and 40 μM [^{14}C]-labeled

substrate (anthranilate or tryptophan) was incubated at ambient temperature for 30 min. The reaction was quenched by 600 µL 10% trichloroacetic acid with addition of 100 µL of 1 mg mL^{-1} BSA. The mixture was vortexed and centrifuged at 13,000 rpm for 3 min. The pellet was then washed twice with 600 µL 10% trichloroacetic acid, dissolved in 250 µL formic acid, added into 10 mL Ultima Gold scintillation fluid and subjected to a Beckman LS 6500 scintillation counter.

Synthesis of Ant-Ant-L-Trp-SNAC. Tripeptide Ant-Ant-L-Trp was custom synthesized by GenScript USA Inc.(Piscataway, NJ). 30 mg Ant-Ant-L-Trp (1.0 eq), 140 mg PyBOP (4.0 eq), 37 mg K$_2$CO$_3$ (4.0 eq) and 145 µL N-acetylcysteamine (SNAC) (20 eq) were dissolved in 20 mL H$_2$O:THF (1:1). The solution was stirred at room temperature for two hours. The mixture was concentrated *in vacuo*, dissolved in acetonitrile, and purified by prep-HPLC (Luna, C18 250 x 21.2mm, 10 µm, 100 Å) using a chromatographic gradient: 20-40% B, 5min; 40-80% B, 25 min; 80-100% B, 5 min (A: H$_2$O; B: acetonitrile, 10 mL min^{-1}, monitor at 340 nm). The peak with expected mass (m/z calculated for Ant-Ant-Trp-SNAC C$_{29}$H$_{29}$N$_5$O$_4$S [M+H]$^+$ 544.2013, found 544.2023) was collected and lyophilized. The final yield is 19.6 mg (53%).

HPLC-Based Time Course Study of Product Formation by AspA. Master reactions (500 µL) contained 1 µM AspA, 3 mM ATP, 2 mM MgCl$_2$, and 1 mM amino acid substrates (Ant and L-Trp) and AspA in 50 mM Tris-HCl buffer (pH 7.5) were carried out at 25 °C and 100 µL aliquots at 1, 2, 3, and 4 h time point were quenched by adding 1 mL of ethyl acetate. The initial product turnover rates were calculated with mean values ± SD. by using the data points within the linear range. The ethyl acetate layer was dried and redissolved in methanol (100 µL), and 20 µL samples were subjected to LC-MS analyses. Peak areas (at 280 nm) of the asperlicin C, aperlicin

D and **1** were converted to concentrations and were used to calculate initial enzymatic rate (μM h^{-1}).

Computational Details. Calculations were carried out with the B3LYP hybrid functional [184, 185] and 6-31G(d) basis set. Full geometry optimizations and transition structure (TS) searches were carried out with the Gaussian 09 package [186]. The possibility of different conformations was taken into account for all structures, and only the lowest energy structures are discussed. Frequency analyses were carried out at the same level used in the geometry optimizations, and the nature of the stationary points was determined in each case according to the appropriate number of negative eigenvalues of the Hessian matrix. The harmonic oscillator approximation in the calculation of vibration frequencies was replaced by the quasiharmonic approximation developed by Cramer and Truhlar [187]. Scaled frequencies were not considered since significant errors in the calculated thermodynamic properties are not found at this theoretical level [188, 189]. Where necessary, mass-weighted intrinsic reaction coordinate (IRC) calculations were carried out by using the Gonzalez and Schlegel scheme [190, 191] in order to ensure that the TSs indeed connected the appropriate reactants and products. Bulk solvent effects were considered implicitly by performing single-point energy calculations on the gas-phase optimized geometries, through the SMD polarizable continuum model of Cramer and Thrular [192] as implemented in Gaussian 09. The internally stored parameters for water were used to calculate solvation free energies (ΔG_{solv}).

2.5   Investigation of the Biosynthetic Pathway of Ardeemin

A version of this section was published as: Haynes, S. W., Gao, X., Tang, Y., Walsh, C. T., "Complexity Generation in Fungal Peptidyl Alkaloid Biosynthesis: a Two Enzyme Pathway

to the Hexacyclic MDR Export Pump Inhibitor Ardeemin." *ACS Chem. Biol.* 2013, DOI: 10.1021/cb3006787. Copyright 2013, American Chemical Society.

2.5.1 Introduction

Members of the fungal genus *Aspergillus* make a variety of peptidyl alkaloids based around the non-proteinogenic aryl β-amino acid anthranilate (Ant) [31, 108, 112, 193]. They show a wide range of biological activities, including tremorgens such as tryptoqualanines [143] developmentally regulated fumiquinazolines [107], asperlicins that act as cholycystokinin antagonists [172, 194], and ardeemins that block MDR export pumps in tumor cells (**Figure 31**) [173, 195]. The fungi build these complex multicyclic quinazolindione scaffolds from two to four amino acids in short efficient pathways.

The peptidyl alkaloid frameworks arise from amino acid building blocks selected and elongated by nonribosomal peptide synthetase (NRPS) machinery [109], which are then further processed by dedicated post assembly line tailoring enzymes for oxidations, acylations, and alkylations (e.g. prenylations). The nonproteinogenic aryl β-amino acid anthranilate along with the proteinogenic tryptophan are key constituents of the multicyclic scaffolds of aszonalenins [26], fumiquinazolines,[38] and the tryptoqualanines [143] .

We have recently deciphered the fungal code for anthranilate-activating NRPS adenylation domains in *Aspergillus* species and demonstrated that anthranilate is the chain-initiating unit [38], In bimodular NRPS enzymes, Ant activation is followed by L-Trp to generate aszonalenin [26] Or, in a trimodular NRPS system [38, 179], the ~450 kDa proteins follow Ant with both Trp and L-Ala and convert the three building blocks into the tricyclic (6,6,6) fumiquinazoline F scaffold (**Figure 32**). We have shown this scaffold to be a common early

intermediate in both the generation of the heptacyclic fumiquinazoline C in *A. fumigatus* [107, 111, 122] and the production of the neurotoxic tryptoquialanines [143] by *Penicillium aethiopicum* [179].
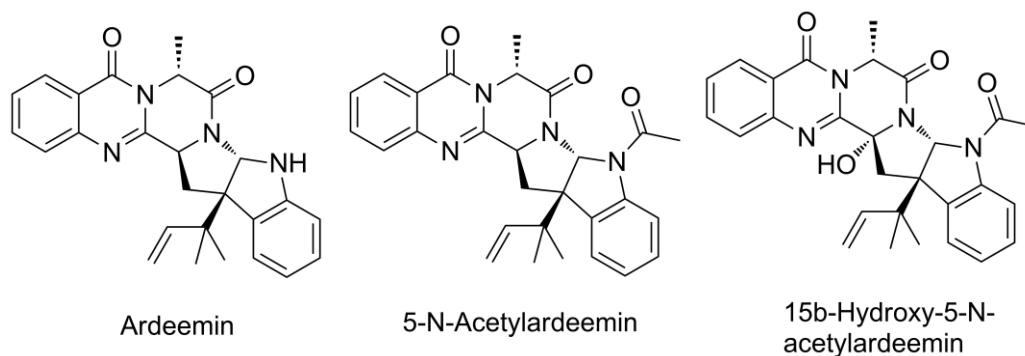

Figure 31. Structures of ardeemins

The hexacyclic ardeemins (**Figure 31**) were isolated from *Aspergillus fischeri* on the basis of their ability to reverse multidrug resistance phenotypes both in isolated cells and in mammary carcinoma xenografts [196, 197] by binding to P-glycoprotein export pumps [198, 199]. The ardeemin name comes from this property: the <u>a</u>bility to <u>r</u>everse <u>d</u>rug insensitivity. Ardeemins were shown to result in sensitization to vinca antitumor alkaloids by up to 700 fold [200]. Danishefsky and colleagues have led synthetic efforts towards the ardeemins and termed them "reverse prenyl hexahydropyrrolo[2,3-b]indole alkaloids" [201].

Inspection of the ardeemin scaffold indicates a tripeptide origin, with the same three building blocks as the fumiquinazoline system; anthranilate, alanine, and tryptophan, but the connectivity and the tailoring enzyme operations must be distinct from the fumiquinazoline system. After construction of the core structure, a subsequent intramolecular, complexity-building, cyclization should be mediated in ardeemin by prenylation with reverse regiochemistry at the β-carbon of the pyrrole ring of a Trp moiety. The immediate precursor in such a ring-

closing prenylation should be a 6,6,6-tricyclic pyrazinoquinazolinone, related to but distinct from fumiquinazoline F (termed here ardeemin fumiquinazoline = ardeemin FQ) by different placement of the Ala- and Trp-derived side chains on the same tricyclic scaffold. The distinction could arise from the putative ardeemin-forming NRPS acting in the order, Ant-Ala-Trp, rather than the order Ant-Trp-Ala seen in the fumiquinazoline and tryptoqualanine pathways (**Figure 31 and Figure 32**) [38, 143, 179] .
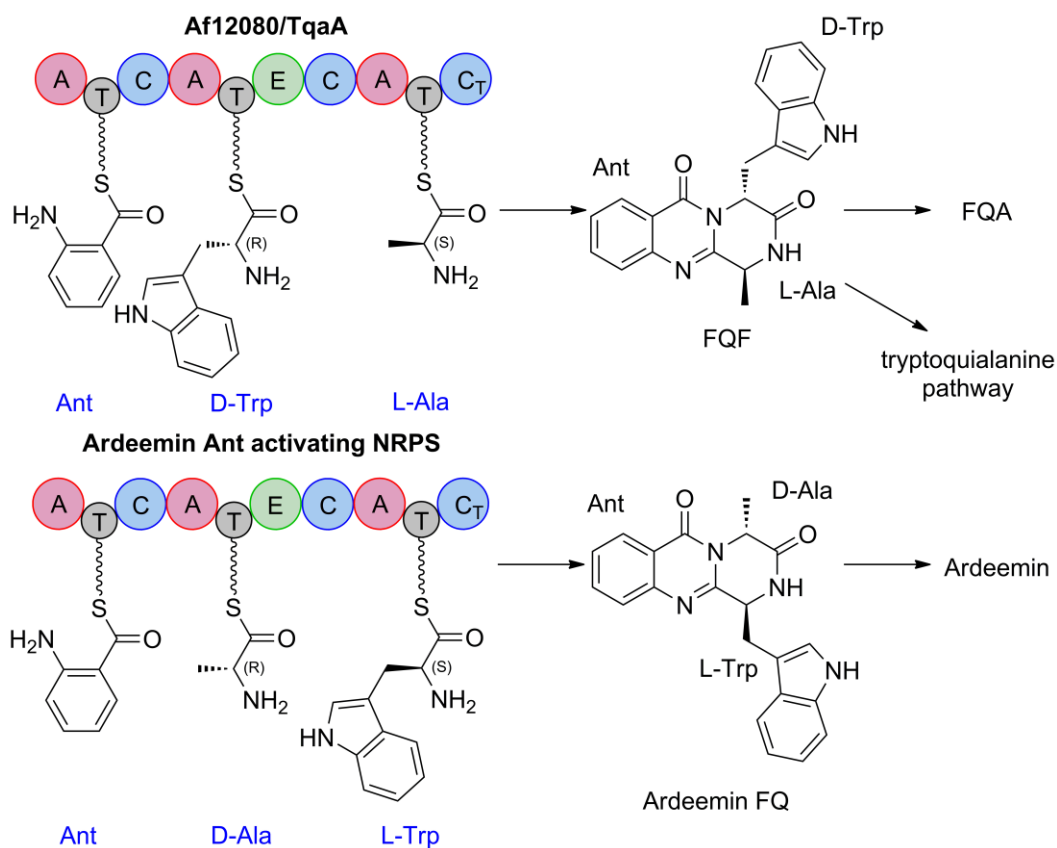

Figure 32. Differentiation in NRPS mediated biosynthesis of FQF and Ardeemin FQ : formation of regioisomeric quinazolinones by swapping the order of A domains

In this study we have validated these predictions by the identification of the ardeemin biosynthetic gene cluster in *A. fischeri*, detection of the ardeemin tricyclic intermediate from a

trimodular NRPS enzyme, and generation of hexacyclic ardeemin by the action of a single prenyltransferase on the tricyclic ardeemin FQ intermediate.

2.5.2 Results and Discussions

Bioinformatic Prediction of Ardeemin Biosynthetic Gene Cluster

We anticipated the ardeemin cluster (*ard*) should contain one gene encoding a trimodular NRPS, whose first adenylation domain should be specific for activation of anthranilate [38]. We also anticipated there might be an associated prenyltransferase. One candidate cluster in *A. fischeri* fulfills these criteria. As shown in **Figure 33A** there is a predicted trimodular NRPS with the domains A-T-C-A-T-E-C-A-T-C_T that we

Figure 33. Identificaiton of the ardeemin (*ard*) gene cluster. (A) Organization of the *ard* cluster; (B) HPLC analysis of metabolites produced by *A. fischeri* strains (1) WT *A. fischeri*; (2) Ardeemin FQ standard; (3) *A. fischeri* with ArdA KO (4) *A. fischeri* with ArdB KO

have termed ArdA (A= adenylation, C = condensation, T = thiolation, E = epimerization domain and C_T is a terminal cyclization domain [175, 179]). Based on the amino acid recognition code we have previously deciphered [38] for the anthranilate-activating NRPS initiation modules for aszonalenein, for fumiquinazoline F, and for tryptoquialanine [38, 143], the first module of ArdA will activate anthranilate. Module 2 (C-A-T-E) has a predicted epimerization domain; just what is required for the incorporation of D-Ala into the ardeemin backbone. Module 3 would then activate and incorporate L-Trp.
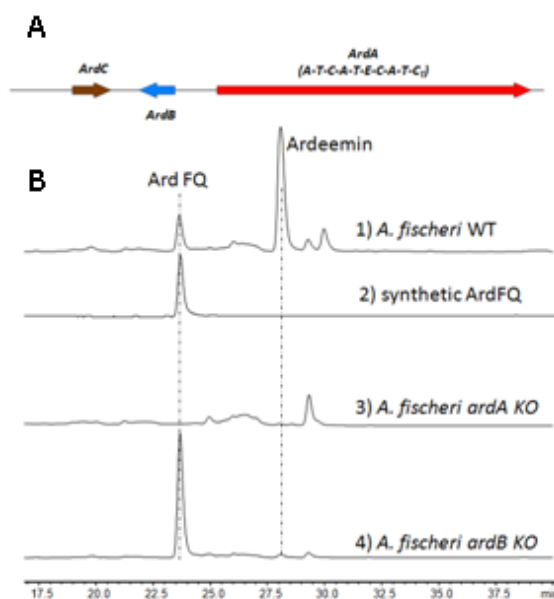
110

Immediately upstream of the *ardA* gene is an ORF we have labeled *ardB*, encoding a predicted dimethylallyl diphosphate-utilizing enzyme, a presumed prenyltransferase. Also, just upstream of this ORF is a predicted acetyltransferase, encoded by *ardC*, in line with the observations of *N*-acetylardeemin as part of the suite of *A. fischeri* natural products (**Figure 31**) [197]. A comparable pair of prenyltransferase (AnaPT, 31% identity/49% similarity to ArdB) and acetyltransferase (AnaAT, 35% identity/53% similarity to ArdB) enzymes accompanies the bimodular NRPS in the acetylaszonalenin biosynthetic gene cluster [26]. Notably however, the examination of this cluster provides no explanation of origin of the other member of the ardeemin suite of natural products, 15b-hydroxy-5-n-acetyl ardeemin. There is no candidate enzyme present in or adjacent to this cluster to account for the incorporation of the hydroxyl group.

Genetic Deletion of *ardA* and *ardB* Validates their Roles in Ardeemin Biosynthesis.

Genetic validation of *ardABC* as the ardeemin biosynthetic gene cluster was achieved by gene deletions and examination of the effect on secreted metabolites. As shown in **Figure 33B** trace 1, extracted metabolites from the parent *A. fischeri* culture yields UV detectable peaks with the masses of 5-N-acetylardeemin (calculated = 469.2234; observed = 469.2232), ardeemin (calculated = 427.2129; observed = 427.2130) and also of the predicted ardeemin FQ (calculated = 359.1503; observed = 359.1506).

Disruption of the *ardA* gene by insertion of the *bar* marker, which confers resistance to glufosinate [202], resulted in loss of both metabolites (**Figure 33B** trace 3). Thus, the NRPS enzyme ArdA operates at the earliest stage of ardeemin assembly. A similar insertion into the *ardB* gene resulted in loss of mature ardeemin but accumulation of ardeemin FQ (**Figure 33B**

trace 4), in accord with our assignment of ArdB as a prenyltransferase that draws ardeemin FQ forward to ardeemin. These knockout results support a simple two-enzyme pathway, ArdA followed by ArdB, to produce the hexacyclic scaffold of ardeemin from the three amino acids Ant, D-Ala and L-Trp.

Heterologous Expression of the ArdA Protein in yeast: *in vivo* and *in vitro* findings.
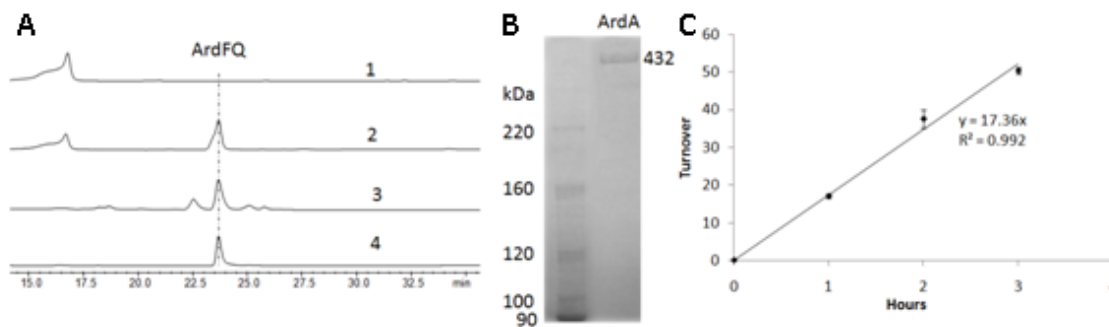

Figure 34. Characterization of ArdA. A) *In vivo* and *in vitro* reconstitution of ArdA. Shown are HPLC analyses (λ = 292 nm) of compounds from extraction of reaction mixtures containing 2 mM of the amino acid building blocks and (1) no enzyme, (2) 10 µM TqaA. Trace 3) is shown is HPLC analysis (λ = 292 nm) of metabolites extracted from 3-d cultures of BJ5464-NpgA expressing ArdA. Trace 4 is the ArdFQ synthetic standard. B) SDS-PAGE of the heterologously expressed ArdA protein (432 kDa). C) HPLC-based time course study of ArdA in the turnover of ArdFQ.

The *ardA* gene, encoding the three module 432 kDa ArdA protein was cloned in five pieces and assembled using a modified yeast *in vivo* recombination method as previous described, with removal of the one detected intron, to yield an N-terminal FLAG-tagged protein for expression in the vacuole protease deficient *Saccharomyces cerevisiae* strain BJ5464-NpgA that we have previously used for cloning the comparably sized trimodular TqaA [179] (**Figure 34A**). Yeast cells expressing ArdA indeed generated the tricyclic ardeemin FQ metabolite as judged by LCMS analysis (**Figure 34B**, trace 1). Purification of the tagged ArdA from the yeast cells yields a soluble high molecular weight (~430 kDa) protein at ~1 mg/L (**Figure 34B**) that is sufficient to convert Ant, L-Ala and L-Trp into ardeemin FQ, clearly distinct from the fumiquinazoline F

(FQF) isomer (**Figure 34A**, trace 2). Rate data for purified ArdA indicate linear production over a 3-hour incubation period at a turnover number of 17 ± 0.96 hr^{-1} (**Figure 34C**). This is comparable to the turnover number of 23 hr^{-1} for TqaA as it produces the regioisomeric FQF [179].

Intriguingly, the trimodular NRPS ArdA has the same domain organization as TqaA and Afl2050 from the tryptoqualanine [143] and fumiquinazoline clusters [38], respectively. In those two NRPS trimodular cases genetics and biochemistry have validated the activation of anthranilate, L-Trp, and L-Ala; where the specificity of the three modules in TqaA and Afl2080 is Ant$_1$-L-Trp$_2$-L-Ala$_3$, and the resulting product is the tricyclic fumiquinazoline F [38, 143, 179]. In contrast, for ArdA it must be Ant$_1$-L-Ala$_2$-L-Trp$_3$; seemingly resulting from a functional swap in the order/specificity of module 2 and module 3. Unfortunately, examination of the amino acid recognition 'code' of A domains from module 2 and 3 of TqaA and ArdB reveals little insight in to any shared genetic heritage of these domains or a clear indication of how this functional swap may have come about. Notably, all three of these trimodular NRPSs have four domains in module 2 (C_2-A_2T_2-E), where E is an epimerization domain. In agreement with the placement of that epimerization domain both residues activated by module 2 undergo epimerization, resulting in the incorporation of D-Trp$_2$ in FQF and D-Ala$_2$ in ardeemin FQ. All three NRPS enzymes end with a terminal C_T domain in module 3. We have shown for TqaA that this terminal C_T domain is required for cyclization/release of the tethered tripeptidyl-*S*-NRPS intermediate [179]. For ArdA that intermediate would be the linear tripeptidyl thioester, covalently tethered as Ant$_1$-L-Ala$_2$-L-Trp$_3$-*S*-T domain (T_3). **Figure 32** shows the two distinct tethered tripeptidyl thioesters and compares their cyclization to the regioisomers FQF and ardeemin FQ by their respective

113

trimodular NRPS enzymes, TqaA and ArdA. We propose that cyclization to give ardeemin FQ and FQF is mediated by an analogous series of events. Firstly, attack by the free amine of Ant_1 on the thioester carbonyl of the corresponding (isomeric) linear tripeptide yields a 6,10 bicyclic nascent product. These bicyclic intermediates then undergo a transannular cyclodehydration to generate the tricyclic quinazolinedione framework characteristic of FQF and ardeemin FQ.

2.5.3 Conclusions

Ardeemins are hexacyclic peptidyl alkaloids isolated from *Aspergillus fischeri* as agents that block efflux of anticancer drugs by (MultiDrug Resistance) MDR export pumps. To evaluate the biosynthetic logic and enzymatic machinery for ardeemin framework assembly, we sequenced the *A. fischeri* genome and identified the *ardABC* gene cluster. Through both genetic deletions and biochemical characterizations of purified ArdA and ArdB we show this ArdAB enzyme pair is sufficient to convert anthranilate (Ant), L-Ala and L-Trp to ardeemin. ArdA is a 430 kDa trimodular nonribosomal peptide synthase (NRPS) that converts the three building blocks into a fumiquinazoline (FQ) regioisomer termed ardeemin FQ. ArdB is a prenyltransferase that takes tricyclic ardeemin FQ and dimethylallyl diphosphate to the hexacyclic ardeemin scaffold via prenylation at C_2 of the Trp-derived indole moiety with intramolecular capture by an amide NH of the fumiquinazoline ring. The two-enzyme ArdAB pathway reveals remarkable efficiency in construction of the hexacyclic peptidyl alkaloid scaffold.

Therefore, we can see that nature generates two dramatically different architecturally complex products, ardeemin and FQA, from related isomeric precursors using comparable enzymatic logic, activation by incoming electrophile (flavin hydroperoxide or allylic cation)

114

setting up for a subsequent nucleophilic attack on C_2 of the indole substituent. The indole ring prenylation strategy that sets up the electrophilic species captured intramolecularly by an amide nitrogen during the biosynthesis of ardeemin is also exemplified in several other peptide contexts. These include the formation of the tricyclic pyrroloindole moiety in the cyclic hexadepsipeptide kutzneride antifungal agents [203, 204] and also in Bacillus competence factor peptides [205-208]. In the case of ardeemin the amide nitrogen responsible for the intramolecular capture of the prenylated indole is part of a tricyclic moiety so the three ring (pyrazinoquinazolinone) unit joins the two ring indole unit by forming a bridging five ring. The net result is dramatic scaffold complexity and architectural constraint generation, all from the action of only two enzymes. Comparable logic in the fungal pathway to aszonalenin also employs only two enzymes, the first a bimodular (rather than trimodular) NRPS, using anthranilate and Trp, to yield the 6, 7-benzodiapinedione [26]. The second prenylating enzyme therefore generates a 6-7-5-5-6 pentacyclic array in that case rather than a hexacyclic framework.

The construction of scaffold complexity in the ardeemin peptidyl alkaloid is remarkably efficient. Only two enzymes are in play to convert three amino acid building blocks into the highly constrained fused hexacyclic product architecture. The first enzyme uses anthranilate as a starter and that planar β-amino acid creates architecture prone to cyclization that gives the 6-6-6 ardeemin FQ-type scaffold. Nature has performed a functional swap between the FQ and ardeemin FQ trimodular NRPS assemblies, switching Ala and Trp between positions two and three to give regioisomeric tricyclic pyrazinoquinazilonedione products, FQ and ardeemin FQ, that are then vectored down distinct maturation pathways. These NRPS assembly lines are remarkable pathway-initiating catalysts. When paired with either indole oxygenase and/or indole

115

prenyltransferase tailoring enzymes, they form branch points to distinct complex peptidyl indole alkaloid molecular frameworks in short two enzyme pathways. These anthranilate-derived tripeptides have been morphed into a range of multicyclic architectures that populate interesting chemical and biological space [209]. Mixing and matching distinct anthranilate-initiating NRPS assembly lines with tailoring enzymes that perform electrophilic chemistry on a partner tryptophan side chain may allow engineering of new biologically active architecturally constrained scaffolds.

2.5.4 Materials and Methods

Growth of *A. fischeri* (WT and mutants) and Analysis of Metabolite Production. *A. fischeri* cultures were grown in liquid culture in both shake flasks and stationary cultures using glucose minimal media (GMM). Shake flask cultures were prepared by inoculating 50 mL of GMM in 250 mL baffled flasks with A. fischeri spore stock (see supporting information), the resulting cultures were grown at 37 °C for 5 days. Stationary cultures were prepared by inoculating 10 mL of GMM in 60 x 15mm Petri dishes, the resulting cultures were grown at 28 °C for 4 ~ 5 days. Media from cultures was then extracted with an equal portion of ethyl acetate. Ethyl acetate was concentrated *in vacuo* to dryness and the residue resuspended in 50 % acetonitrile/water and filtered for analysis by LC-MS.

Cloning of ArdA. Cloning of the intact *ardA* gene into the expression vector was performed using the modified yeast-based homologous recombination method as described previously [179]. The gene encoding ArdA was divided into four pieces (P1–P4) with a maximum size of ~4 kb. The only intron (475-547 base pairs (bp)) of the ArdA-encoding gene within P1 was removed by PCR with reverse transcription (RT–PCR). P2 to P4 were PCR amplified from the genomic

116

DNA of *A. fisheri*. Each successive piece was designed to overlap (35–40 bp) with the two flanking ones, and the 5′ end of P1 and 3′ end of P4 overlap with the expression vector. The vector fragment was generated by digesting pXW55 (an expression vector with an engineered Flag tag on its N terminus and a His$_6$ tag on its C terminus) with SpeI and PmlI [210]. The individual pieces (1 μg per piece) and the vector (~0.5 μg) were transformed together using the EasyComp Transformation Kit (Invitrogen) into *S. cerevisiae* strain BJ5464-NpgA to allow assembly of the entire gene *in vivo*. The assembled plasmid was miniprepped from *S. cerevisiae* using a yeast plasmid miniprep II kit (Zymo) and was further verified by PCR and restriction enzyme digestion.

In vitro assay of ArdA. To test the *in vitro* activities of ArdA, 2 mM of each amino acid building block, 6 mM ATP, 4 mM MgSO$_4$ and 10 μM of purified ArdA in 50 mM Tris-HCl buffer (pH 7.5) to a total volume of 100 μl. After a 12-h incubation, reaction was quenched and extracted with 1 ml of ethyl acetate. The organic layer was dried, redissolved in methanol and injected onto a Shimadzu 2010 EV LC mass spectrometer using positive and negative electrospray ionization and a Phenomenex Luna 5 μm, 2.0 mm × 100 mm C18 reverse-phase column. Samples were separated on a linear gradient of 5–95% (v/v) CH$_3$CN in water (0.1% (v/v) formic acid) for 30 min at a flow rate of 0.1 ml min^{-1} followed by isocratic 95% (v/v) CN in water (0.1% (v/v) formic acid) for 15 min. Under these conditions Ardeemin FQ was eluted at 23.5 min.

HPLC-based Time Course Study of Product Formation by ArdA. Reactions (500 μl) contained 1 μM ArdA, 3 mM ATP, 2 mM MgCl$_2$, and 1 mM amino acid substrates (Ant, l-Trp, and l-Ala) in 50 mM Tris-HCl buffer (pH 7.5). Reactions were incubated at 25 °C and 100 μl

aliquots were taken at 1, 2, 3 and 4 hours after addition of enzyme and extracted with 1 ml of ethyl acetate. Only data points within the linear range were used to calculate the initial product formation rates. The fitted slope of the linear portion of initial turnover is 17.36 hr-1, with R^2=0.992. Data represent mean values ± s.d. The organic layer was dried and redissolved in 100 μl methanol and 20 μl samples were injected for LC-MS analyses. Integration of the Ardeemin FQ peak (at 292 nm) was used to generate a plot of product peak area vs. time in order to approximate enzymatic rate. Initial rate data (obtained as integration area per hour) was converted to μM per hour using standard curves generated from 20 μl injections of Ardeemin FQ sample of known concentration.

2.6　　Directed Evolution and Structural Characterization of the Simvastatin Synthetase

A version of this section was published as: Gao, X., Xie, X., Pashkov, I., Sawaya, R. M., Laidmen, J., Zhang, W., Cacho, R., Yeates, T. O., Tang, Y. "Directed evolution and structural characterization of a simvastatin synthase." *Chem. Biol.*, 2009, 16, 1064-1074. Copyright © 2009, Elsevier Ltd.

2.6.1　Introduction

LovD is an acyltransferase found in *Aspergillus terreus* and is responsible for converting the inactive precursor monacolin J acid (MJA) into the cholesterol-lowering drug lovastatin (LV, acid form lovastatin acid: LVA) via acylation of the α-*S*-methylbutyrate side chain [83, 85] (**Figure 35**). The importance of the hydrophobic α-*S*-methylbutyryl side chain for binding of LVA to HMG-CoA reductase has been structurally confirmed [211]. Chemical modification of the LV side chain to α,α-dimethylbutyrate yielded the semisynthetic derivative simvastatin (SV, acid form simvastatin acid: SVA), which is the active pharmaceutical ingredient in the

118

blockbuster drug Zocor[®] [212]. Semisynthesis of SV from LV is a multiple-step chemical process and is therefore an intensely pursued target for devising an efficient biocatalytic approach [213, 214]. As a result, LovD is a prime candidate to serve as such a biocatalyst.

LovD is a 413-amino acid protein predicted to have an α/β hydrolase fold based on primary sequence analysis [83]. Among enzymes of known structure that are homologous to LovD is cephalosporin esterase, EstB (PDB ID 1CI9, 26% sequence identity) from *Burkholderia gladioli* [215]. The likely general base Tyr188, as well as a conserved SXXK patch that contains the active site nucleophile Ser76, were indicated through alignment of LovD with EstB (**Figure 36A**) [216]. During LVA biosynthesis, the α-*S*-methylbutyrate side chain is synthesized by the lovastatin diketide synthase (LDKS) LovF, and is then transferred by LovD regioselectively to the C8 hydroxyl of MJA via an unprecedented polyketide offloading mechanism [85]. The protein-protein interaction between LovD and the acyl carrier protein (ACP) domain of LovF facilitates this highly efficient tailoring reaction in *A. terreus*. We have previously explored the substrate promiscuity of LovD and have shown that it can also synthesize SVA by using the small molecule substrate α-dimethylbutyryl-*S*-methyl-mercaptoproprionate (DMB-SMMP) as an acyl donor [87] (**Figure 35**). Using *Escherichia coli* as an expression host, a whole-cell biocatalytic platform for converting MJA to SVA was established that can produce SVA with low throughput [87]. However, as with many enzymes that have been removed from their natural context, LovD is catalytically suboptimal as a biocatalyst and suffers from poor thermal stability [217]. The catalytic activity of SVA synthesis using DMB-SMMP is attenuated ~1,300 fold when compared to the natural substrate attached to LovF [149], indicating there is ample opportunity for optimization by protein engineering efforts. Furthermore, the structural basis of

119

LovD function and substrate selection had not been elucidated, limiting our ability to rationally optimize the binding of the unnatural dimethylbutyryl substrate and improve LovD efficiency as a SV synthase.
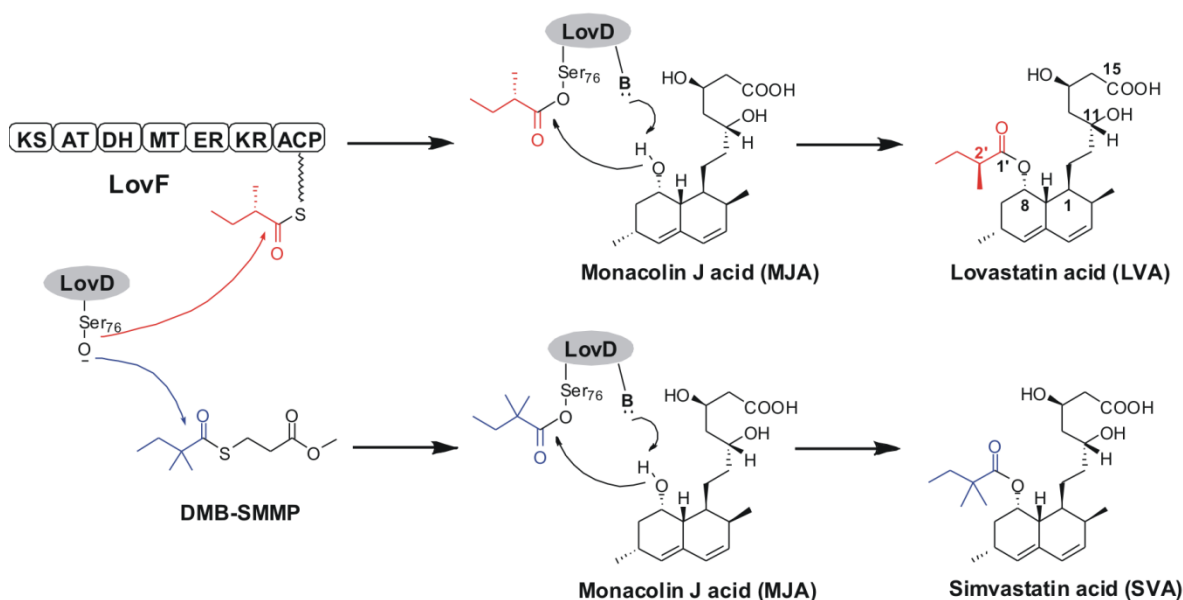

Figure 35. Reactions catalyzed by LovD. LovD is responsible for converting MJA into LVA via acylation of the α-S-methylbutyrate side chain and can also synthesize SVA using DMB-SMMP as an acyl donor.

In this article, we employed directed protein evolution [218] to improve the SV synthase activity of LovD. After seven rounds of screening, LovD mutants with significantly improved catalytic activities and higher thermal stability were isolated. In parallel, seven X-ray crystal structures including the parent LovD G0, an improved mutant G5, and the co-crystal structures of G5 with MJA, LVA and SVA were obtained. The crystal structures provide atomic resolution details regarding the mechanism of catalysis, substrate and product binding, protein-protein interactions with LovF, and a likely explanation for the effects of beneficial mutations on catalysis.
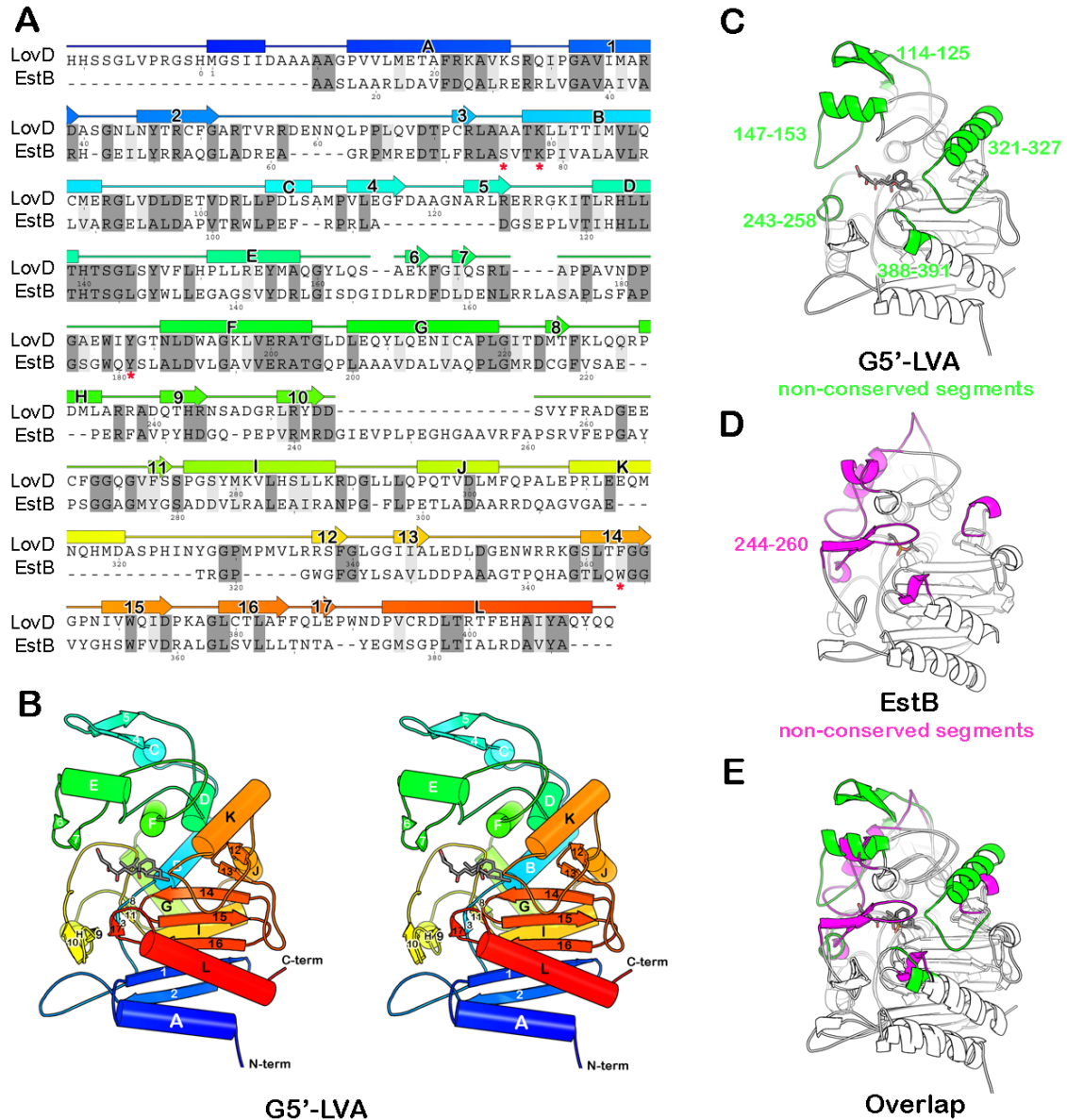
2.6.2   Results and Discussions

Figure 36. The crystal structure of LovD and its relationship to EstB. (A) A structure-based sequence alignment between LovD and EstB. Secondary structure elements assigned from the structure of LovD are shown above the sequence. The colors are ramped from blue at the N-terminus to red at the C-terminus. The active site residues in EstB are indicated by an asterisk "*" below the amino acid. (B) A ribbon diagram showing the G5'-LVA complex. (C) Structure of LovD. Highlighted in green are segments that are not conserved in EstB. These five loops project around the circumference of the active site like the fingers in a catcher's mitt. (D) Structure of EstB. Highlighted in magenta are segments that are not conserved in LovD. (E) The overlay of LovD and EstB structures. Notably absent from LovD is a loop that covers the active site in EstB (residues 244-260).

Development of an Agar-based Diffusion Screening Method

121

We developed a high throughput screening method to assay for *E. coli* expressing LovD mutants with improved properties in the synthesis of SVA from MJA and DMB-SMMP. The assay relied on the growth inhibition of *Neurospora crassa* by statins, a property that was previously exploited in the screening of high LV producing *A. terreus* strains [219]. We found that SVA can inhibit the growth of *N. crassa* at sub-microgram quantities, while inhibition by JA requires hundred milligram quantities. To demonstrate the sensitivity and feasibility of the assay, an *E. coli* culture expressing wild type LovD was supplied with 10 mM MJA and 15 mM DMB-SMMP. At different time points, 2 µL aliquots were directly spotted on a Sabouraud's dextrose agar (SDA) plate embedded with *N. crassa* at a density of 0.3~0.5 x 10^8 spores/L. After 16 hours of incubation at 30ºC, different inhibition zones were observed for samples containing different degrees of conversion of MJA to SVA (as verified by HPLC) (**Figure 37A**). Based on this screening strategy, any significant contribution to whole cell LovD activity, such as improvements in solubility, catalytic efficiency and stability can lead to a detectable phenotypical change.

Screening of LovD Variants with Enhanced Whole Cell Activities

The starting LovD (generation zero or G0) used in directed evolution is the previously characterized double mutant C40A/C60N [220]. G0 was rationally engineered to be less prone to disulfide-mediated aggregation and was used for crystallization studies [220]. The mutant libraries were created by either saturation mutagenesis or error-prone Polymerase Chain Reaction (ep-PCR) that generated an average of 2.5 amino acid changes per round [221]. During each round of screening, the mutant library was ligated into pET28(a) and electroplated into YT2 competent cells [88]. The individual mutants were cultured in 96-well plates, followed by

induction of LovD expression, addition of MJA and DMB-SMMP, and spotting onto *N. crassa* embedded plates. **Table 3** shows the gradual improvement in whole cell activity obtained following four rounds of ep-PCR (G1, G2, G4, G6), one round of saturated mutagenesis (G7), and two iterations of combining individual beneficial mutations (G3, G5), with the best mutant G7 displaying ~11 fold increase in whole cell activity as a SV synthase compared to G0 (**Figure 37B**). The mutant G2.1 contains amino acid changes at D12G and G275S. Construction of the corresponding single mutants using G1 as template showed that D12G alone had a large negative effect, while G275S alone had a weak positive effect compared to G1. This result suggests the two mutations in G2.1 act synergistically to enhance LovD activities. Combination of mutations from G2.1 and the A190T mutation in G2.2, which was recovered from the same round of ep-PCR, yielded the next generation mutant G3. Ep-PCR using G3 as a template yielded G4.1 and G4.2, each containing a different double mutation combination of A10V/K26E and H161Y/K227R, respectively. Site directed mutagenesis confirmed that both A10V and K227R had negative effects on the activities of LovD. Removal of these mutations and combination of K26E and H161Y yielded an improved mutant G5, which was ~6-fold improved in whole cell activity compared to G0. At this point, structural studies were performed on the G5 mutant to provide insights into the accumulated beneficial mutations. In parallel, an additional round of ep-PCR afforded G6 that contained the beneficial mutations V334D and L361M. Saturation mutagenesis was employed to optimize the combined effects of mutations at positions 334 and 361. The best mutant recovered was G7, of which the whole cell activity was increased an additional 20%. Surprisingly, we found that while position 334 was altered to phenylalanine, the previously deemed beneficial L361M mutation reverted back to leucine. Site-directed mutation

of L361M in G7 confirmed that leucine was indeed the more favorable residue in the context of the V334F mutation.
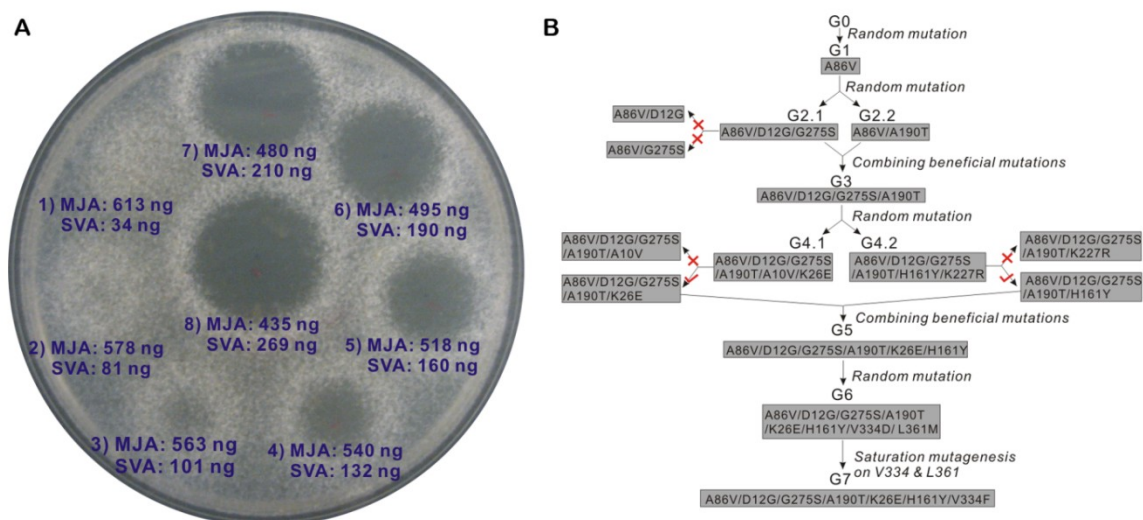

Figure 37. Directed evolution of LovD as a simvastatin synthase. (A) An agar diffusion-based assay was used to quantify the amount of SVA in the whole-cell activity experiments. *N. crassa* was embedded in the agar prior to spotting the reaction mixture. The numbers (1-8) designate different incubation times of 2.5, 4, 5.5, 7, 8, 9, 10, 12 hours following addition of MJA and DMB-SMMP to *E. coli* expressing wild type LovD. (B) Directed evolution of LovD mutants towards higher whole cell activities. There are a total of seven generations of LovD mutants. Four generations were derived from random mutagenesis including G1, G2.1, G2.2, G4.1, G4.2, and G6. Two generations were derived from combination of beneficial mutations from previous generation mutants including G3 and G5. G7 was derived though saturated mutagenesis of G6 at positions V334 and L361. All mutants with two amino acid changes were subjected to site directed mutagenesis to determine beneficial or deleterious mutations (G2.1, G4.1 and G4.2). (×) indicates that the mutant had lower whole cell activity comparing to the previous generation. (√) indicates that the mutant had higher whole cell activity comparing to the previous generation.

In Vitro Characterization of LovD Variants

To dissect the contributions that led to the increases in whole cell activity, kinetic parameters (k_{cat}, K_M), soluble protein levels and thermal stability of all the improved mutants were characterized and listed in **Table 3**. The binding affinities (K_M) of LovD mutants toward MJA and DMB-SMMP were each within a narrow range (0.7 mM to 0.9 mM for MJA and 0.6 mM to 0.7 mM for DMB-SMMP). The lack of improvement in K_M towards either substrate is not surprising considering the high concentrations of substrates used in the screening assay (~10 K_M of MJA and DMB-SMMP). The observed improvements in whole cell activity are mainly

due to increases in the k_{cat} of the mutants and levels of soluble proteins. The k_{cat} and soluble protein levels were simultaneously increased ~3 fold and ~1.5 fold from G1 to G3, respectively. Impressively, the protein expression levels of G3 reached 205 mg/L. In contrast, improvements in k_{cat} were the sole contribution to the increases in whole cell activities from round 4 to round 7. Most notably, a single V334F mutation from G5 to G7 nearly doubled the catalytic turnover rate. We also observed a gradual increase in the melting temperature (T_m) of the LovD mutants from 39.5°C to 48.5°C by using circular dichroism. The increases in thermal stability were also reflected in the whole cell activities of the mutants when expressed at elevated temperatures (**Figure S49**). Whereas the G0-G3 mutants have no detectable activity when expressed at 32°C, the later generation mutants retained significant SV synthase activities, with the G7 mutant exhibiting comparable activity to that of G0 at 25°C. Furthermore, the G7 mutant remained active even when expressed at 37°C. The increased thermal stabilities of the mutants have important practical implications in using LovD as a biocatalyst for SV semisynthesis.

To examine the activities of the mutants towards synthesis of the natural biological product LVA, we performed kinetic assays using □-methylbutyryl-SMMP (MB-SMMP) and MJA. A similar trend in the improvements of k_{cat} toward LVA synthesis was observed (**Figure S50**), indicating the LovD relative substrate specificity towards the acyl group (either MB or DMB) has not changed. Interestingly, when LovF was used in the kinetic assay for LVA synthesis (**Figure 35**), we observed a progressive loss of activities of the LovD mutants (**Figure S50**). The G7 mutant exhibited a 27-fold decrease in activities towards LovF compared to the G0 parent, most likely attributed to the deterioration of the required protein-protein interactions for catalysis [149]. Therefore, the mutations accumulated during directed evolution may have

gradually altered the conformation of LovD to impair its communication with LovF, while not affecting binding of the SMMP-bound acyl group.

Table 3. Amino acid substitutions and characterization of LovD variants

	Mutations	Whole cell Activity[*]	k_{cat} (min^{-1})	K_M of MJA (mM)[†]	K_M of DMB-SMMP (mM^{-1})[‡]	Soluble protein (mg/L)[§]	T_m(°C)[¶]
G0		1	0.66±0.03	0.77±0.17	0.67±0.12	138±11	39.5±0.4
G1	A86V	1.2	0.79±0.03	0.74±0.16	0.66±0.19	140±5.4	41±0.7
G2.1	A86V D12G G275S	1.9	1.14±0.03	0.91±0.17	0.62±0.10	184±8.7	40.5±0.4
G2.2	A86V A190T	1.8	1.20±0.09	0.74±0.21	0.69±0.17	168±17	41±0.4
G3	A86V D12G A190T G275S	3.6	1.86±0.09	0.77±0.11	0.70±0.19	205±23	41±0.4
G4.1	A86V D12G A190T G275S A10V K26E	4.8	2.13±0.03	0.70±0.24	0.66±0.16	183±18	43.5±0.7
G4.2	A86V D12G A190T G275S H161Y K227R	5.2	2.16±0.12	0.80±0.24	0.64±0.16	221±9.3	42.5±1.9
G5	A86V D12G A190T G275S K26E H161Y	6.4	2.61±0.03	0.74±0.03	0.69±0.14	206±5.7	46.5±0.4
G6	A86V D12G A190T G275S K26E H161Y V334D L361M	9.3	3.30±0.06	0.70±0.07	0.63±0.15	212±3.9	47±0.1
G7	A86V D12G A190T G275S K26E H161Y V334F	11.2	4.80±0.06	0.70±0.04	0.69±0.17	214±6.3	48.5±0.7

[*]The whole-cell activity of LovD mutants are compared to G0, which has a conversion rate of 1.7 mM/hr and is normalized to 1. [†]K_M of MJA is derived at 25°C when DMB-SMMP is fixed at 2 mM. [‡]K_M of DMB-SMMP is derived when MJA is fixed at 2 mM. [§]The amounts of soluble proteins are measured from purified protein levels. [¶]Tm is measure by circular dichroism. All results represent mean values of triplicate determinations ± SD.

Overall Structure of LovD

Seven LovD crystal structures were determined to help illuminate the mechanism of the LovD-catalyzed reaction and possible basis for improved catalysis. These include (1) G0, (2) selenomethionyl G0 (G0-Semet), (3) the improved mutant G5, (4) G5 in complex with substrate

126

Table 4. Statistics of X-ray data collection and atomic refinement (numbers in parentheses refer to the outer shell of data).

Protein	G0	SeMet G0	G5	G5	G5'	G5'	G5'
Ligand	none	none	none	MJA	MJA	SVA	LVA
.Data collection							
Space group	P1	C2	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions							
a, b, c (Å)	56.7,79.7,104.1	209.5,85.2,104.0	58.2, 75.0, 131.6	58.5, 75.2, 133.5	58.0, 75.2, 132.7	58.5, 75.1, 131.8	59.0, 75.2, 131.7
□□□□□□ (°)	94.1,91.6,106.8	90.0,117.5,90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
protomers/asymmetric unit	4	4	1	1	1	1	1
Resolution (Å)	100.00-3.40	80.00-2.50	90.00-2.00	60.00-2.00	60.00-2.05	60.00-2.00	60.00-2.00
	(3.66-3.40)	(2.59-2.50)	(2.07-2.00)	(2.07-2.00)	(2.12-2.05)	(2.07-2.00)	(2.07-2.00)
R_{merge}(%)	20.8 (43.4)	11.8 (45.3)	4.9 (37.8)	15.6 (39.6)	6.8 (40.2)	6.9 (31.4)	11.8 (39.6)
I / □I	5.1 (2.9)	10.2 (1.7)	47.2 (6.5)	13.0 (4.8)	38.3 (6.6)	19.5 (3.9)	13.33(4.64)
Completeness (%)	99.3(98.9)	89.9 (46.7)	99.9 (100.0)	99.8 (100.0)	99.9 (99.8)	97.7 (99.6)	98.5 (99.5)
Redundancy	3.7(3.7)	5.3 (3.0)	11.6 (11.7)	6.2 (6.4)	15.4 (12.9)	4.9(4.9)	6.4(6.5)
wavelength (Å)	0.9793	0.9793	0.9794	0.9794	0.9792 &1.5418	0.9717	0.9795
Refinement							
Resolution (Å)	103.7- 3.4	65.9- 2.50	65.80- 2.00	46.18- 2.00	49.86- 2.05	49.51- 2.00	46.42- 2.00
	(3.5- 3.4)	(2.56- 2.50)	(2.05- 2.00)	(2.05- 2.00)	(2.10- 2.05)	(2.05- 2.00)	(2.06- 2.00)
No. reflections	22457 (1669)	47728 (1657)	37539 (2837)	38325 (2910)	35223 (2686)	36938 (2714)	37005 (2671)
R_{work} / R_{free} (%)	23.1/27.5	24.8/29.0	17.9/20.6	17.9/21.3	17.7/20.9	17.6/20.9	16.2/18.9
	(31.9/35.5)	(35.6/40.5)	(20.4/24.4)	(21.5/26.1)	(19.8/23.7)	(23.1/25.1)	(18.3/21.8)
No. atoms							
Protein	12328	12547	3289	3332	3355	3351	3343
Ligand/ion 1	0	65 (sulfate)	13 (PEG)	48 (MJA)	24 (MJA)	31(SVA)	30
Ligand/ion 2	0	0	6 (glycerol)	12 (formate)	8 (dithiothreitol)	0	0
Water	0	21	151	275	157	179	230
B-factors (Å^2)							
Protein	6.3	51.7	34.4	27.6	35.4	32.2	31.4
Ligand/ion 1	n/a	68.0 (SO$_4$)	58.6 (PEG)	58.8 (MJA)	46.5 (MJA)	45.5 (SVA)	38.3 (LVA)
Ligand/ion 1	n/a	n/a	58.2 (glycerol)	38.6 (formate)	74.2 (dithiothreitol)	n/a	n/a
Water	n/a	39.4	36.9	32.8	37.6	36.1	37.2
R.M.S. deviations							
Bond lengths (Å)	0.011	0.006	0.009	0.009	0.008	0.008	0.008
Bond angles (°)	1.3	0.995	1.199	1.230	1.202	1.193	1.193
Ramachandran Plot (%)							
most favored	88.0	90.0	91.2	89.4	89.7	90.3	90.6
additionally allowed	11.4	9.7	8.5	10.3	10.0	9.4	9.1
generously allowed	0.6	0.3	0.3	0.3	0.3	0.3	0.3
disallowed	0.0	0.0	0.0	0.0	0.0	0.0	0.0
PDB ID code	3HL9	3HLB	3HLC	3HLD	3HLE	3HLF	3HLG

$R_{merge}=\Sigma|I-<I>|^2/\Sigma I^2$, where I is the observed intensity. Both summations involve all input reflections for which more than one symmetry equivalent is averaged. $R_{work}=\Sigma||F_o|-|F_c||/\Sigma|F_o|$, where F_o and F_c refer to observed and calculated structure factors, respectively. R_{free} is similar to R_{work}, but is based on a subset of the reflections, which were withheld from refinement for cross validation. Numbers in parentheses refer to the outer shell of data.

MJA, the G5 with S76A active site mutated (called G5') in complex with (5) LVA, (6) SVA, and (7) MJA. The resolution limits of the structures range from 2.5 to 2.0 Å except for G0. The native G0 structure was resolved at 3.4 Å, but was improved to 2.5 Å in the G0-Semet variant. Refinement statistics are provided in **Table 4**.

The crystal structure of LovD G0-Semet revealed a variation of the α/β hydrolase fold [222, 223]. It consists of two domains. The first domain (residues 1-92 and 204-413) is a central seven-stranded antiparallel β-sheet flanked by α-helices on either face (**Figures 36B and S51A**). A *cis*-peptide bond is formed between Glu388 and Pro389, contributing to a kink in the sheet. The second domain is smaller (residues 93-203) and primarily α-helical. A deep and narrow cleft (11 x 6 Å) is formed at the interface between the two domains. At the bottom of the cleft is the catalytic Ser76 that acts as the nucleophile in the acyltransfer reaction.

Comparison between LovD and EstB

Encircling the active site cleft is a broad, ring-shaped ridge, which is absent from the homologous enzyme EstB. Their structures are superimposable with RMS deviation of only 1.5 Å over 270 pairs of α-carbons (about 2/3 of the structure) (**Figure 37E**). The similarity is striking for the core of the two enzymes, but they differ notably in the loops peripheral to the active site, both in size and architecture. In LovD, these loops give the impression of a ring-shaped ridge or catcher's mitt over the active site with fingers composed of five loops: residues 114-125, 147-173, 243-258, 321-327, and 388-391 (**Figure 36C**). The first and last of these loops are longer in LovD than EstB by 11 and 19 residues, respectively. The second loop is displaced 7 Å from the active site compared to EstB, extending the grasp of the "mitt". Most notably absent from the LovD molecule is the 23-residue loop that if present would obstruct the grasp of the mitt and cover the active site entrance (corresponding to residues 244-260 in EstB)

(Figure 36D). The shape and diameter of the ridge surrounding the active site (a circle of 17 Å diameter) satisfies the requirement of accommodating LovD's natural binding partner, the ACP of LovF [149]. Moreover, a positively charged tunnel leading from the active site of LovD and the positively charged ridge surface further suggest that it binds to ACP, the surface of which tends to be negatively charged (**Figure S51B**) [224]. The distance between the rim of LovD and active site Ser76 is ~20 Å, which is roughly the same as the length of the phosphopantetheine (Ppant) arm of the ACP domain of LovF.

Crystal Structures of the Mutant G5

Knowledge of the LovD G0-Semet structure made it possible to locate in three-dimensions the residues that when mutated (in combination) were shown in this study to accelerate the catalytic activity of the LovD G5 mutant, which has a ~4 fold improvement in k_{cat}. These residues are scattered widely over space, forming no mutual contacts. Nor do they share a common physical environment, being located in both buried and solvent exposed regions. Moreover, distances of these residues to the active site Ser76 are relatively large, ranging from 10 to 32 Å for Thr190 and Gly12, respectively (**Figure 38**).
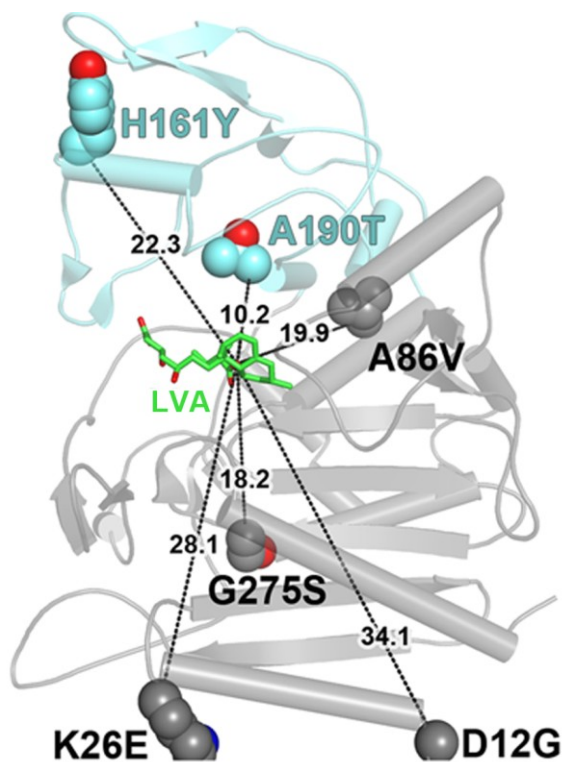

Figure 38. Structure of the G5 mutant provides insight into improved catalysis. Positions of the amino acid changes present in the improved mutant G5, highlighting their generally large distances from the active site. Distances are drawn from the amino acid alpha carbons to the nucleophilic hydroxyl (C8) of LVA (shown in green).

Gly12, respectively (**Figure 38**). The lack of connectivity of the residues to each other and to the active site is a common phenomenon in numerous directed evolution experiments [225-227].

The crystal structure of G5, however, offered evidence to suggest that the increased activity afforded by its six mutations can be attributed to their ability to stabilize a more closed form of the active site cleft. Comparison of the G0-Semet structure with G5 revealed a rotation about the domain-domain hinge of 5°, narrowing the cleft by about 0.5 Å and producing motions up to 3 Å for atoms furthest from the hinge (**Figure 39A**). Subsequent structures of G5 bound to the LVA showed larger movements along the same trajectory arising from a 14° hinge rotation. That observation suggests that the beneficial mutations in the G5 variant help promote a conformational change required for catalysis.
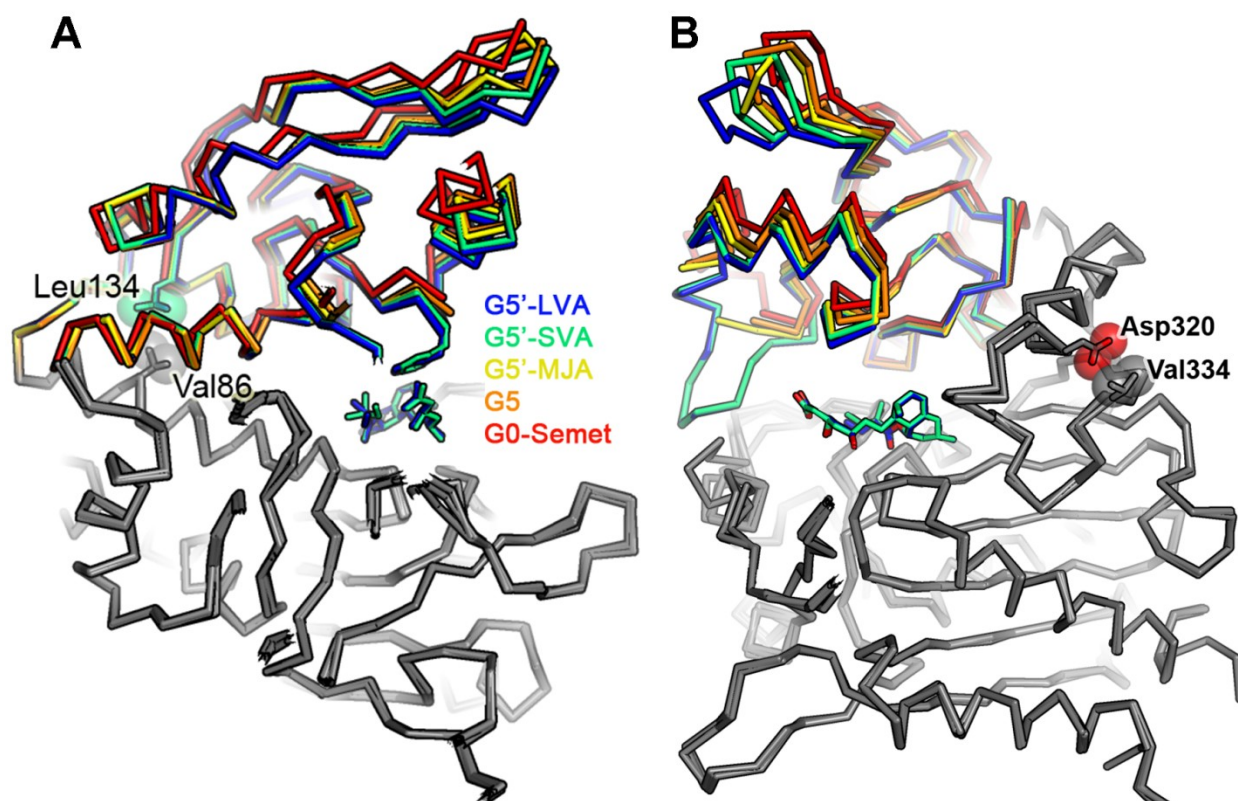
Figure 39. Side view of the overlap of five LovD structures. This representation indicates a hinge rotation between two domains due to G5 mutations and the presence of bound ligands. (A) Two important residues (Val86 and Leu134), which may stabilize closure of the hinge, are shown in spheres. (B) Another two residues (Val334 and Asp320) are directly in contact with each other. Mutation V334D in G6 would result in electrostatic repulsion between Asp334 and Asp320 while mutation V334F in G7 could result in steric clash between the phenyl side chain and Asp320. Both could stabilize closure of the hinge.

Stabilization of the closed conformation of LovD might enhance activity by positioning residues critical for catalysis. For example, when the large domains of G0-Semet, G5, and G5-MJA are superimposed (residues 14-92 and 204-405), it is evident that domain rotation from G0-Semet to G5 closes the gap between the guanido group of Arg173 and the C15 carboxylate of MJA by 0.5 Å . This rotation could be attributed to the G5 mutations (and perhaps to a difference in crystal packing) but not to ligand binding, since the comparison is between two unliganded structures. Ligand binding produces a further rotation from G5 to G5-MJA which closes the gap between Arg173 and MJA by an additional 2.2 Å, so that a hydrogen bond is formed between the two groups. Similarly, G5 mutations bring Phe148 and Tyr188 side chains from the G0-Semet position closer to substrate (G5-MJA), although their motion is smaller since they lie closer to the hinge axis (**Figure S52A**).

The A86V mutation in particular appears responsible for stabilizing closure of the hinge. Its two additional methyl groups buried in the boundary between domains act as a wedge pushing against Leu134 on the distal side of the hinge axis, thereby closing the active site cleft on the proximal side of the hinge axis (**Figure 39A**). The beneficial effects of the K26E and G275S mutations are less obvious. The K26E mutation might improve stability of the enzyme by breaking up a patch of positively charged residues (R22, K23, K26, and R28) on the surface of helix A (**Figure S53A**) [228]. The G275S mutation appears to improve stability of the enzyme by adding a hydrogen bond with the N-terminal end of helix I and decreasing torsional flexibility of the backbone (**Figure S53B**).

Co-crystallization of LovD with MJA, LVA and SVA

Structures of LovD G5' in complex with substrate and products illustrate the mode of binding of these ligands and suggest a catalytic mechanism for acyl transfer. The substrate MJA

binds with its C8-hydroxyl group deep inside the cleft between domains, forming hydrogen bonds with Ser76, Tyr188, and a fortuitously bound formate molecule (used as a cryoprotectant) (Figure 6A). The proximity of the C8-hydroxyl to the Ser76 hydroxyl is consistent with the expectation that both hydroxyl groups initiate a nucleophilic attack on the same acyl group during different steps in the reaction sequence (**Figure S54B**). The two faces of the decalin ring system of MJA are sandwiched between the aromatic rings of Trp390 and Tyr188 (**Figure S51C**). Additional hydrophobic and van der Waals interactions with the edges of the ring system are observed with Phe363, Ile325, Tyr327, Phe148, Ler149, and the peptide planes of Gly364, Gly365, and Gly366 (**Figure S51C**). The hydrophilic tail of MJA (i.e. the C1 substituent on the decalin ring) extends away from the active site into bulk solvent. The C11 hydroxyl group hydrogen-bonds with the Glu388 side chain and the backbone amide of Trp390, the latter being mediated by a water molecule. The C15 carboxylic acid forms a salt bridge with Arg173.

The position of the α-*S*-methylbutyryl group is revealed in the crystal structure of the LovD G5' mutant in complex with LVA. As in the MJA complex, the decalin ring and hydrophilic tail of LVA bind with similar geometry. Interestingly, the additional methylbutyryl group extends parallel to the MJA hydrophilic tail (Figure 6B). The proximity of the two tails gives LVA a hairpin shape, with the decalin ring forming the hairpin turn between the two tails. The hydrophobic side chain binding position is likely the site at which the acyl donor binds. The proximity of the tails also suggests how MJA competitively inhibits the acyl transfer reaction when it binds prior to the methylbutyryl substrate [86]. Because the hydrophilic tail of MJA partly obstructs access of the methylbutyryl group to the active site, ordered binding of the substrates is required.

Structural comparisons between complexes of LovD G5' with LVA and SVA suggest some strain is involved in accommodating the non-natural product, SVA (Figure 6C). SVA contains an additional methyl group compared to LVA, located on the ☐-S-methylbutyryl moiety. Superimposition using only α-carbons in the large domains of the two structures shows nearly identical arrangements of atoms in the large domains and the decalin rings (Figure 6D). But contact between the additional methyl group (attached at C2') and the side chain of Phe148 appears to push open the cleft between domains. As a result, Phe148 moves approximately 0.8 Å away from its position in the LVA complex (**Figure S52B**). There is also a 30° rotation about the C1'-C2' bond of the α-dimethylbutyryl moiety. The consequence of these rotations for the relative catalytic rates of the two substrates appears minor; the movements near the atoms directly involved in acyl transfer are small. This is consistent with the ability of LovD to catalyze acyl transfer using an α-dimethylbutyryl group as substrate instead of the natural α-methylbutyryl group. However, further amino acid mutations, such as the aforementioned Phe148, could improve the fit to the α-dimethylbutyryl substrate, or other variations. The structures presented here provide a framework for such design efforts.

In this work, we show that seven amino acid changes led to the ~ 11 fold increase in the SV synthase activity of LovD. This level of enhancement is significant considering G0 was already an adequate SV synthase following our previous efforts in substrate and strain optimization. Although the kinetic activity of G7 is far below that of the natural reaction catalyzed by LovD using acyl-LovF, G7 is a robust mutant for high volume synthesis of SVA using the whole cell platform. Indeed, when applied in a high density fermentation environment, more than 30g/L of MJA can be quantitatively converted to SVA within one day. The relatively

few rounds of direct evolution to achieve the activities of G7 also demonstrate that LovD is highly evolvable as a biocatalyst.
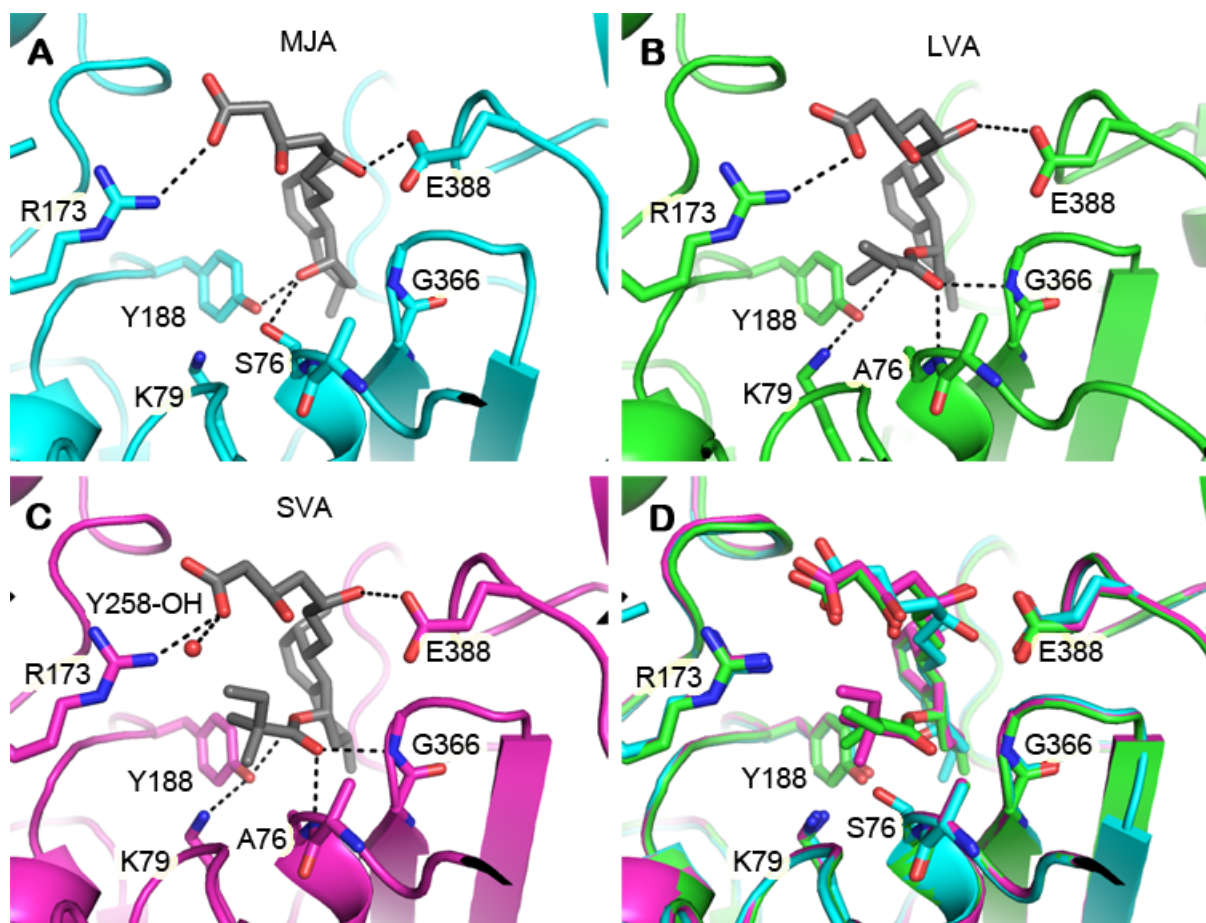

Figure 40. Comparison of the LovD active site bound with different ligands. (A) G5 in complex with substrate MJA. (B) G5' in complex with product LVA. (C) G5' in complex with product SVA. (D) Overlay of the three structures showing conformational changes, particularly at the nucleophilic serine, associated with binding to different ligands. Dashed lines represent hydrogen bonds. The active site entrance is at the top of each figure. Some residues involved in ligand binding are not shown.

The X-ray crystal structures solved in this work provides insight into different facets of LovD enzymology. Among these, the spatial arrangement of the LovD catalytic triad Ser76-Lys79-Tyr188 was captured and is shown to be consistent with that of the esterase EstB (**Figure S54**). Tyr188 (as the phenolate) appears to be the general base in initiating the two nucleophilic attacks required for completion of the acyl transfer reaction [229]. The first nucleophilic attack is by Ser76 on the α-*S*-methybutyryl group and the second attack is by the C8 hydroxyl of MJA on

the acylated enzyme intermediate. In both reactions, the attacking hydroxyl group must be activated by deprotonation (**Figure S54C**). Tyr188 is well-positioned to deprotonate both hydroxyl groups, forming hydrogen bonds with Ser76 in the apo-enzyme and with MJA in the G5-MJA complex. Lys79 is also well-positioned to aid in activating the two hydroxyl groups by forming a hydrogen bond relay with Tyr188 in the G5-MJA complex. Site-directed mutation of either Tyr188 or Lys79 to alanine resulted in complete loss of activity.

Details of the α-*S*-methylbutyryl binding pocket suggest how the transition states for the acyl transfer reactions are stabilized. As in the MJA complex, the C8 oxygen of LVA maintains a hydrogen bond to Tyr188, but the neighboring water molecule has been displaced by the carbonyl oxygen of the α-*S*-methylbutyryl group. This carbonyl oxygen forms a pair of hydrogen bonds with the backbone amides of Ala76 (i.e. Ser76 in the G0-Semet) and Gly366. The geometry of the hydrogen bonds appears ideal, with the amide hydrogen atoms pointed directly at the two lone electron pairs on the carbonyl oxygen. These hydrogen bonds would appear well suited to stabilize a tetrahedral transition state. The closest protein contacts with the aliphatic portion of the methylbutyryl group are aliphatic or aromatic carbons: Ala75 (Cβ), Phe148 (Cζ), Tyr146 (Cζ), and Asn270 (Cβ). Notably, the α-*S*-methylbutyryl aliphatic carbons are also surrounded by three positively charged side chains, Arg73, Lys79, and Arg173, all within 4.1 Å. These positive charges might help stabilize the negative charge of the oxyanion hole that forms during acyl transfer. Indeed, the pocket's affinity for negative charges is demonstrated by the fortuitous presence of a bound formate anion in this position in the G5-MJA complex.

The crystal structures provide plausible explanations for the basis of enhanced catalytic efficiency. Upon ligand binding, LovD undergoes a conformational change analogous to the

closing of the catcher's mitt. Movement of the domains shown in **Figure 39A** positions the catalytic residues in closer proximity to each other and to the ligands, and serves to enhance the rate of catalysis. The G5 structure suggested that beneficial mutations afforded an alternative way to pre-position the active site residues to increase the catalytic efficiency. Although the V334D and V334F mutations of G6 and G7, respectively, were discovered subsequent to the structural work of G5, the molecular explanation above can also be applied to rationalize the beneficial nature of the additional single mutations. Val334 is located in the middle of a loop between helix K and sheet 12. The two side chain methyl groups of Val334 are directly in contact with the side chain of Asp320 on helix K, which serves as one of the domain-domain hinges and is in contact with the loop where Tyr188 is located (**Figure 39B**). Mutation of V334 could therefore result in movement of the domains around the helix K hinge. For example, the mutation V334D in G6 would result in electrostatic repulsion between Asp334 and Asp320. Similarly, the mutation V334F in G7 could result in steric clash between the phenyl side chain and the side chain of Asp320, further closing the active site cleft and bringing key residues (e.g. Tyr188) into more optimal positions for catalysis. On the other hand, the more compact conformations of LovD mutants are less compatible with binding to LovF, which apparently favors the open conformation of LovD.

Crystallization of highly evolved LovD mutants

Collaborate with the industrial company Codexis®, we obtained a serious of evolved LovD mutants (simh1001, simh3189, simh6028 and simh9014). The best mutants simh9014 has 1000-fold higher of whole cell activity than LovD WT.

LovD WT, simh1001, simh6208 and 9014 are monomers in solution.

We employed the gel filtration chromatography to examine whether the LovD proteins exist as monomers or oligomers in solution. All the proteins were purified by using Ni-NTA agarose affinity resin, loaded onto a Superdex 200 (GE Healthcare) column and eluted with a flow rate of 0.5 ml/min in 50 mM Tris pH 8.0 and 2 mM DTT. The only peak showed in all the traces corresponds to the molecular weight of 46 kDa, which indicates that all the LovD proteins are monomers in solution, although we observed different packing patterns in a single crystal (4 monomer subunits in WT and simh9014 and 2 monomer subunits in simh6208) (**Figure S55**).

Crystallization of simh6208 and 9014.

LovD simh6208 was cloned into pET28a vector with N-terminus 6Xhis tag. However, no diffractable crystals of simh9014 could be obtained from this version, so we moved the 6Xhis tag to C-terminus. Both LovD simh6208 and 9014 were purified as previously described [230]. The proteins were then dialysed overnight into 50 mM Tris pH 8.0, 150 mM NaCl and 10 mM DTT (using Spectra/Por molecular porous membrane tubing MWCO 6~8,000 Da). LovD proteins were concentrated to 20 mg/mL (using Amicon Ultra 15 MWCO 30,000 Da) for crystallization trays.

Crystals of simh6208 were grown at RT by a hanging drop vapor diffusion method using a 1:1 protein to reservoir solution ratio for a total drop size of 2 μl. Diffraction quality crystals were obtained in 1~2 days when using 50mM sodium citrate, 16% propanol, 18% PEG 4,000, 1mM DTT, as a reservoir solution. In preparation for data collection crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen.

Crystals of simh9014 were grown at RT by a hanging drop vapor diffusion method using a 2:1 protein to reservoir solution ratio for a total drop size of 3 μl. Diffraction quality crystals were obtained in 1~2 days when using 0.2 M MgCl$_2$, 0.1M Bis-Tris pH 5.5, 18% PEG3,350 and

10mM DTT, as a reservoir solution. In preparation for data collection crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen.

Data collection and Structure Determination.

The simh9014 crystal belonged to space group P_1 with four protein molecules in the asymmetric unit. X-ray diffraction data were collected at the Advanced Photon Source (Argonne National Laboratory), beamline 24-ID-C, using an ADSC Quantum 315 CCD detector. Crystals were cryo-protected by a quick dip in a solution consisting of 6.5 µl reservoir and 3.5 µl 50% (w/v) D-(+)-trehalose. Crystals were cryo-cooled to 100 K for during the data collection. One-hundred-twenty 1.0° oscillation frames were collected at a wavelength of 0.9791 Å. Data reduction and scaling were performed using XDS [231]. Diffraction to 3.2 Å resolution was observed.

The simh6208 crystal belonged to space group $P2_1$ with two protein molecules in the asymmetric unit. X-ray diffraction data were collected at the same beamline as simh9014. Crystals were cryo-protected by a quick dip in a solution consisting of 6.5 µl reservoir and 3.5 µl glycerol. Crystals were cryo-cooled to 100 K for during the data collection. One-hundred-eighty 1.0° oscillation frames were collected at a wavelength of 0.9641 Å. Data reduction and scaling were performed using XDS. Diffraction to 1.8 Å resolution was observed.

The crystal structures were determined by the molecular replacement method using the program PHASER [232] and search model LovD G5 mutant, PDB entry 3HLC [230]. The models were refined using REFMAC5 [233] and Buster/TNT [234] with TLS parameterization of domain disorder [235]. After each refinement step, the models were visually inspected in COOT [236], using both 2Fo-Fc and Fo-Fc difference maps. The models were validated with the following structure validation tools: PROCHECK [237], ERRAT [238]and VERIFY3D [239].

For the simh9014 structure, 90.4% of the residues are within the most favoured region of the Ramachandran plot, 9.3% were in additional allowed regions, and 0.3% were in generously allowed regions. There were no residues in disallowed regions. For the simh6208 structure, 91.9% of the residues are within the most favoured region of the Ramachandran plot, 7.8% were in additional allowed regions, and 0.3% were in generously allowed regions. There were no residues in disallowed regions. The Errat scores, 93.7% and 94.5%, for simh9014 and simh6208 structures respectively, indicates that these percentages of residues fall below than 95% confidence limit of being erroneously modeled. Verify3D reports 100% and 96.2% of the residues have an averaged 3D-1D score greater than 0.2 for simh9014 and simh6208 structures, respectively. Data collection and refinement statistics are reported in **Table S11**. The structures were illustrated using Pymol [240].

Overall structure/comparisons of LovD WT with simh6208 and simh9014

The X-ray crystal structures of simh6208 and simh9014 were solved to facilitate the interpretation of the possible mechanism of the improved enzymatic activities. The resolution limit is 3.2 Å in the simh9014 variant, but improved to 1.8 Å for simh6208. Refinement statistics are provided in Table S2. Both structures displayed α/β hydrolyase fold as our previously solved LovD structures, which consist of a central seven-stranded antiparallel β-sheet flanked by α-helices on either side [230]. Simh6208 has a total of 20 mutations compared to the WT LovD, while simh9014 kept all the mutations from simh6208 (the same Q297 was mutated to E in simh6208, while G in simh9014) and obtain 9 other mutations in the later three rounds of directed evolution. All the mutations are scattered all over the protein and most of them are remote from the active site residue Ser76 (**Figure S56**, S164G is not shown due to disorder in simh6208; I4N, A9V, A261H are not shown in simh9014). Interestingly, three of total eight

mutations (K26E, G275S and L361M) in the previous G6 variant were converged with those in simh6208, indicating the necessities of these mutations to improve the catalytic activity.

2.6.3 Conclusions

Enzymes from natural product biosynthetic pathways are attractive candidates for creating tailored biocatalysts to produce semisynthetic pharmaceutical compounds. LovD is an acyltransferase that converts the inactive monacolin J acid (MJA) into the cholesterol-lowering lovastatin. LovD can also synthesize the blockbuster drug simvastatin using MJA and a synthetic α-dimethylbutyryl thioester, albeit with suboptimal properties as a biocatalyst. Here we used directed evolution to improve the properties of LovD towards semisynthesis of simvastatin. Mutants with improved catalytic efficiency, solubility and thermal stability were obtained, with the best mutant displaying an ~11-fold increase in an *Escherichia coli* based biocatalytic platform. To understand the structural basis of LovD enzymology, seven X-ray crystal structures were determined, including the parent LovD, an improved mutant G5, and G5 co-crystallized with ligands. Comparisons between the structures reveal that beneficial mutations stabilize the structure of G5 in a more compact conformation that is favorable for catalysis.

Simvastatin is a semisynthetic derivative of lovastatin and is the active pharmaceutical ingredient of the cholesterol-lowering drug Zocor® [212]. Previously we discovered a simvastatin synthase LovD to synthesize simvastatin from MJA and DMB-SMMP in one step. The efficiency of this reaction will determine whether it can be used for industrial production. So it is crucial to improve the properties of LovD. In this work, several better mutants were obtained through an high throughput agar-based diffusion assay and the best mutant "G7" displaying an ~11-fold increase in the *E. coli* based biocatalytic platform compared to the parent. Catalytic efficiency, solubility and thermostability were improved simultaneously showing the power of

our selection system. More strikingly, we have determined seven X-ray crystal structures including the parent LovD G0, an improved mutant G5, and the co-crystal structures of G5 with MJA, LVA and SVA. The structure information not only aided our understanding of the catalytic mechanism of LovD, but also afforded a great insight into how mutations affected the overall properties of LovD. Comparing the structures between LovD G0 and G5 suggests the beneficial mutations help promote a more compact conformation required for catalysis. The co-crystallization of LovD with substrate MJA, product LVA and SVA reveals how acyl transfer reaction proceeds via a ping-pong mechanism, how MJA becomes a competitive inhibitor and how the catalytic cavity is adapted to accommodate its non-natural product. Our work, therefore, therefore, can have significant impact on biocatalyst development and provides deep insights into fundamental understanding of enzymology.

2.6.4   Materials and Methods

Ep-PCR and Construction of Mutant Library. Ep-PCR procedure was modified from established protocols [221]. The reaction consisted of 0.35 mM dATP, 0.4 mM dCTP, 0.2 mM dGTP, 1.35 mM dTTP, 4 mM MgCl$_2$, 0.25 mM MnCl$_2$, and 2.5 U Taq polymerase. The reaction mixture was submitted to 25 cycles PCR: 94°C for 1 min, 55°C for 1 min and 72°C for 3 mins. The resulting PCR products were digested with *Dpn*I, further digested with *EcoR*I and *Nde*I, and ligated to pET28(a). The ligation mixture was transformed to YT2 and plated on LB agar containing 35 mg/L kanamycin.

Selection of High Activity Mutants. *N. crassa* was grown on SDA slants for 10 days and spores were harvested with 1% Tween-80. 100 ml of molten SDA was seeded with 0.3-0.5×10^8 spores and poured into a 230×230 mm plate. Colonies from mutation library were cultured in 96 well plates containing 250 μl LB medium (with 35 mg/L kanamycin). The cells were grown at 37°C

to saturation and transferred to duplicated plates. Protein expression was induced with 0.1 mM IPTG at OD_{600} of 0.5 and the expression was performed at 25°C for 16 h. 5 mM MJA and 10 mM DMB-SMMP were added to initiate the reaction. After a certain reaction time (45 mins to 4 hours depending on the activity of the parent), cells were removed with centrifugation (2,000 g, 4°C, 5 min). The amount of supernatant spotted on the SDA plate was typically 1~3 µl. The plates were incubated at 30°C for 16~18 hours. The improved mutants were selected following visual comparison of the inhibition zones.

Site-Directed Mutagenesis. Site-directed mutations were performed using the standard Quickchange® strategy using relevant templates. The primers were ordered from IDT (Integrated DNA Technologies). All mutations were verified by DNA sequencing (Laragen, Los Angeles, CA).

Saturation Mutagenesis. The LovD G6 gene was randomly mutated at positions of V334 and L361. Since the two residues are close to each other, the two random mutations were introduced in a single pair of primers. Two segments were amplified by PCR and linked together using slice-by-overlap extension (SOE) PCR to give intact LovD gene, which was subsequently introduced to pET28(a).

Determining Whole-cell Biocatalysis Activity. Parent LovD G0 and all mutants were cultured in parallel for comparison. A single colony of the freshly transformed YT2 competent cells was used to inoculate a 5 mL LB culture supplemented with 35 mg/L kanamycin. Following overnight growth at 37°C, 100 µl of the culture was inoculated into 50 mL LB medium supplemented with 35 mg/L kanamycin. When OD_{600} reached 0.4~0.6, 0.1 mM IPTG was added to the cultures and expression of all LovD variants was performed at 25°C for 16 hr. To mimic the high density fermentation conditions, the cells were then concentrated 10-fold before

addition of substrates. A 10 ml aliquot of each culture was collected by centrifugation (4°C, 2,000 g, 10 min). The cell pellet was gently resuspended in 1 ml of the medium supernatant, followed by addition of 70 µl of MJA (300 mM stock) to a final concentration of 15 mM. The concentrated culture was then divided into seven 200 µl aliquots and 1 µl of pure DMB-SMMP was added to each sample to a final concentration of 20 mM. The small cultures were then shaken at 300 rpm at 25 °C. At each time point, a complete extraction of one culture aliquot was performed by adding 10 µl of 20% SDS for cells lysis, followed by extraction with 500 µl ethyl acetate containing 1% trifluoroacetic acid TFA. The organic phase was removed, evaporated, and redissolved in 500 µl acetonitrile for HPLC analysis. The whole-cell activity was determined by fitting the linear regions of the conversion time course plot.

Kinetic Assay of LovD Variants towards MJA and DMB-SMMP. To obtain K_M values for MJA and k_{cat}, the DMB-SMMP concentration was fixed at 2 mM, while the concentration of MJA was varied from 0.25 to 5 mM. To obtain K_M values for DMB-SMMP and k_{cat}, the MJA concentration was fixed at 2 mM, while the concentration of DMB-SMMP was varied from 0.5 to 5 mM. Dimethyl sulfoxide (DMSO) was added to a final concentration of 10% to facilitate the solubilization of DMB-SMMP. At different time points of the kinetic assay, an aliquot of the reaction mixture was removed, quenched with 1% TFA and extracted with EA containing 1% acetic acid. The organic phase was separated, dried, resolubilized by acetonitrile (ACN) and analyzed by a Beckman Gold HPLC using a reverse phase C18 column (Alltech Apollo 5µ, 150 mm x 4.6 mm) and a linear gradient: 60% ACN in water (0.1% trifluoroacetic acid [TFA]) to 95% ACN in water (0.1% TFA) for 10 min, 1 mL/min. Conversion of MJA to SVA was measured by integration of the peaks at 238 nm.

Kinetic Assay of LovD Variants towards MB-SMMP. To compare the k_{cat} of LovD mutants towards lovastatin synthesis using MB-SMMP as the substrate, both MJA and MB-SMMP were fixed at 2 mM. DMSO was added to a final concentration of 10% to facilitate the solubilization of MB-SMMP. At different time points of the kinetic assay, an aliquot of the reaction mixture was removed, quenched with 1% TFA and extracted with EA containing 1% acetic acid. The organic phase was separated, dried, resolubilized by ACN and analyzed HPLC using the same program described above.

In Vitro Assay of LovD Variants towards LovF [149]. 50 μM LovF was incubated with 1 μM LovD variants, 2 mM MJA, 2 mM malonyl-CoA, 2 mM S-(5'-adenosyl)-L-methionine chloride (SAM), 2 mM NADPH in 100 mM PBS, pH 7.4. At 1 and 2 hours time points, an aliquot of the reaction mixture was removed, quenched with 1% TFA and extracted with EA containing 1% acetic acid. The organic phase was separated, dried, resolubilized by ACN and analyzed HPLC using the same program described above.

Comparing Expression Levels of Soluble LovD. Each expression plasmid encoding *LovD* mutant was transformed into *E. coli* BL21(DE3). The transformant was cultured in 50 mL LB medium containing 35 mg/L Kan at 37°C to optical density (OD_{600}) value of 0.4~0.6. Protein expression was induced with 0.1 mM IPTG and the subsequent expression was performed at 25°C for 16 h. Cells were collected by centrifugation (2,000g, 4°C, 15 min), resuspended in 7 mL Buffer A (50 mM Tris–HCl, pH 8.0, 2 mM DTT, 2 mM EDTA), and lysed by sonication. Cell debris and insoluble proteins were removed by centrifugation (20,000g, 4°C, 1 h). To the cleared cell lysate, 0.5 mL of Ni-NTA resin (Qiagen, Valencia, CA) was added to each sample. The mutants were then purified using a step gradient of Buffer A with increasing concentration of imidazole (10, 20, and 250 mM). LovD variants were eluted with 5 mL Buffer A containing

250 mM imidazole. The protein concentrations were qualitatively assessed by SDS–PAGE and quantitatively determined by the Bradford protein assay using bovine serum albumin (BSA) as the standard.

T_m **Measurement by Circular Dichroism.** Samples were prepared by adding 50 µg of proteins to 250µl 10 mM Tris-HCl buffer (pH 7.0). The sample was placed in a quartz cuvette with a 1 cm path length and heated in a Peltier-controlled cell at a rate of 1°C per min. Ellipticity was monitored at 222 nm in a Jasco spectropolarimeter (Jasco Inc., Easton, MD). The midpoint of the denaturation curve was determined with Microcal Origin 5.0 software (OriginLab Corporation, Northampton, MA).

ACCESSION NUMBERS. The coordinates and structure factors of G0, SeMet G0, G5, G5-MJA, G5'-MJA, G5'-LVA and G5'-SVA have been deposited into the Protein Data Bank under code 3HL9, 3HLB, 3HLC, 3HLD, 3HLE, 3HLF and 3HLG, respectively.

3 Conclusions

In conclusion, we have identified the gene clusters from several fungal strains which are involved in the biosynthesis of fungal peptidyl alkoilds, including tryptoquialanines, fumiquinazolines, asperlicins and ardeemin. The chemical logic and enzymatic machinery for generation of the architecturally complex of these peptidyl alkaloids scaffolds from simple building blocks is revealed. By systematically inactivating every gene (15 genes in tryptoquialanine pathway, 3 genes in asperlicin pathway and 2 genes in ardeemin pathway), followed by isolation and characterization of the intermediates, we were able to establish the enzymatic sequence of each pathway. The intact trimodular nonribosomal peptide synthetase (NRPS), TqaA protein (450kDa) was heterologous expressed in *Saccharomyces cerevisia.* TqaA was demonstrated to use L-tryptophan, L-alanine and the nonproteinogenic anthranilate to synthesize the common intermediate, fumiquinazoline F (FQF). Unlike bacterial NRPSs use thioesterase domains to perform the cyclization, fungal NRPSs producing macrocyclic peptides terminate with a condensation-like (C_T) domain that may perform the macrocyclization reaction. To probe the role of such a C_T domain, we dissected the cyclization steps of TqaA in transforming the linear anthranilate-D-tryptophan-L-alanyl tripeptide into FQF. Extensive biochemical and mutational studies confirmed the essential role of the C_T domain in catalyzing cyclization in a thiolation domain–dependent fashion. Our work provides evidence of a likely universal macrocyclization strategy used by fungal NRPSs. Similarly, the bimodular 276 kDa nonribosomal peptide synthetase AspA from *Aspergillus alliaceus*, heterologously expressed in *Saccharomyces cerevisiae*, converts tryptophan and two molecules of the aromatic β-amino acid anthranilate into a pair of tetracyclic peptidyl alkaloids asperlicin C and D in a ratio of 10:1. The first module of AspA activates and processes two molecules of Ant iteratively to generate a

tethered Ant-Ant-Trp-*S*-enzyme intermediate on module two. Release is postulated to involve tandem cyclizations, in which the first step is the macrocyclization of the linear tripeptidyl-*S*-enzyme, by the C_T domain to generate the regioisomeric tetracyclic asperlicin scaffolds. Computational analysis of the transannular cyclization of the 11-membered macrocyclic intermediate shows that asperlicin C is the kinetically favored product due to the high stability of a conformation resembling the transition state for cyclization, while asperlicin D is thermodynamically more stable. After gaining the information from the biosynthetic pathways, we next focus on engineering the acyltransferase, LovD from the lovastatin (LV) biosynthetic pathway to make the blockbuster cholesterol-lowering drug Zocor, simvastatin (SV). Several better mutants were obtained and the best mutant ''G7'' displayed an 11-fold increase in whole-cell biosynthesis of SV compared with the parent WT (G0). Catalytic efficiency, solubility, and thermostability were improved simultaneously. We have also determined seven X-ray crystal structures including the parent LovD G0, an improved mutant G5, and the cocrystal structures of G50 with MJ, LV, and SV. Comparing the structures between LovD G0 and G5 suggests the beneficial mutations help promote a more compact conformation required for catalysis. The cocrystallization of LovD G50 with substrate MJA, product LVA and SVA reveals how acyl transfer reaction proceeds via a ping-pong mechanism, how MJA becomes a competitive inhibitor, and how the catalytic cavity is adapted to accommodate its nonnatural product. Our work, therefore, can have significant impact on biocatalyst development and provides insights into fundamental understanding of enzymology.

4 Appendices

4.1 Supporting Information for Section 2.1

X-ray Crystallographic data collection

A colorless platelet-like crystal of **1**, approximate dimensions .40mm x .20mm x .10 mm, was used for intensity data. The diffraction data were measured at room temperature on a Bruker SMART Apex2 CCD-based X-ray diffractometer system equipped with a Mo-Kα radiation (λ= 0.71073 Å). The detector was placed at a distance of 6.00 cm from the crystal. A total of 1798 frames were collected with a scan width of 0.5° in ω, with an exposure time of 30 sec./frame. The frames were integrated with the Bruker SAINT software package using a narrow-frame integration algorithm. [1a] The integration of the data using a orthorhomic unit cell yielded a total of 29867 reflections to a maximum 2Θ angle of 56.64°, of which 3665 were independent (Rint = 5.3%). The final cell constants of a = 14.158 (10) Å, b = 21.005 (14) Å, c = 21.005 (12) Å, V = 5453(6) Å^3, are based upon the refinement of the XYZ centroids of 6528 reflections. The structure was solved and refined using the Bruker SHELXTL (Version 6.14) Software Package,[1b] in the space group C222$_1$. The absolute configuration of **1** was fixed based on the possible *R* configuration at C12 position and is consistent with the crystal structure of **4** established by X-ray crystallography.[2] All atoms were refined anisotropically and hydrogen atoms were placed at the calculated positions. The final anisotropic full-matrix least-squares refinement on F^2 converged at R$_1$ = 4.091%, wR$_2$= 16.55% and a goodness-of –fit of 1.042.

A colorless needle -like crystal of **18**, approximate dimensions .12mm x .02mm x .02 mm, was used for intensity data. The diffraction data were measured at 100K on a Bruker SMART Apex2 CCD-based X-ray diffractometer system equipped with a Mo-Kα radiation (λ= 0.71073 Å). The detector was placed at a distance of 6.00 cm from the crystal. A total of 1798 frames were collected with a scan width of 0.5° in ω, with an exposure time of 120 sec./frame. The

frames were integrated with the Bruker SAINT software package using a narrow-frame integration algorithm.[1a] The integration of the data using a hexagonal unit cell yielded a total of 41893 reflections to a maximum 2Θ angle of 52.72°. The final cell constants of a = 21.974 (10) Å, c = 10.8958 (12) Å, V = 4556(6) Å^3, are based upon the refinement of the XYZ centroids of 5907 reflections. The structure was solved and refined using the Bruker SHELXTL (Version 6.14) Software Package,[1b] in the space group P6$_3$. The absolute configuration of **18** was fixed based on the possible *S* configuration at C27 position and *R* configuration at C12 position and is consistent with incorporation of L-Ala and D-tryptophan moieties. All atoms were refined anisotropically and hydrogen atoms were placed at the calculated positions. The final anisotropic full-matrix least-squares refinement on F^2 converged at R$_1$ = 7.03%, wR$_2$= 17.13% and a goodness-of –fit of 1.082.

Table S1. Domain architectures of putative NRPSs in *P.aethiopicum* and comparison to the closest homologs in *A.clavatus* and *P. chrysogenum*

	Contig No.	Contig length	Location	NRPS Length (aa)	Domains	Closest homolog in BlastP	Closest *P. chrysogenum* homolog	Closest *A. clavatus* homolog
1	contig00857	73656	6614-13830	2368	A-T-C-A-T-C	Pc21g15480 (87%)	Pc21g15480 (87%)	ACLA 059530 (31%)
2	contig01275	41088	37-12346	4074	C*A-T-C-A-T-E-C-A-T-C	ACLA 017890 (63%)	Pc13g14330 (33%)	ACLA 017890 (63%)
3	contig00865	23983	10454-23163	4209	C-A-T-C-A-T-C-A-T-C	Pc13g14330 (53%)	Pc13g14330 (53%)	ACLA_059530(45%)
4	contig00311	157072	142932-149218	2082	A-T-C-A-T-C	Pc22g20400 (94%)	Pc22g20400 (94%)	AFLA_010620 (68%)
5	contig00107	69917	32541-38452	1921	A-T-C-A-T-C	Pc16g03850 (93%)	Pc16g03850 (93%)	ACLA_061000(42%)
6	contig00904	94631	40241-47585	2409	A-T-C-A-T-C	Pc21g12630 (84%)	Pc21g12630 (84%)	ACLA_076770(33%)
7	contig00185	110569	31302-42930	3864	A-T-C-A-T-C	NFIA_044240 (83%)	Pc21g10790 (31%)	ACLA_095980(34%)
8	contig00284	128491	104811-121323	5397	A-T-C-A-T-C-A-C-A-T-C-A-T-C	AFLA_066720 (54%)	Pc21g10790 (30%)	ACLA_059530(28%)
9	contig00894	67132	8793-27331	6048	A-T-E-C-A-T-C-A-T-E-C-A-T-C-A-T	Pc16g04690 (93%)	Pc16g04690 (93%)	ACLA_025160(57%)
10	contig00025	17604	543-13015	4102	A-T-C-A-T-C-A-C	ABR23346.1 (32%, partial identity)	Pc21g12630 (31%)	ACLA_093780(32%)
11	contig01022	16832	9044 - 12370	1108	A-T-C	ACLA_017900 (67%)	Pc21g10790 (32%)	ACLA_017900(67%)
12	contig01242	215775	55563 - 58001	812	A	Pc22g22580 (85%)	Pc22g22580 (85%)	ACLA_093780(30%)
13	contig01239	68878	23001 - 25741	809	A	Pc21g12840 (91%)	Pc21g12840 (91%)	ACLA_098420(32%)
14	contig00921	69917	32996 - 36780	1215	A	Pc20g12670 (82%)	Pc20g12670 (82%)	ACLA_095980(32%)
15	contig00239	135391	104823 - 111215	2106	A-T-C-A-T-C	Pc21g01710 (94%)	Pc21g01710 (94%)	ACLA_061190(64%)
16	contig00759	116158	26436 - 41994	5096	A-C-A-T-C-A-T-C-A-T-C-A-T-C-A-T-C	Pc13g05250 (89%)	Pc13g05250 (89%)	ACLA_079690(53%)

The NRPS genes in *P. aethiopicum* were identified from the local genome database using A domains of the acetylaszonalenin synthetase of *Neosartorya fischeri NRRL 181*, which activates an anthranilate and a tryptophan. A total of 16 NRPS genes were found. The strategy for the search of *P. aethiopicum* NRPSs genes responsible for the production of **1** was based on assumptions that 1) PaeNRPSs that are highly similar and orthologous (> 80% identity with identical domain architecture) to those in *P. chrysogenum* are unlikely to produce **1** and thus can be excluded; 2) since the PaeNRPSs for **1** in *P. aethiopicum* and **2** in *A. clavatus* are likely to be close homologs of each other, PaeNRPSs genes that are highly similar to those in *A. clavatus* are good candidates. Combining these two search parameters allow us to narrow down the candidate to two (PaeNRPS1275 and PaeNRPS1022). Incidentally, the homologs of the two NRPSs in *A. clavatus* (ACLA_17890 and ACLA_17900) are adjacent to each other in the genome and have a combined total of four modules, correspond to the four amino acid units required to build **1** and **2**. The close similarity of PaeNRPS1275 to the tryptophan- and anthranilate-activating A domains (both 47% identity) of the acetylaszonalenin synthetase of *N. fischeri NRRL 181* supports that PaeNRPS1275 is a highly plausible candidate.

Primer	Sequence
tqaAKO_P1	5'-AATTCCACCAAGCCCTTACA-3'
tqaAKO_P2	5'-ATGACGGTCCAATCTCTCGA-3'
tqaAKO_P3	5'-GCCCGTCACCGAGATTTAGGGTCCTGCCTCCAAGCAGGCG-3'
tqaAKO_P4	5'-CAATATCATCTTCTGTCGACGTCGCCGAGTAAAAGCCCGT-3'
tqaAKO_P5	5'-GATCAAGTTTGTGCTTGCCG-3'
tqaAKO_P6	5'-AATTAGATTTACCGCAGTCA-3'
tqaBKO_P1	5'-AAACCATAAAATCAATTGGA-3'
tqaBKO_P2	5'-ATTTTCAACCCACGAGACAA-3'
tqaBKO_P3	5'-GCCCGTCACCGAGATTTAGGCATTATACAGGTGAACATAT-3'
tqaBKO_P4	5'-CAATATCATCTTCTGTCGACGAAACCTGCGAACTACCGGA-3'
tqaBKO_P5	5'-GTAAGTCTTGGTTGGGGCCT-3'
tqaBKO_P6	5'-GTTCGACCGATTCGAGCAGC-3'
tqaCKO-P1	5'-GAAATCACTCGACAAGGTGGAC -3'
tqaCKO-P2	5'-AGGATGACGCCTCAGGAAATGT-3'
tqaCKO-P3	5'-TGCCCGTCACCGAGATTTAGGCTTCAAGCTCGGGTGTGGAAGG-3'
tqaCKO-P4	5'-TCAATATCATCTTCTGTCGACGAGACCTGCGGATCATTGGGAG-3'
tqaCKO-P5	5'-AGCGTAACGTAAACCTCCCACC-3'
tqaCKO-P6	5'-CTGAAGAAATGAGCCCATAGCA-3'
tqaDKO-P1	5'-TTCCCGGAGACTAGAACTGGCT-3'
tqaDKO-P2	5'-TGCCACGGTAGCAATAGTCAAC-3'
tqaDKO-P3	5'-TGCCCGTCACCGAGATTTAGGGTCCCACAGTATGATCCAAGCAG-3'
tqaDKO-P4	5'-TCAATATCATCTTCTGTCGACCTACAGTCGGATGCTATGGGCT-3'
tqaDKO-P5	5'-GATGCTCAAATGTCGCCAGGTC-3'
tqaDKO-P6	5'-CTGAGGGAGTCAACATGGCAAG-3'
tqaEKO-P1	5'-TCGAACGATGGCTGCAAATCTT-3'
tqaEKO-P2	5'-GCCGATGGAGAGGAGGAAATAC-3'
tqaEKO-P3	5'-TGCCCGTCACCGAGATTTAGGTGGGTTGTTCCACGCAGAAGAC-3'
tqaEKO-P4	5'-TCAATATCATCTTCTGTCGACCAAGGCCGACAATGATGACAGT-3'
tqaEKO-P5	5'-AGGGCTTCCCATTTGATTTGGT-3'
tqaEKO-P6	5'-GTGGTGTCCAATGTGGACCAAA-3'
tqaFKO-P1	5'-ATCGCCTTCCGAAGATAAGAGA-3'
tqaFKO-P2	5'-ATTTCTTCCTGTAACCCTCGTC-3'
tqaFKO-P3	5'-TGCCCGTCACCGAGATTTAGGAGGCGGTATGCGGATAACTGAG-3'
tqaFKO-P4	5'-TCAATATCATCTTCTGTCGACGACTGGCGATTTGACACACTTG-3'
tqaFKO-P5	5'-TGAAGACAGTTCGCTACCGAGC-3'
tqaFKO-P6	5'-CAATTCCCCAGCACTACTGAAG-3'
tqaGKO-P1	5'-ACAATACGGTCTTCCAGGCGCT-3'
tqaGKO-P2	5'-AGCTTGATTATCGCAGCAACCA-3'
tqaGKO-P3	5'-TGCCCGTCACCGAGATTTAGGAGGTGATGTCGGTTCCAAAGGA-3'
tqaGKO-P4	5'-TCAATATCATCTTCTGTCGACCCAATGAGCACGGACACGAAAC-3'
tqaGKO-P5	5'-GTTTGTGACCAAGGAGGAGGTG-3'
tqaGKO-P6	5'-GACTGTGGATGATGGAGAGCGT-3'
tqaHKO-P1	5'-ACCGTAGGCGTTCCACGTCATG-3'
tqaHKO-P2	5'-TGCTGGCTTTCATCAGATTGGT-3'
tqaHKO-P3	5'-TGCCCGTCACCGAGATTTAGGTCCGGTTCCCTAATCCTGACTG-3'
tqaHKO-P4	5'-TCAATATCATCTTCTGTCGACATCTACGTGCGTGGAGTCACAG-3'
tqaHKO-P5	5'-TCTCCAAATCGAGGGCAAGCAT-3'
tqaHKO-P6	5'-GATGGTGGGAGCTGATGGTTGG-3'
tqaIKO-P1	5'-TGCAATGAAGGCATCTGCACTG-3'
tqaIKO-P2	5'-GCCGACATCTTTGACCATCCTG-3'
tqaIKO-P3	5'-TGCCCGTCACCGAGATTTAGGTGGCCAGTATCCAAACATGGGT-3'
tqaIKO-P4	5'-TCAATATCATCTTCTGTCGACAAATGGGCATTGGGAAGAACCT-3'
tqaIKO-P5	5'-TGGCAAGCGACAACAAATCATC-3'

tqaIKO-P6	5'-GCAGTGTACGACCGAGTGGTG-3'
tqaKKO-P1	5'-GCAGCAAATTGCGAAAGCTCTC-3'
tqaKKO-P2	5'-ACCGTCGATGAGGCATCCTTTC-3'
tqaKKO-P3	5'-TGCCCGTCACCGAGATTTAGGGGTTTCCTTTCCATCCGGCTCTT-3'
tqaKKO-P4	5'-TCAATATCATCTTCTGTCGACGACATAGCACTTGCAGCCATTG-3'
tqaKKO-P5	5'-TGGCCGACTCGCTTACCGACTA-3'
tqaKKO-P6	5'-AGGTTGAGTTGGGTGGCCGACT-3'
tqaLKO-P1	5'-CAGGATTTCGGAATGAACCATG-3'
tqaLKO-P2	5'-GCTGTCGTGAAGGGAAAGCAGA-3'
tqaLKO-P3	5'-TGCCCGTCACCGAGATTTAGGTCCACGGCCAAGTCTCTGGTCT-3'
tqaLKO-P4	5'-TCAATATCATCTTCTGTCGACGAACCTGTCGGATGGATATGGA-3'
tqaLKO-P5	5'-CAGTCTAGTTGGTGATTTCGTG-3'
tqaLKO-P6	5'-TCGTGGAGCCGGTCATAACTTT-3'
tqaMKO-P1	5'-TCTTGTTCAGCGCATCCACTAC-3'
tqaMKO-P2	5'-GACTGACACAATCATGGCAATG-3'
tqaMKO-P3	5'-TGCCCGTCACCGAGATTTAGGTCAGGCTCTAATTTGGAAGATG-3'
tqaMKO-P4	5'-TCAATATCATCTTCTGTCGACTTGGCTGTATGTAGAGGCGCAG-3'
tqaMKO-P5	5'-CTAGTGCGAATTAGTTCCCCA-3'
tqaMKO-P6	5'-GTCATCAATGAGTGGTTGTCGT-3'
orf4KO-P1	5'-TAGCGATGACAAGCATTGCG
orf4KO-P2	5'-GATCAAGCCAGTCGAGGTTG
orf4KO-P3	5'-TGCCCGTCACCGAGATTTAGGCACGTCTGGGTTGGACTATACTG-3'
orf4KO-P4	5'-TCAATATCATCTTCTGTCGACAGTCGGTGAATGTGCTTCCA-3'
orf4KO-P5	5'-TCCAGCAACACAAGTGAAGCT-3'
orf4KO-P6	5'-GAAATGGCAGAATCCCAGTCTT-3'
orf5KO-P1	5'-CTCTAAAGCGCGGCGAGTCG-3'
orf5KO-P2	5'-GATGGAGACATGGTGGTAATTG-3'
orf5KO-P3	5'-TGCCCGTCACCGAGATTTAGGAGTGAGGGCTCATGTCGTGGAG-3'
orf5KO-P4	5'-TCAATATCATCTTCTGTCGACGTTGCATTGGGTCCTACAGTTC-3'
orf5KO-P5	5'-CAAAGCCTCCGGCTTCCTAGC-3'
orf5KO-P6	5'-TGCTATGTCTCGGGTATTCG-3'
orf6KO-P1	5'-CATGTTTGTTGGAGCAAGCCT-3'
orf6KO-P2	5'-TCGTGGGTACAATCTTCCGA-3'
orf6KO-P3	5'-TGCCCGTCACCGAGATTTAGGGGCAACCTGTCCGGGATTTC-3'
orf6KO-P4	5'-CAATATCATCTTCTGTCGACGTGTTCTCGGAGGACGAACA-3'
orf6KO-P5	5'-CTCTGTTGGCAGGCTACCTC-3'
orf6KO-P6	5'-ACTAGCCGGGCTATACAAGGT-3'
TqaB_SacI_F	5'-AA**GAGCTC**ATGTTTGAGCCAATCGAGACTC-3'
TqaB_SacI_R	5'-TT**AAGCTT**TTACCCGTTGATCTTGGAGAGC-3'
TqaB_CDFduet_SacI_F	5'-AAAAAA**GAGCTC**GATGTTTGAGCCAATCGAGACTC-3'
TqaH_NdeI_F	5'-AAAAAA**CATATG**ACAACCGACCGTCAACCAT-3'
TqaH_EcoRI_R	5'-TTTTTT**GAATTC**TCACCCGCGCTTGTATGTATTC-3'
TqaH_XhoI_R	5'-TTTTTT**CTCGAG**TCACCCGCGCTTGTATGTATTC-3'
TqaD_NheI_F	5'-AA**GCTAGTA**TGACAGTCTCTGAAACAAAGA-3'
TqaD_XhoI_R	5'-AA**GCTAGTA**TTACGAAATGTATCTTGCAACT-3'
Seq contig 1022B_R	5'-CTAACGTTTGAAGATATAAC-3'
Seq contig 326B_R	5'-GTCTTATGAGTATTCAGCCT-3'

	1					**20**			

NO.	^{13}C δ(ppm)	^1H δ(ppm) (m, J_{HH}(Hz))	HMBC	HMQC		^{13}C δ(ppm)	^1H δ(ppm) (m, J_{HH}(Hz))	HMBC	HMQC
2	83.9	5.22 (s)	16N-OH, H13	H2		88.4	5.41 (s)	H15, H13, 16N-OH	H2
3	85.5		H2, H13			84.2		H13, H5, H2	
4	134.3		H8, H6, H13			134.3		H8, H6, H13	
5	125.4	7.91(m)	H7	H5		125.4	7.91(m)		H5
6	125.2	7.30 (t, 7.4)	H8	H6		125.4	7.32 (t, 7.5)	H8	H6
7	131.4	7.52 (m)	H5	H7		131.4	7.53 (t, 7.5)	H5, H9	H7
8	114.7	7.52 (m)	H6	H8		115.0	7.50 (t, 7.5)	H6	H8
9	137.1		H5, H7			137.2		H5	
11	169.9		H12, H13			169.8		H12, H13	
12	54.4	5.91 (t, 9.9)	H13	H12		54.5	5.91 (t, 9.9)	H13	H12
13	34.2	3.12 (m)	H12, H2	H13		33.4	3.13 (dd, 13.5, 9.9)	H2	H13
		3.08 (m)		H13			3.04 (dd, 13.5, 9.9)		H13
14	170.8		H29, H30			169.9		H29, H2, H15	H15
15	70.6		H29, H30			67.3	4.07 (q, 7.1)	16N-OH	
16N-OH		7.95 (S)					8.01 (s)		
18	160.4		H20, H12			160.5		H12, H20	
19	119.8		H23, H21			119.8		H23, H21	
20	126.2	8.17 (d, 7.9)	H22	H20		126.2	8.17 (d, 7.9)	H18, H22	H20
21	128.3	7.63 (t, 7.5)	H23	H21		128.1	7.63 (t, 7.6)	H23	H21
22	135.6	7.94 (m)	H20	H22		135.6	7.92 (m)	H20	H22
23	127.4	7.76 (d, 8.2)	H21	H23		127.4	7.76 (d, 8.1)	H21	H23
24	146.0		H20, H22			146.0		H20, H22	
26	154.8		H27, H12, H28			154.9		H27, H28	
27	68.4	6.31 (q, 6.2)	H28	H27		68.4	6.28 (q, 6.3)	H28	H27
CH$_3$COO$^-$	170.1		H27, CH$_3$COO$^-$			170.1		H27, CH$_3$COO$^-$	
CH$_3$COO$^-$	20.6	2.07 (s)		CH$_3$COO$^-$		20.6	2.08 (s)		CH$_3$COO$^-$
28	18.6	1.68 (d, 6.3)	H27	H28		18.6	1.68 (d, 6.3)	H27	H28
29	22.8	1.30 (s)	H30	H29		10.9	1.43 (d, 7.1)	H15	H29
30	16.7	1.34 (s)	H29	H30					

153

Table S4. NMR Data of **14** in CDCl3 Measured on 500 MHz Bruker NMR Spectrometer.
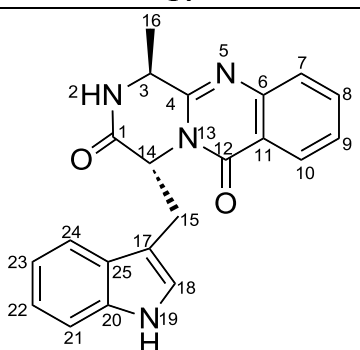
	14			

NO.	^{13}C δ(ppm)	^{1}H δ(ppm) (m, J_{HH}(Hz))	HMBC	HMQC
1	169.3		H2, H14, H15	
2		6.63 (br)		H3
3	49.1	3.13 (q, 6.5, 0.3)	NH2, H16	
4	151.6		H2, H14, H3, H16	
6	147.0		H10, H8	
7	127.3	7.59 (d, 8.1)	H9, H10	H7
8	134.7	7.77 (td, 7.7, 1.7)	H10	H8
9	127.1	7.53 (t, 7.6)	H8, H7	H9
10	126.8	8.36 (dd, 8.0, 1.1)	H8, H9	H10
11	120.2		H7, H9	
12	160.8		H10	
14	57.5	5.68 (m)	NH2, H15	H14
15	27.0	3.70 (dd, 14.8, 5.1) 3.64 (dd, 15.0, 3.6)	H18, H14	H15
16	19.1	1.36 (d, 6.6)	H3	H16
17	109.4		H19, H24, H23, H18, H14, H15	
18	123.5	6.67 (d, 1.7)	H15	H18
19		8.26 (br)		
20	135.9		H19, H24, H22, H18	
21	111.2	7.28 (d, 8.1)	H22	H21
22	122.6	7.12 (t, 7.8)	H24, H21, H23	H22
23	120.0	6.91 (t, 7.6)	H21, H22	H23
24	118.5	7.39 (d, 8.0)	H21, H22	H24
25	127.3		H24, H21, H23, H18, H15	

	18				19			

NO.	^{13}C δ(ppm)	^{1}H δ(ppm) (m, J_{HH}(Hz))	HMBC	HMQC	^{13}C δ(ppm)	^{1}H δ(ppm) (m, J_{HH}(Hz))	HMBC	HMQC
2	79.0	5.00 (d, 9.0)	H3, 16NH	H2	81.8	5.11 (d, 8.7)		H2
3	74.4				75.0		H2, 3OH	
3OH		5.66(s)				5.60 (s)		
4	138.1		H8		138.1		H8	
5	124.5	7.45 (d, 7.6)	H7	H5	124.6	7.45(d, 7.7)	H7	H5
6	124.8	7.12(m)	H8	H6	124.8	7.13 (dd, 9.9, 6.3)	H8	H6
7	129.7	7.33(m)	H5	H7	129.6	7.33 (m)	H5	H7
8	114.4	7.32(m)	H6	H8	114.6	7.33 (m)	H6	H8
9	137.4		H5, H7		137.5		H5	
11	175.3		H13		174.7		H13	
12	53.0	5.35(dd, 8.0, 3.6)	H13, 31NH	H12	53.0	5.38 (dd, 13.0, 9.9)	31NH	H12
13	37.5	2.81 (dd, 14.4, 7.9)	16NH	H13	37.5	2.83 (dd, 14.8, 3.9)	3OH	H13
		2.69 (dd, 14.4, 7.5)		H13		2.66 (dd, 14.7, 7.4)		
14	168.5		H29, H30		168.5		H13, H29	
15	64.4		H29, H30		60.0	3.65 (m)		H15
16NH		2.84 (d, 3.4)				3.17 (m)		
18	160.3		H20		160.3		H20	
19	120.2		H21, H23		120.1		H21, H23	
20	126.2	8.15 (d, 7.9)	H22	H20	126.3	8.15 (d, 7.9)	H22	H20
21	126.7	7.53 (t, 7.5)	H23	H21	126.8	7.54 (t, 7.6)	H23	H21
22	134.5	7.83 (m)	H20	H22	134.5	7.84 (t, 7.7)	H20	H22
23	126.9	7.66 (d, 8.0)	H21	H23	127.0	7.67 (d, 8.1)	H21	H23
24	147.0		H20, H22		146.9		H20, H22,H27	
26	152.9		H27, 31NH, H28		153.0		H27, 31NH, H28	
27	48.8	4.96 (q, 6.3)	H28, 31NH	H27	48.7	4.96 (q, 6.1)	31NH, H28	H27
28	17.3	1.58 (d, 6.5)	H27	H28	17.1	1.58 (d, 6.6)	H27	H28
29	24.4	1.11(s)	H29	H29	17.2	0.98 (d, 7.1)		H29
30	23.5	0.93(s)	H30	H30				
31NH		8.67(s)				8.67 (s)		

Table S6. NMR Data of **24** and **25** in DMSO-d6 Measured on 500 MHz Bruker NMR Spectrometer.

	24				**25**			

NO.	^{13}C δ(ppm)	^1H δ(ppm) (m, J_{HH}(Hz))	HMBC	HMQC	^{13}C δ(ppm)	^1H δ(ppm) (m, J_{HH}(Hz))	HMBC	HMQC
2	79.8	5.22 (d, 6.8)	16NH, H13	H2	84.3	5.31 (s)	16NOH	H2
3	83.8		H2, H13		83.5		H2,H5,H13	
4	134.6		H6, H8, H13		134.4		H5,H8,H13	
5	124.6	7.59(d, 7.3)	H7	H5	124.6	7.58(d, 7.5)	H7	H5
6	125.5	7.33 (t, 7.4)	H8	H6	125.4	7.31 (dd, 10.0, 3.8)	H8	H6
7	131.4	7.52 (m)	H5	H7	131.5	7.53 (m)	H5	H7
8	115.3	7.49 (m)	H6	H8	114.9	7.50 (m)	H6	H8
9	138.0		H7, H8		137.3		H5, H7,H8	
11	170.0		H13		170.1		H13	
12	55.8	6.08 (t, 9.9)	H2,H13	H12	55.9	6.16(t, 10.4)	H13	H12
13	33.4	3.17(dd, 13.1, 9.8)	H2	H13	35.1	3.21 (m)	H2	
		3.08(dd, 13.1, 9.9)		H13		3.18 (dd, 12.6, 9.2)		H13
14	175.0		H29, H30		171.1		H29, H30	
15	64.9		H29, H30		70.6		H29, H30	
16NH		3.22 (d, 8.4)						
16NOH						8.29 (s)		
18	160.1		H20		160.1		H20	
19	121.4		H21, H23		121.6		H21, H23	
20	126.4	8.22 (d, 8.0)	H22	H20	126.4	8.18 (d, 7.9)	H22	H20
21	129.6	7.74 (t, 7.6)	H23	H21	129.7	7.74 (t, 7.6)	H23	H21
22	135.5	7.98 (t, 7.7)	H20	H22	135.5	7.98 (t, 7.7)	H20	H22
23	128.3	7.89 (d, 8.2)	H21	H23	128.4	7.89 (d, 8.0)	H21	H23
24	145.0		H20, H22		145.0		H20, H22	
26	150.0				150.0		H28	
27	195.2		H28		195.3		H28	
28	27.4	2.77 (s)		H28	27.4	2.77 (s)		H28
29	25.6	1.34 (s)	H30	H29	22.8	1.27 (s)	H30	H29
30	24.9	1.42 (s)	NH, H29	H30	16.3	1.32 (s)	H29	H30

Table S7. NMR Data of **28** in DMSO-*d*₆ Measured on 500 MHz Bruker NMR Spectrometer.

28

No	¹³C δ(ppm)	¹H δ(ppm) (m, J*HH*(Hz))	HMBC	HMQC
2	80.6	5.47 (d, 7.3)	NH, H13	H2
3	84.3		H2, H5, NH, H13	
4	134.5		H6, H13	
5	125.8	7.92(m)	H7	H5
6	125.2	7.32 (t, 7.4)	H8	H6
7	131.3	7.52 (m)	H5	H7
8	114.9	7.49 (m)	H6	H8
9	138.0		H5, H8	
11	170.4		H12, H13	
12	54.1	5.91 (t, 9.6)	H13	H12
13	33.4	3.10 (dd, 13.8, 10.6)	H2, H12, H13	H13
		2.90 (dd, 13.8, 8.9)		H13
14	174.9		NH, H28,H29	
15	64.6		H29, H30	
16NH		3.18 (d, 7.4)		
18	160.4		H20	
19	119.8		H23, H21	
20	126.3	8.20 (d, 7.9)	H22	H20
21	127.9	7.62 (t, 7.6)	H23	H21
22	135.5	7.91 (m)	H20	H22
23	127.4	7.74 (d, 8.1)	H21	H23
24	146.1		H20	
26	155.0		H12, H22, H26, H27	
27	68.3	6.19 (q, 6.3)	H28	H27
CH₃COO⁻	170.2		H27, CH₃COO⁻	
CH₃COO⁻	20.6	2.10 (s)		CH₃COO⁻
28	18.7	1.67 (d, 6.3)	H27	H28
29	24.9	1.44 (s)	16NH, H30	H29
30	25.5	1.22 (s)	H29	H30
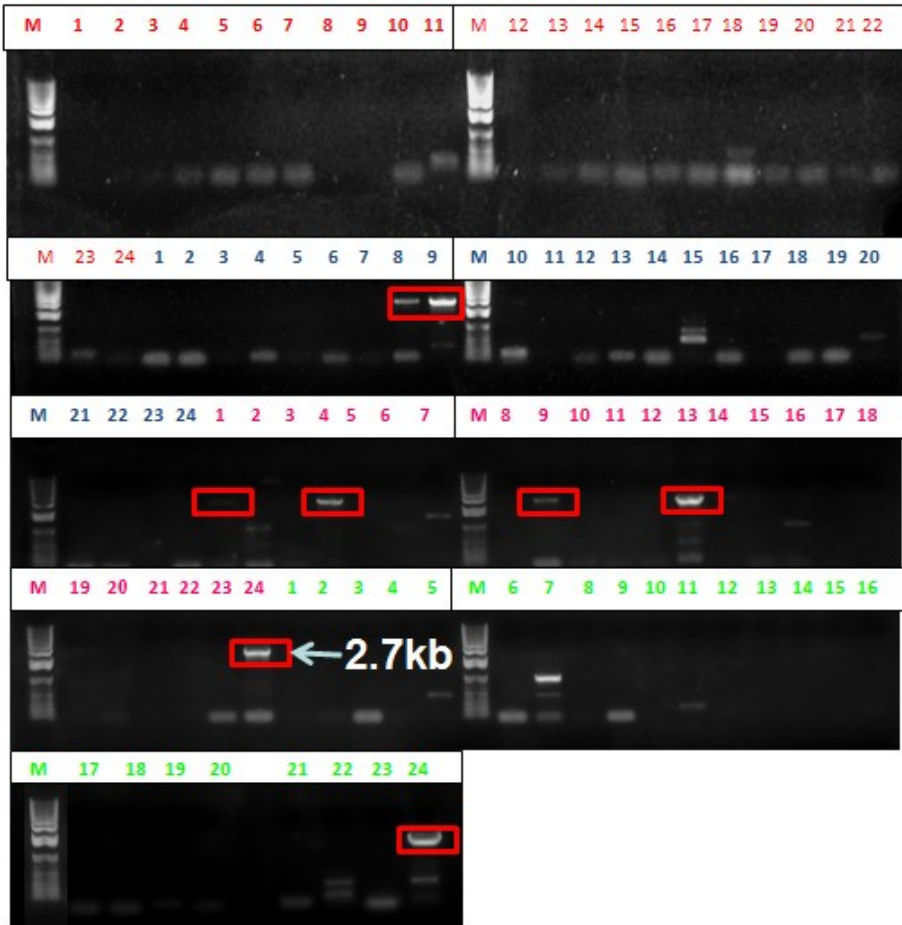
Figure S1. PCR verification of PaeNRPS1275 transformants as an example. 96 transformants were screened by PCR using bar gene primer (BarF) and primer outside of the deletion cassette (PaeNRPS1275KO_P1). Positive transformants showed a band at 2.7kb.
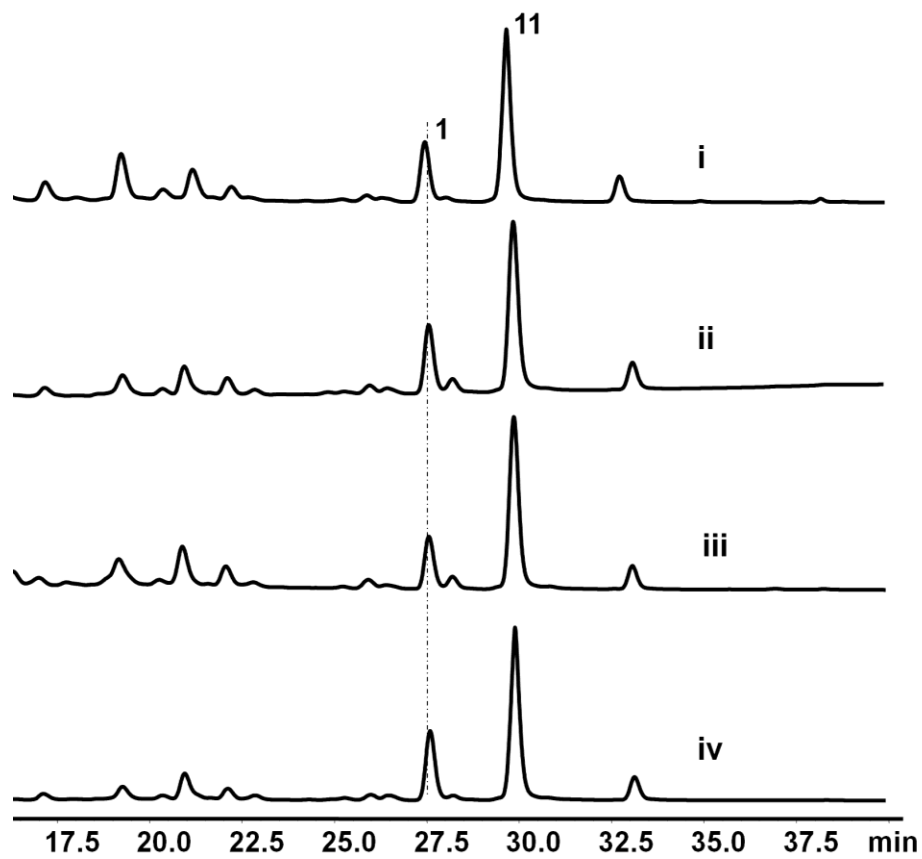
Figure S2. HPLC traces of metabolic extracts from single gene deletion strains of *P. aethiopicum.* Trace i: Δ*gsfA;* trace ii Δ*gsfA/*Δ*orf4*; trace iii: Δ*gsfA/*Δ*orf5;* trace iv: Δ*gsfA/* Δ*orf6*.
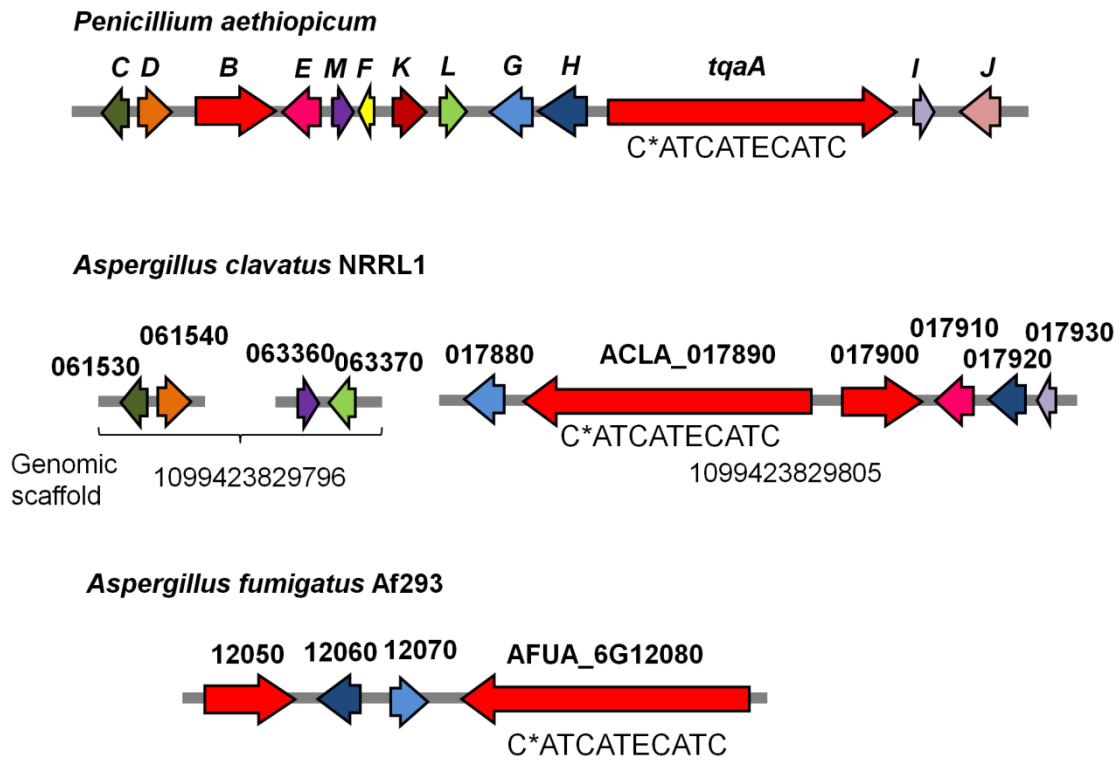
Figure S3. Corresponding *tqa* homologs in *A. clavatus* NRRL1 and *A.* fumigatus Af293 (corresponding homologs are in the same color); domain architecture of TqaA, TqvA (ACLA_017890) and FqaA (AFUA_6G12080). Homologous *tqv* gene cluster are fragmented in three different genomic loci.

160
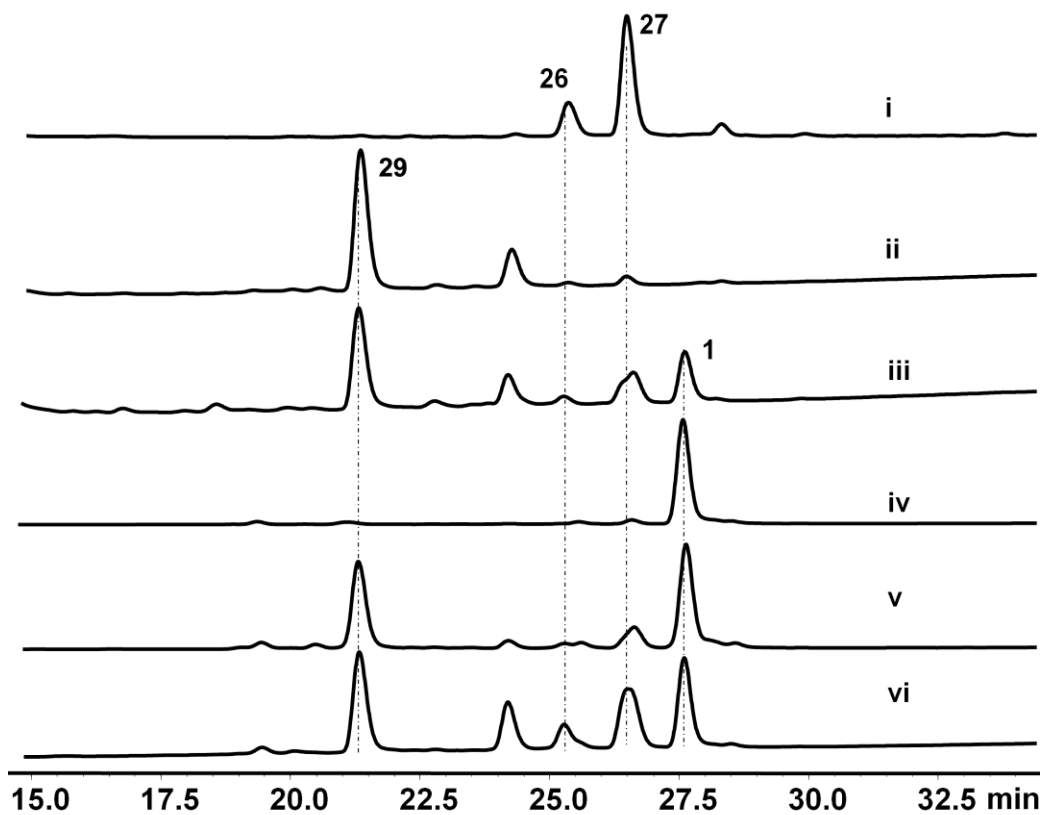
Figure S4. Biochemical characterizations of TqaD. Trace i: Compound **26** and **27** purified from ΔtqaD strain as substrate; extracts from overnight in vitro assays in 50 mM HEPES, pH 7.9, trace ii: control with 1 mM **26** and **27** mixture, 4 mM acetyl-CoA; trace iii: assay with 1 mM **26** and **27** mixture, 4 mM acetyl-CoA, 10 μM TqaD; trace iv: control with 1 mM 1; trace v: assay with 1 mM 1, 10 μM TqaD; trace vi: assay with 1 mM 1, 10 μM TqaD. Reaction samples in trace i to v were extracted by EA only; reaction sample in trace vi is extracted by EA+5% TFA.

161

Figure S5. Substrate binding pocket comparison of the A-domains of TqaB and Af12050 generated via homology modeling using HHpred with PDBid 1AMU as template. (A) A top-down view of the modeled substrate binding pocket and (B) a side-on view generated by rotating 90° about the horizontal plane. Residues are labeled according to their position in the 10AA specificity code or with their respective residue identifier. 2-aminoisobutyric acid (2-AIB, green ball-and-stick) was manually docked into the active site based on the position of the bound Phe of 1AMU. The pro-R (D-configuration) methyl of 2-AIB is indicated by an asterisk. The overlapping ")(" symbol indicates steric clash.

tryptoquialanine A **1**
Exact Mass: 518.18

RT=27.5 min   T: + c ESI MS [200.00-1200.00]

Figure S6. 1 UV spectrum and MS measured during LC-MS for **1**
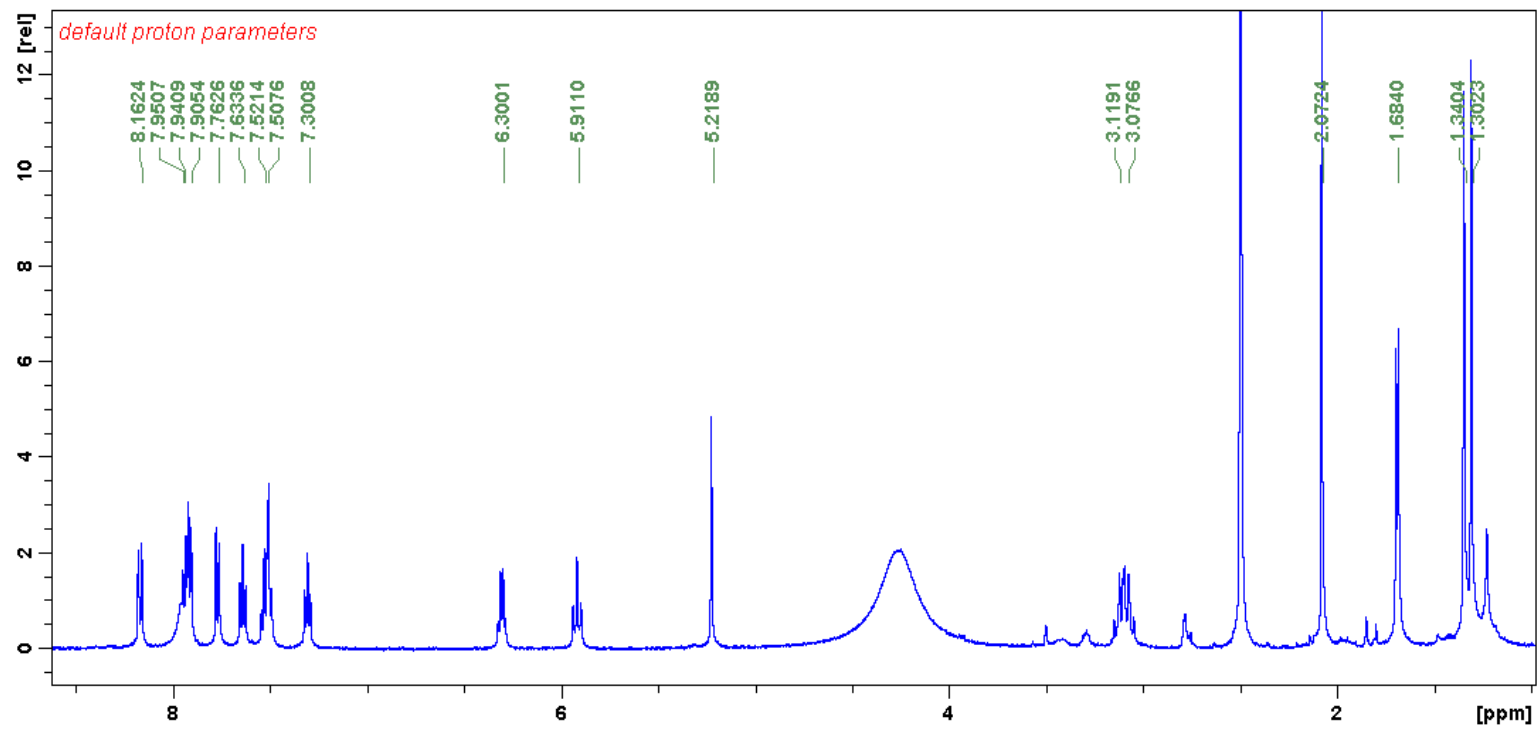
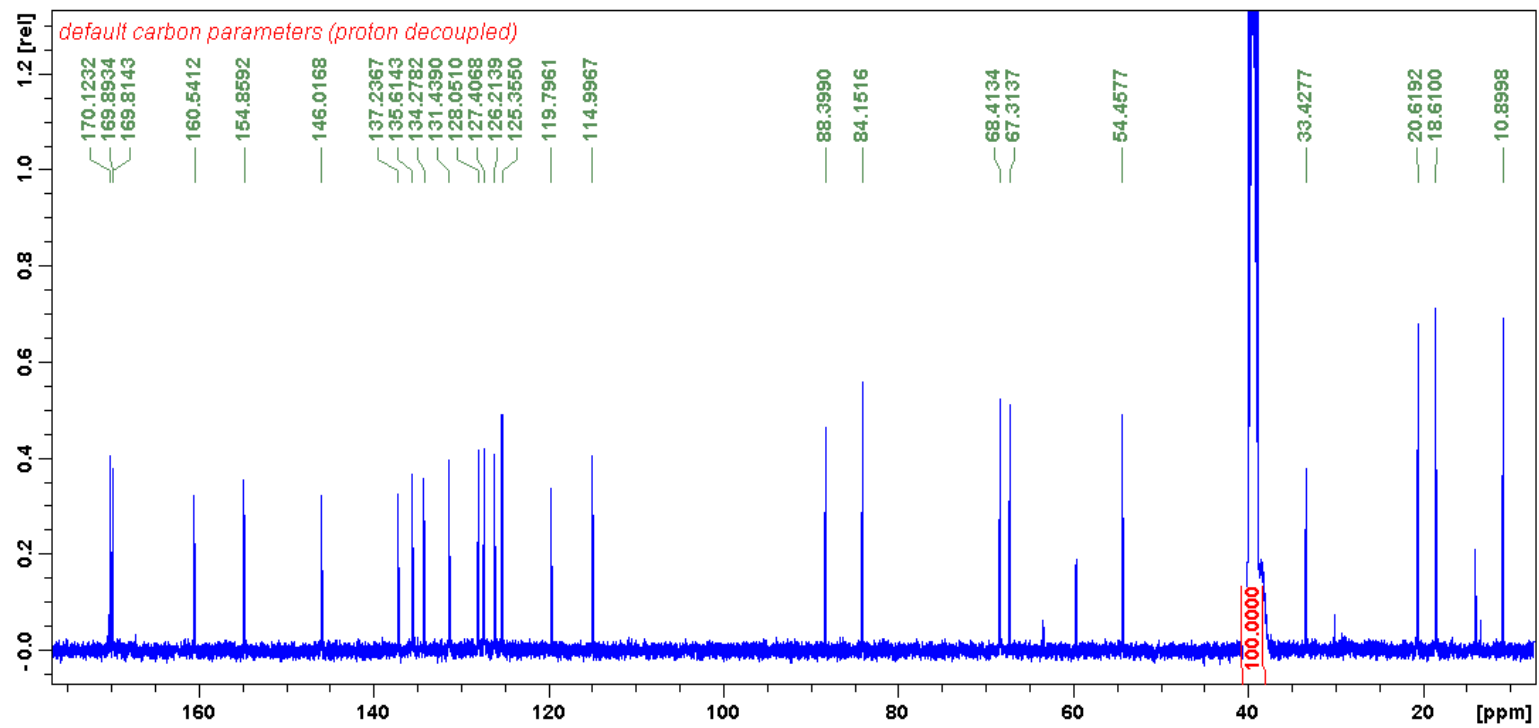**Figure S6.2** ^{1}H NMR spectrum of **1**. Measured in DMSO-d6 at 500MHz
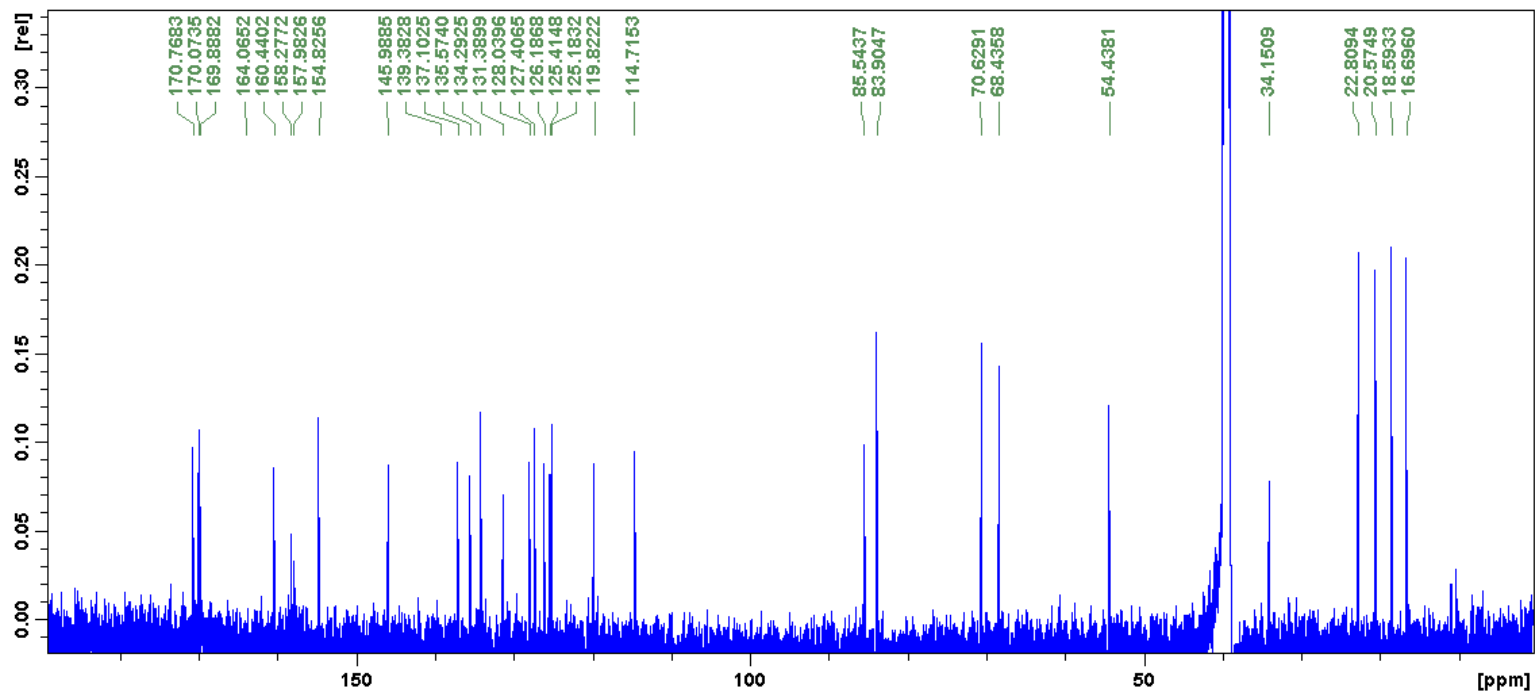
**Figure S6. 3** ^{13}C NMR spectrum of **1**. Measured in DMSO-d6 at 125MHz
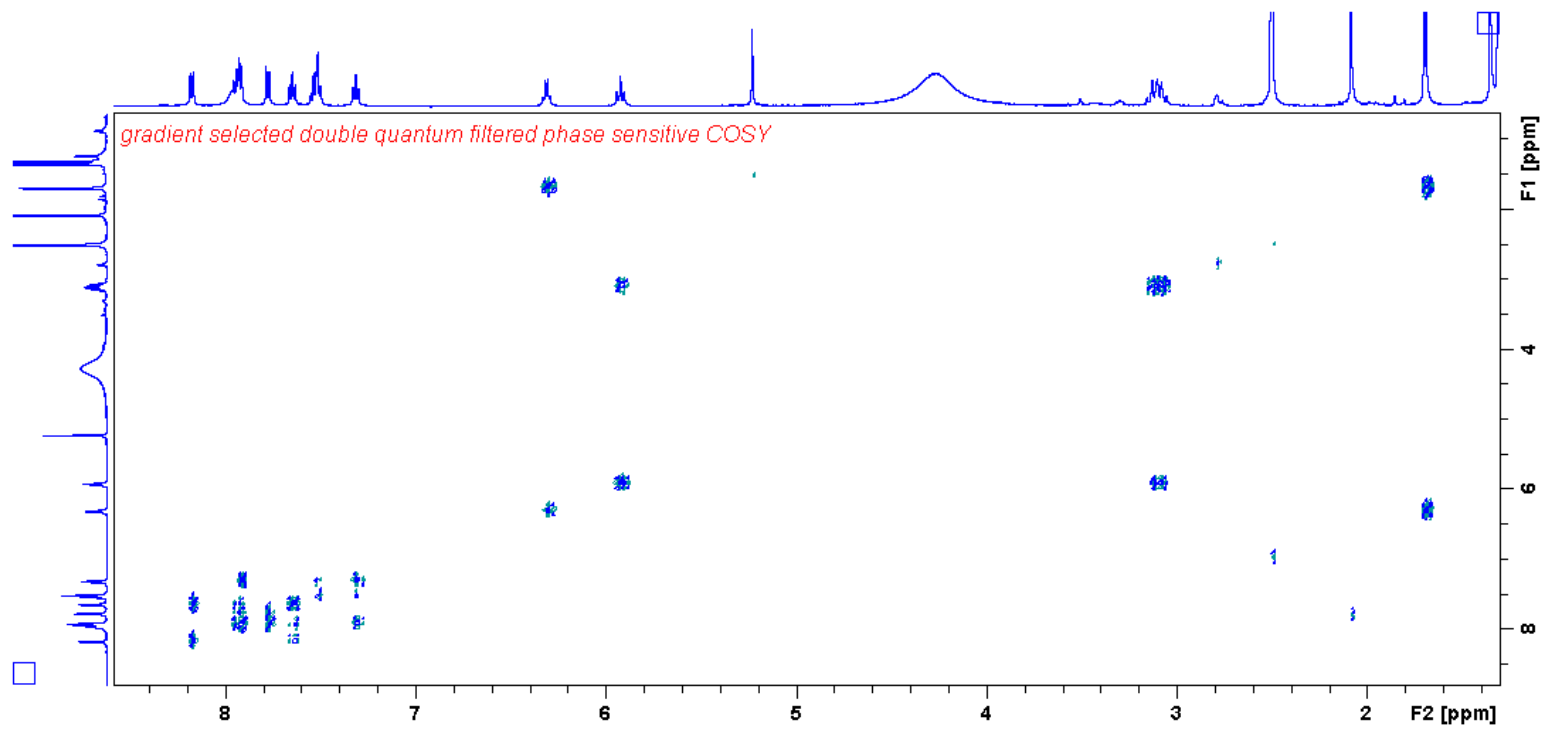
Figure S6.4 ^{1}H-^{1}H COSY spectrum of **1**.
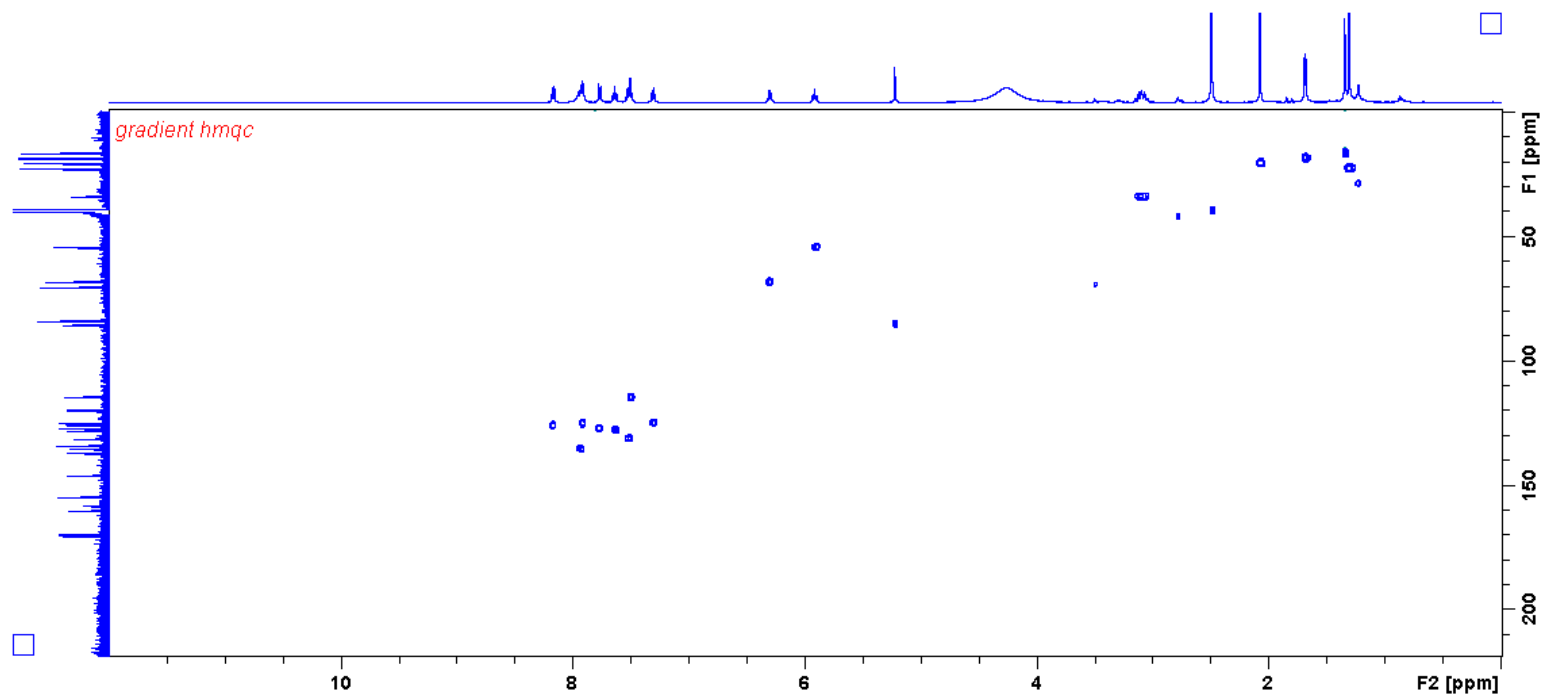
Figure S6. 5 HMQC spectrum of **1**.
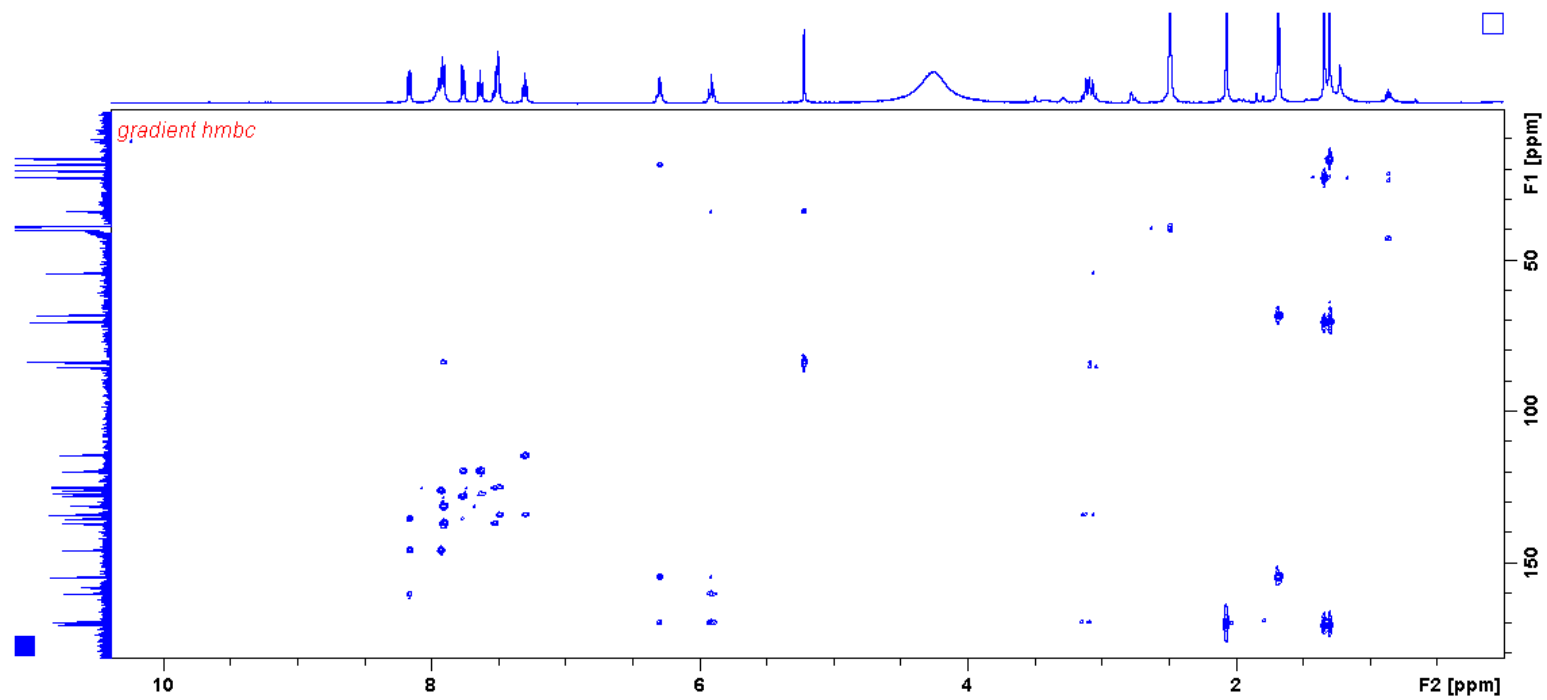
Figure S6.6 HMBC spectrum of **1**.

mAU(x1,000)

204
216
272
427
513

fumiquinazoline F **14**
Exact Mass: 358.14

RT=22.8 min

T: + c ESI MS [200.00-1200.00]

Inten. (x100,000)

359
381
228 248 268 305 332 397 413 443 466 491 517 567 614 656 673 693
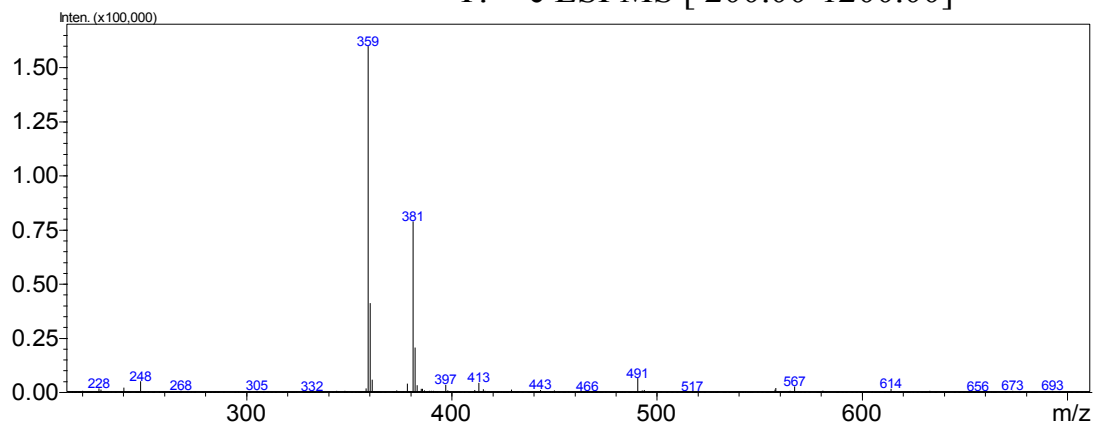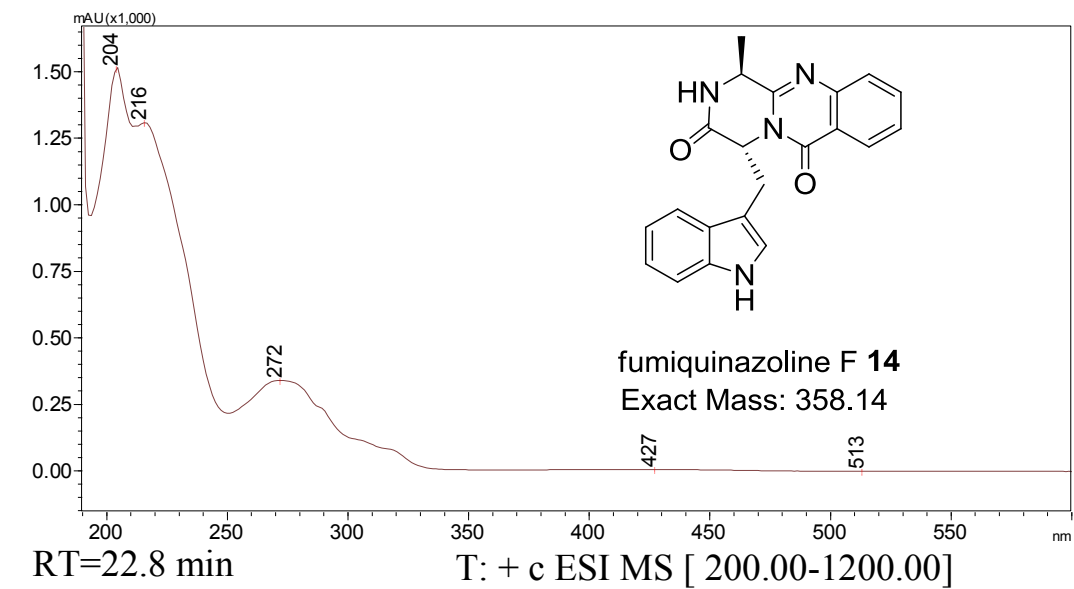
Figure S7. 1 UV spectrum and MS measured during LC-MS for **14**.
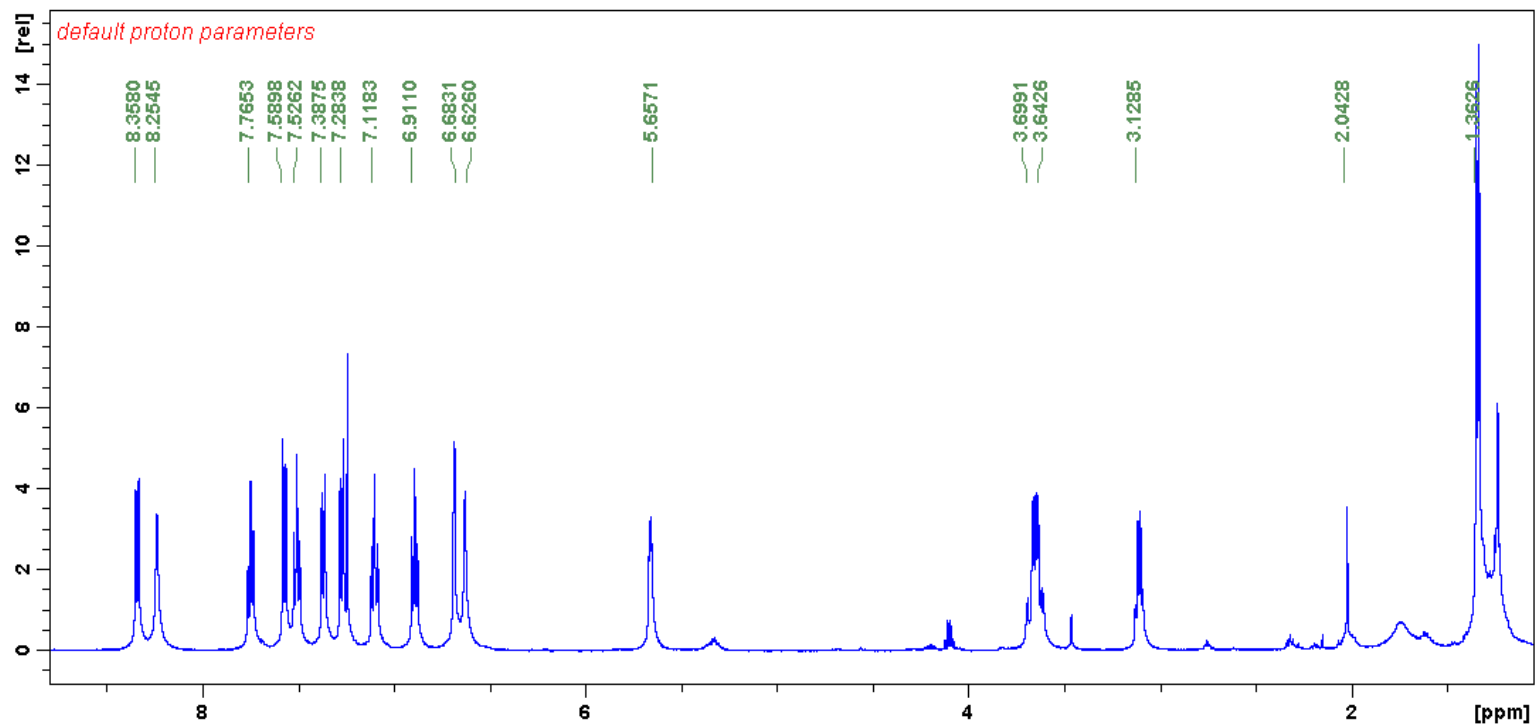
Figure S7.2 ^{1}H NMR spectrum of **14**. Measured in CDCl3 at 500MHz.
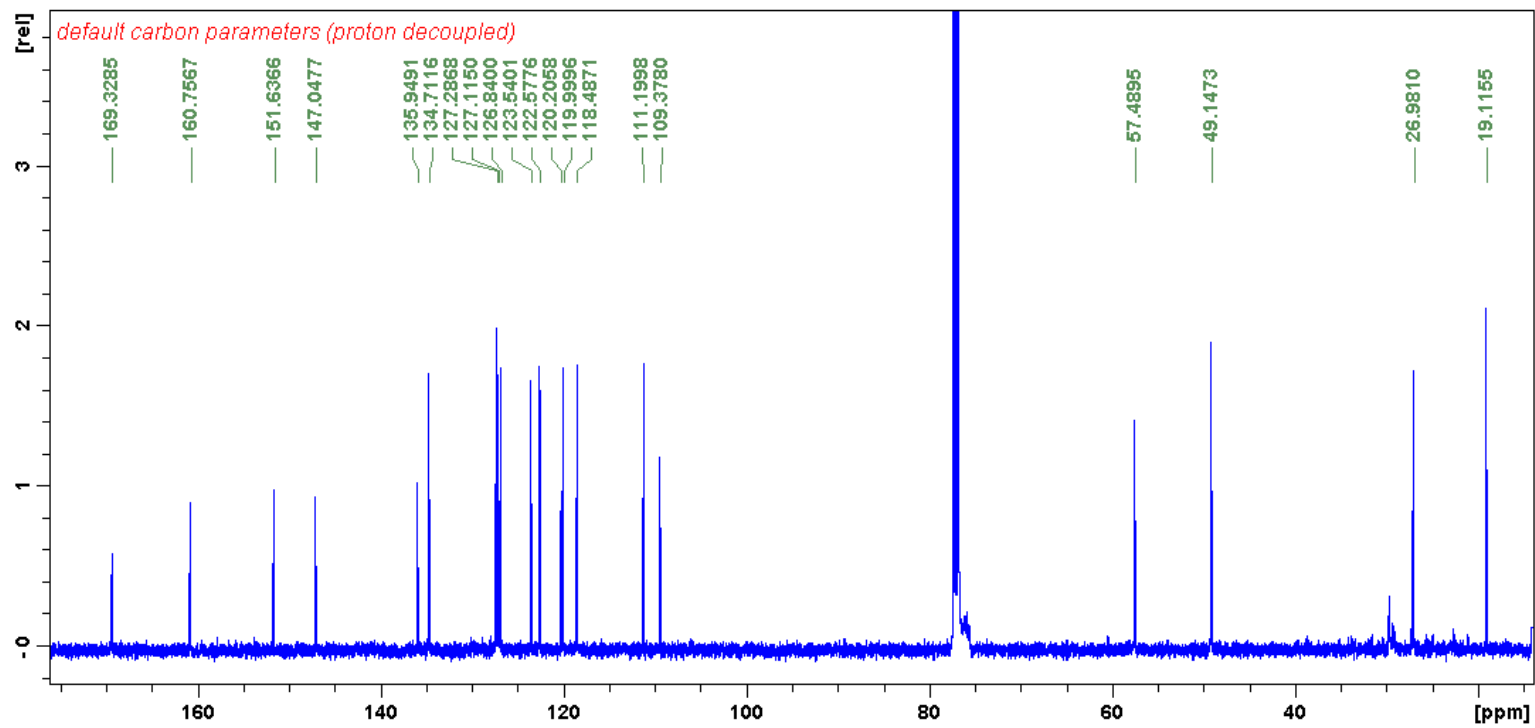
Figure S7.3 ¹³C NMR spectrum of **14**. Measured in CDCl3 at 125MHz.
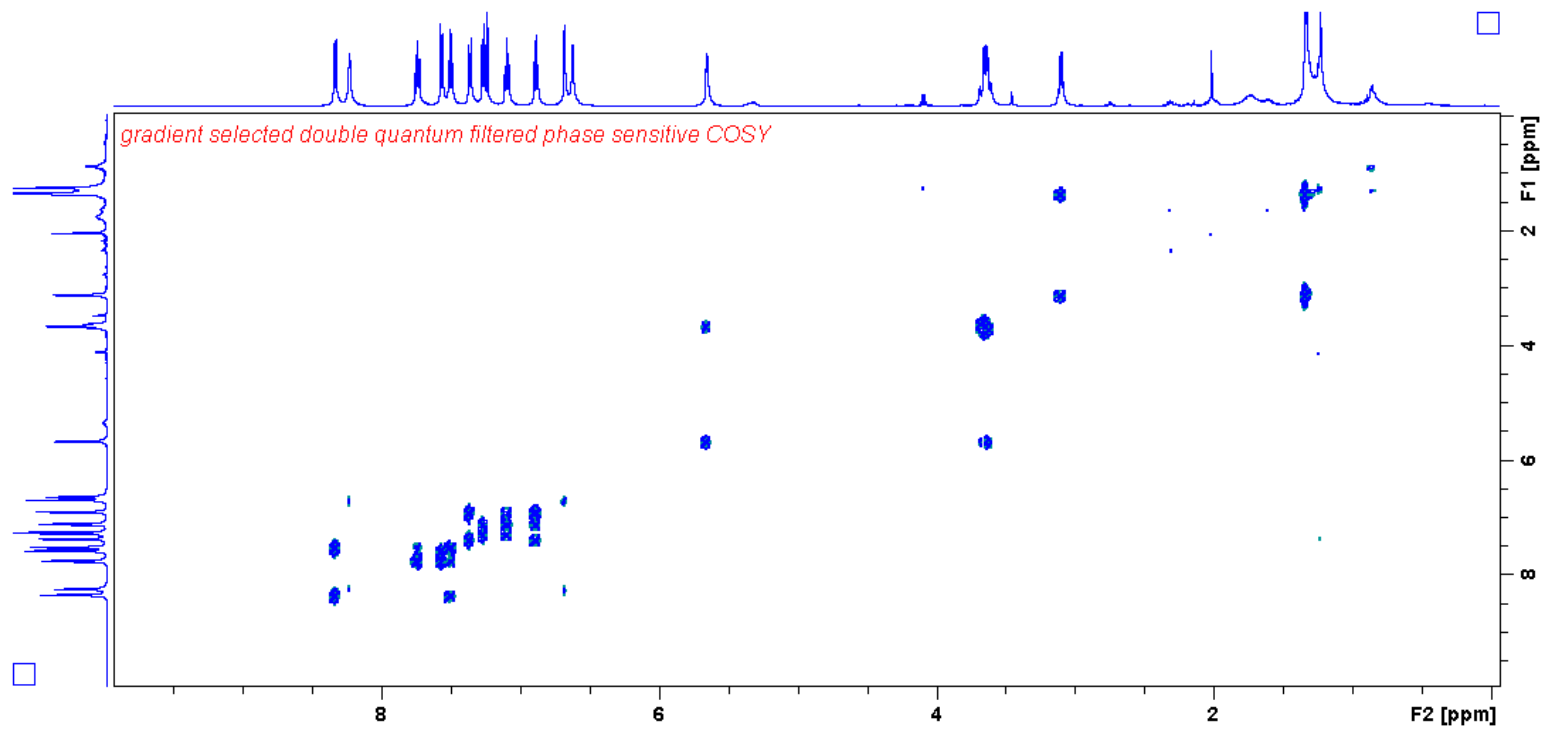
Figure S7.4 [^1]H-[^1]H COSY NMR spectrum of **14**.
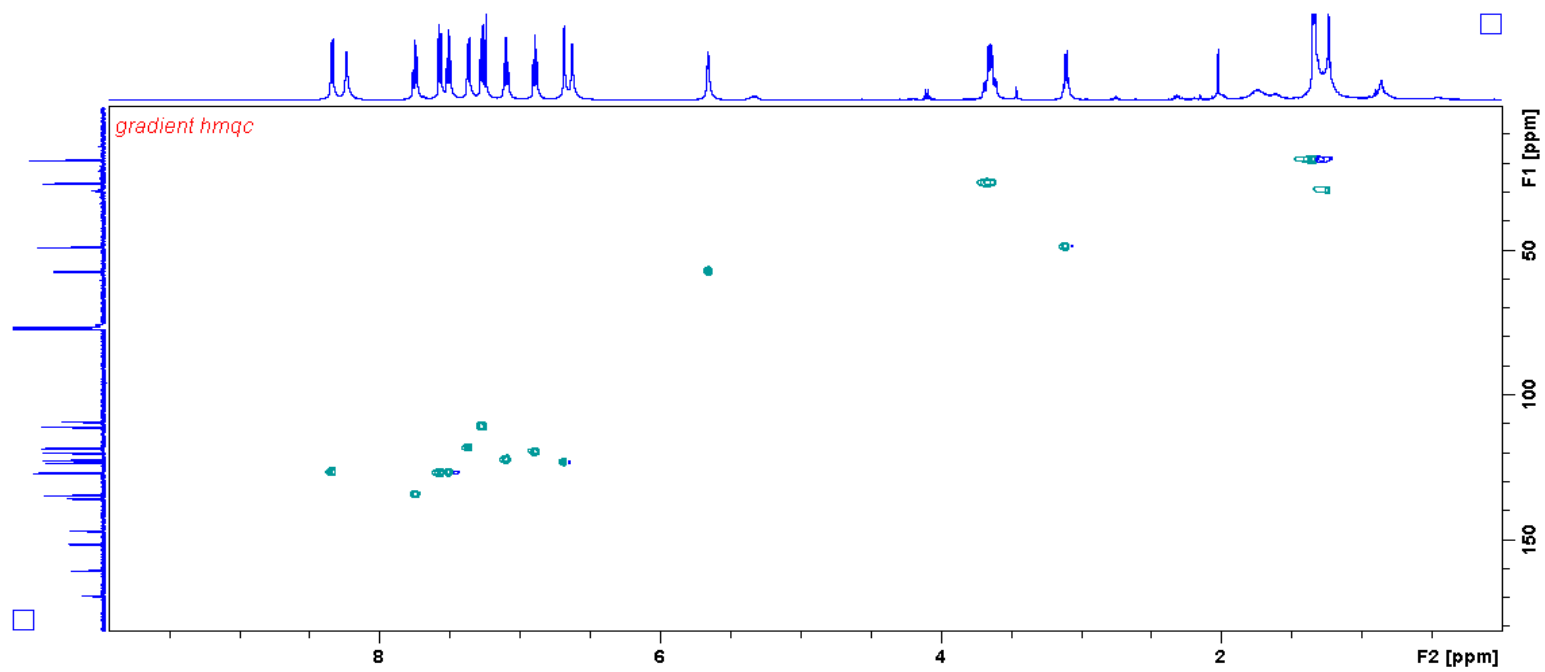
Figure S7.5 ^{1}H-^{13}C HMQC NMR spectrum of **14**.
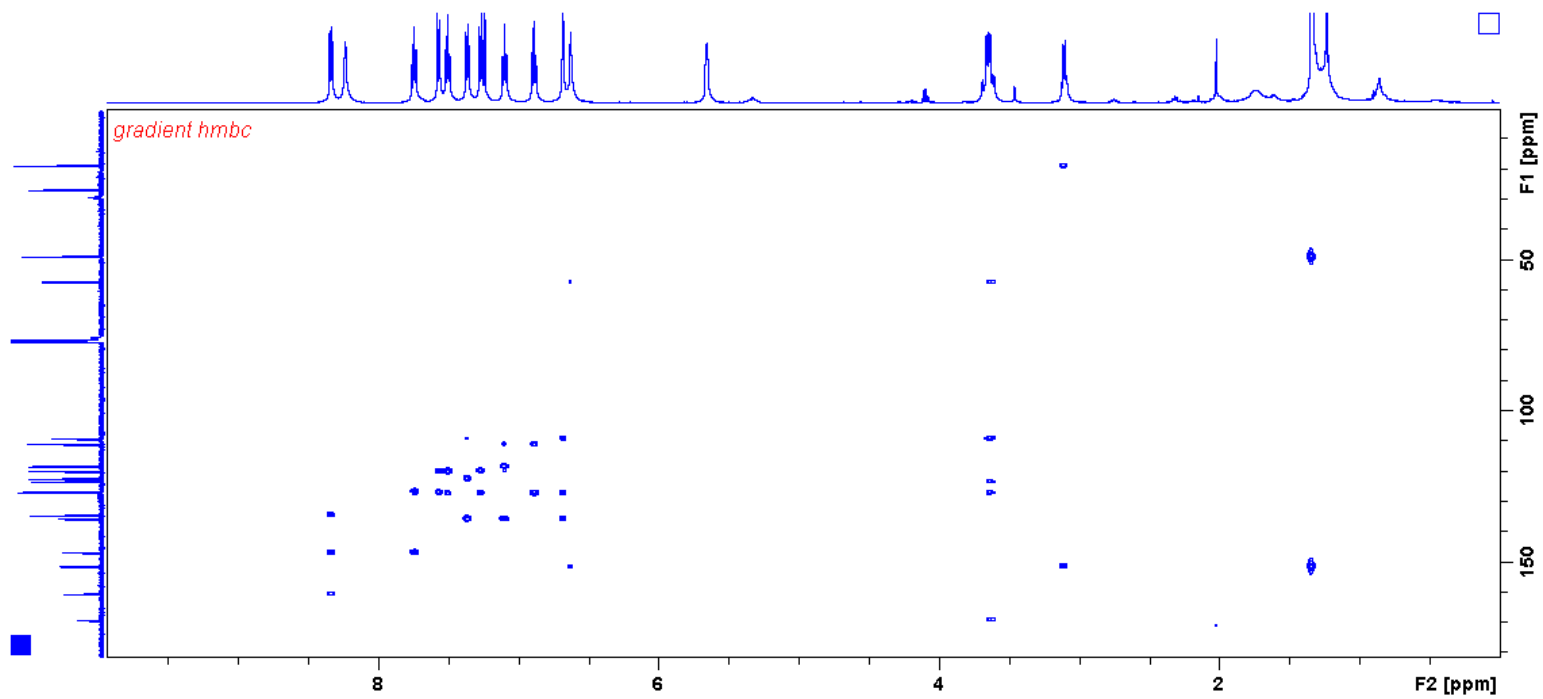
Figure S7.6 ^{1}H-^{13}C HMBC NMR spectrum of **14**.
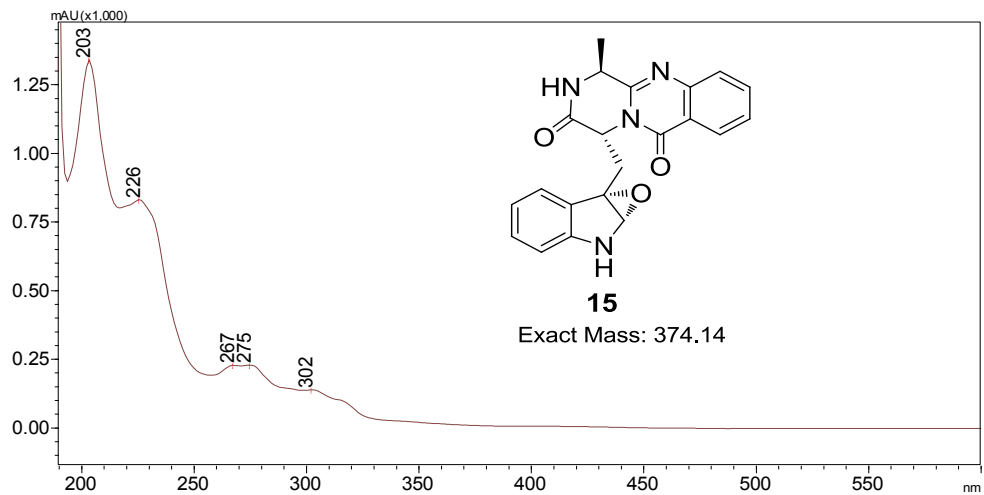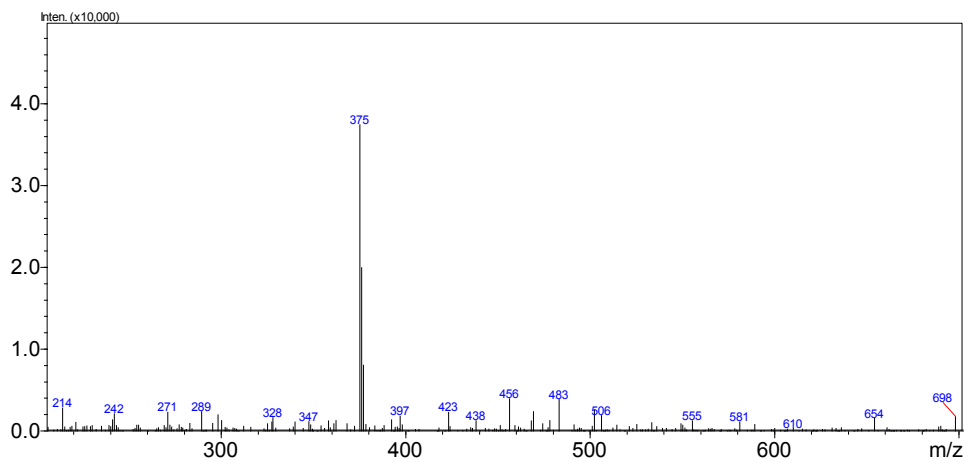
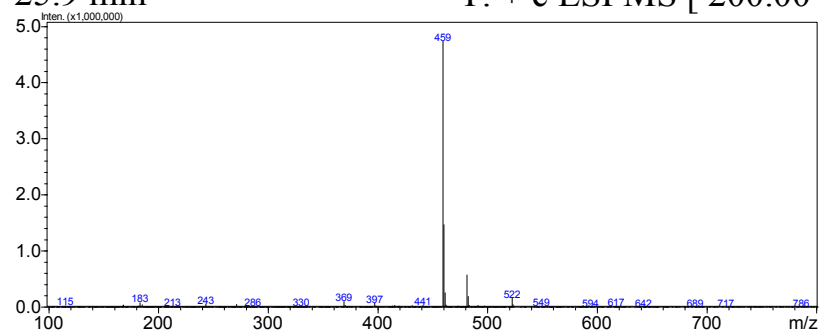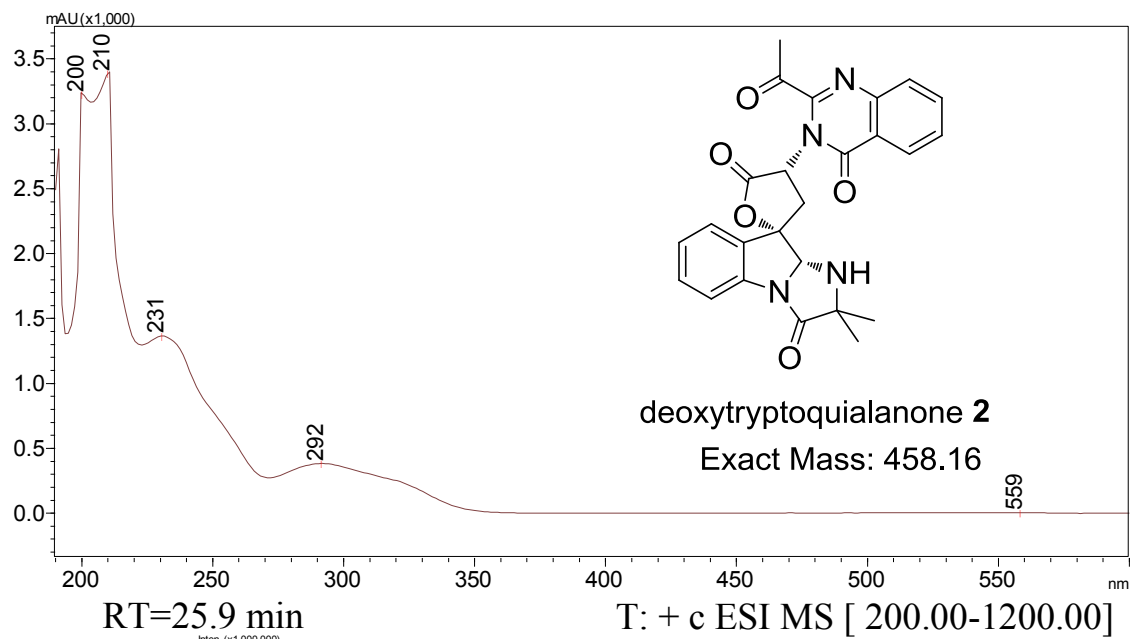**Figure S8.1** UV spectrum and MS measured during LC-MS for **15**
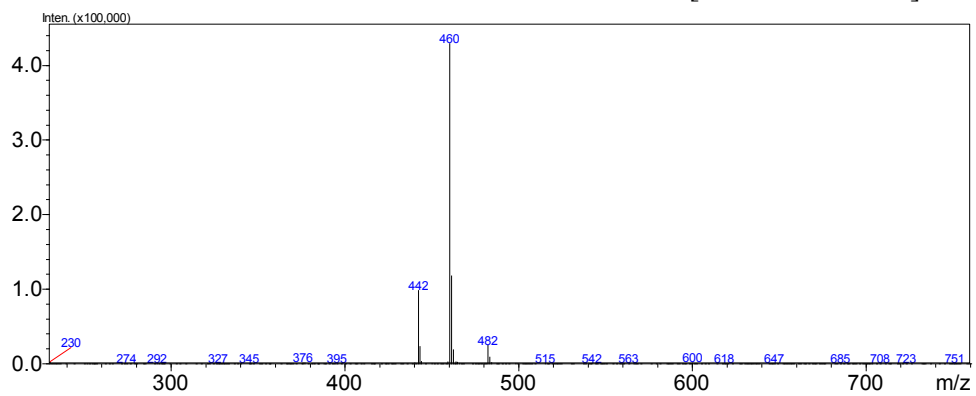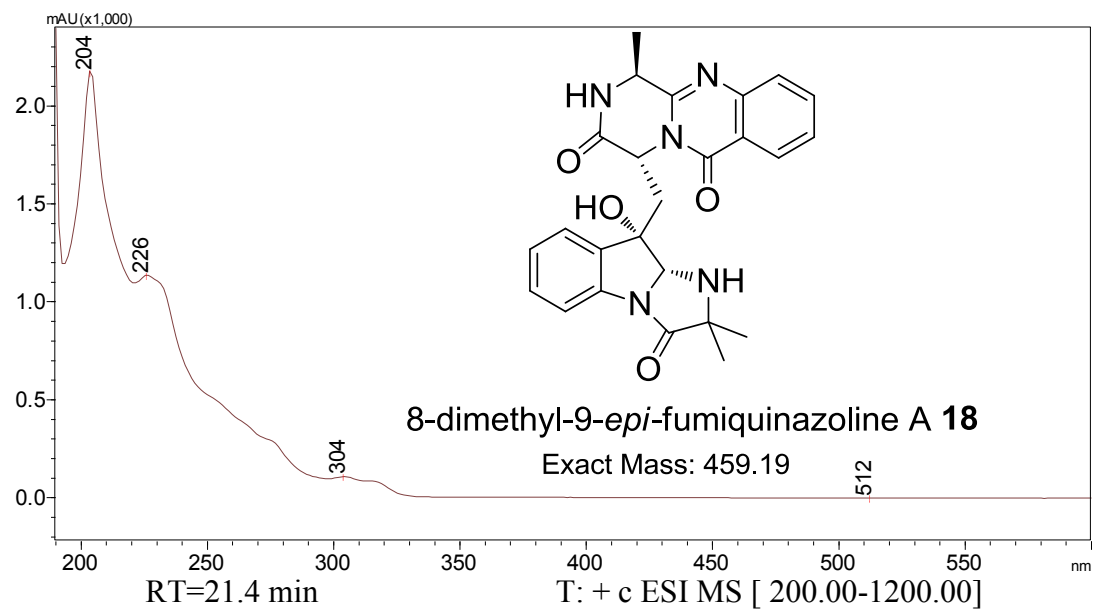
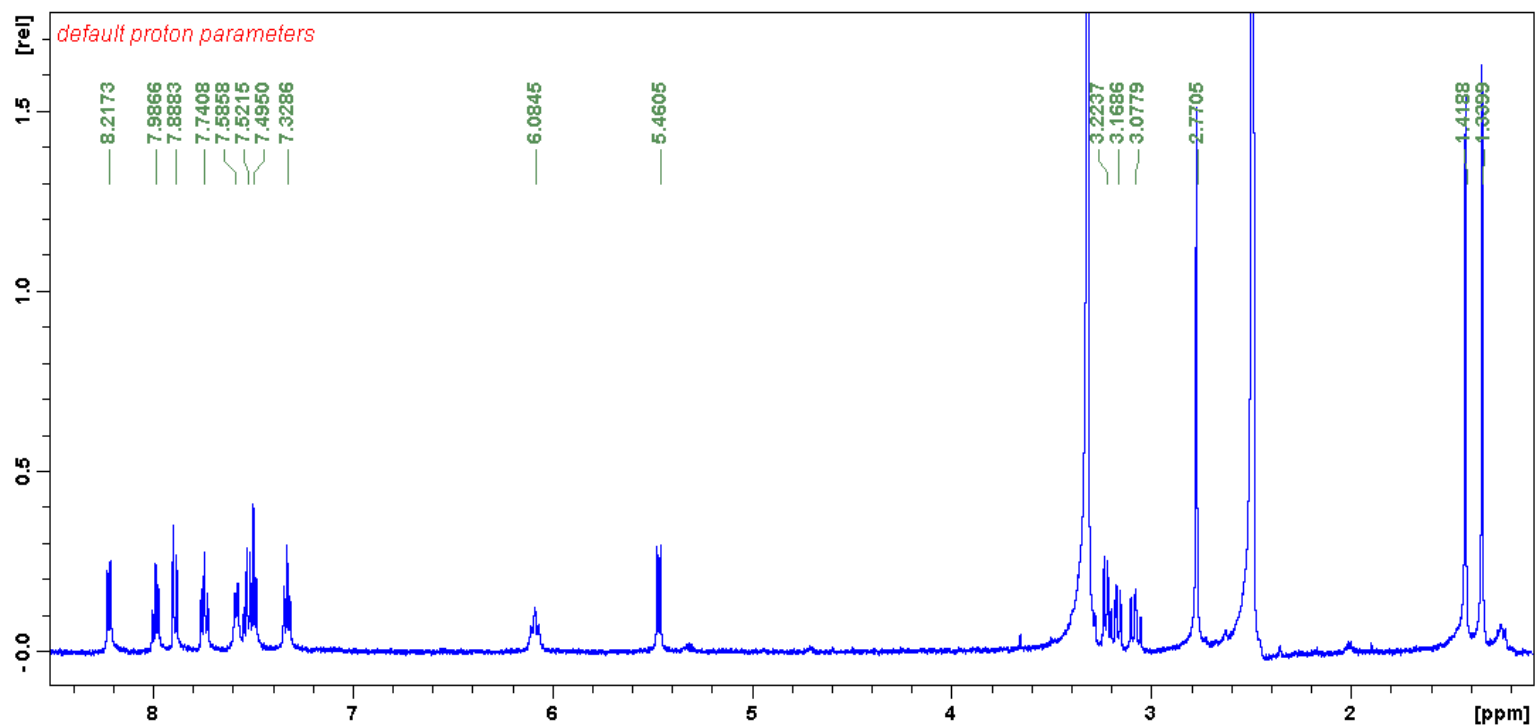**Figure S9.1** UV spectrum and MS measured during LC-MS for **18**
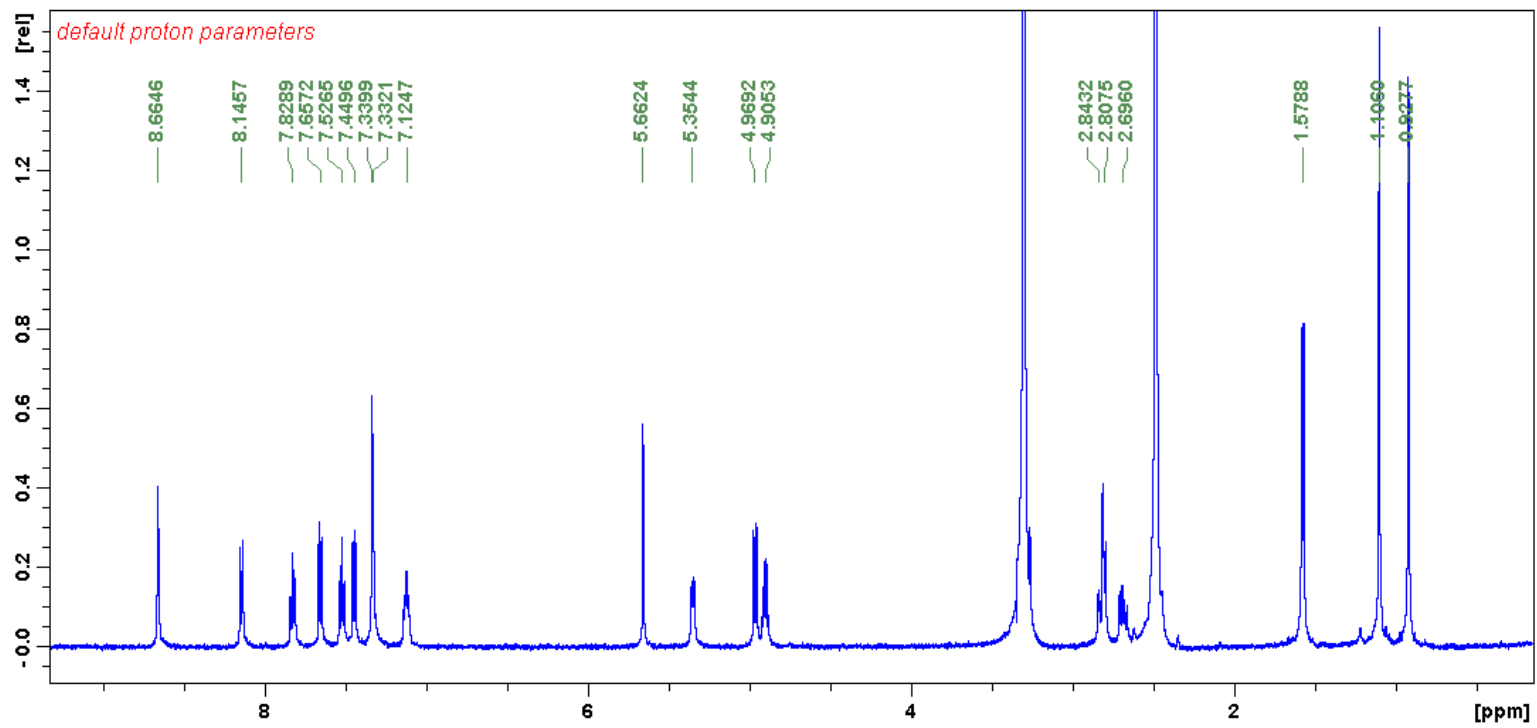
Figure S9.2 ^{1}H NMR spectrum of **18**. Measured in DMSO-d_6 at 500MHz.
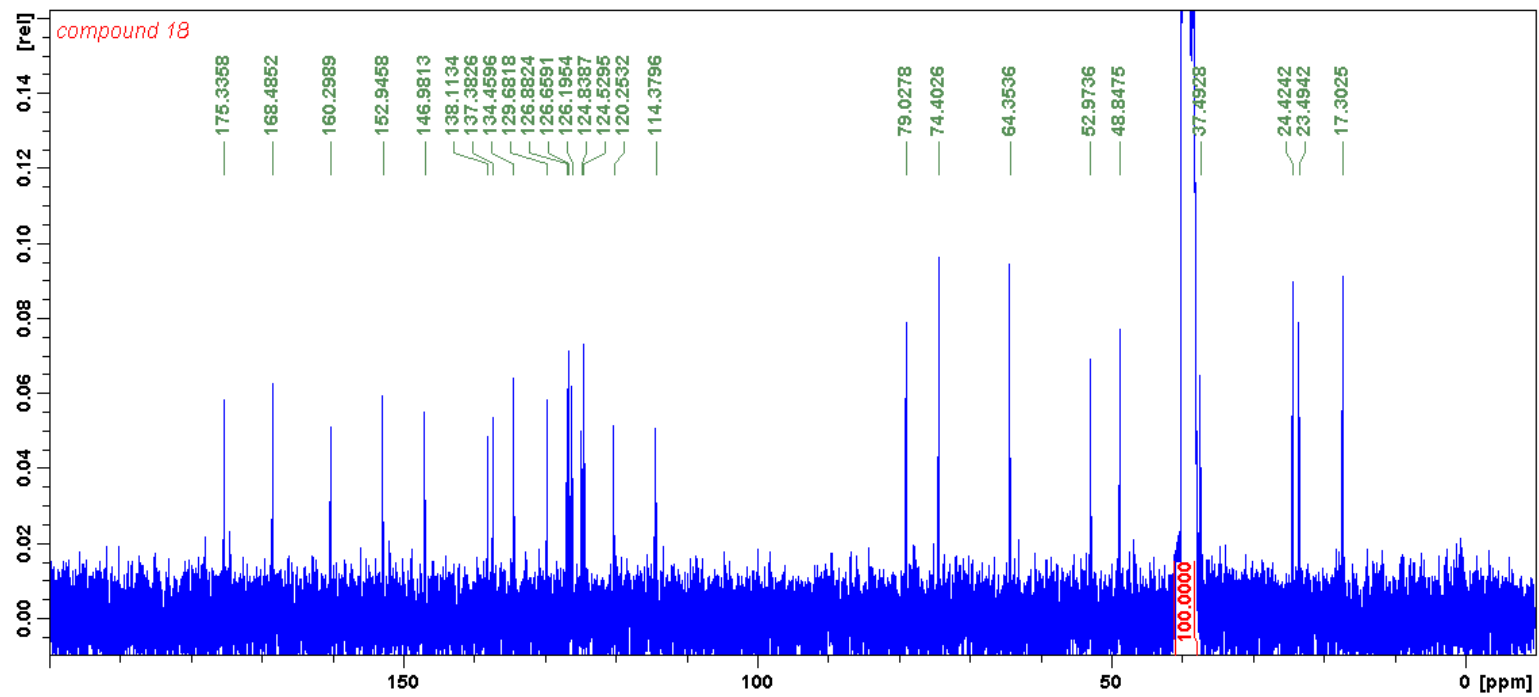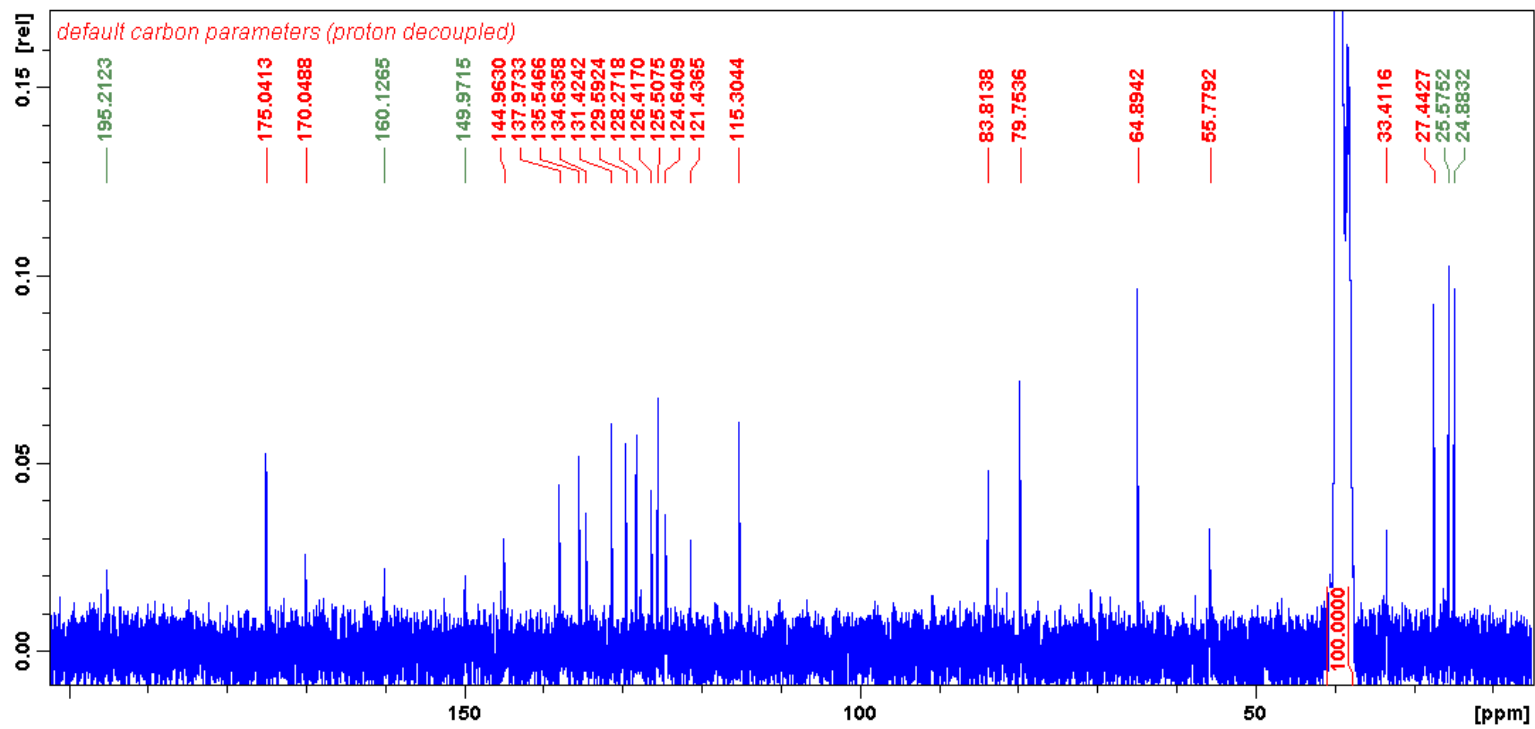
Figure S9.3 ^{13}C NMR spectrum of **18**. Measured in DMSO-d_6 at 125MH.
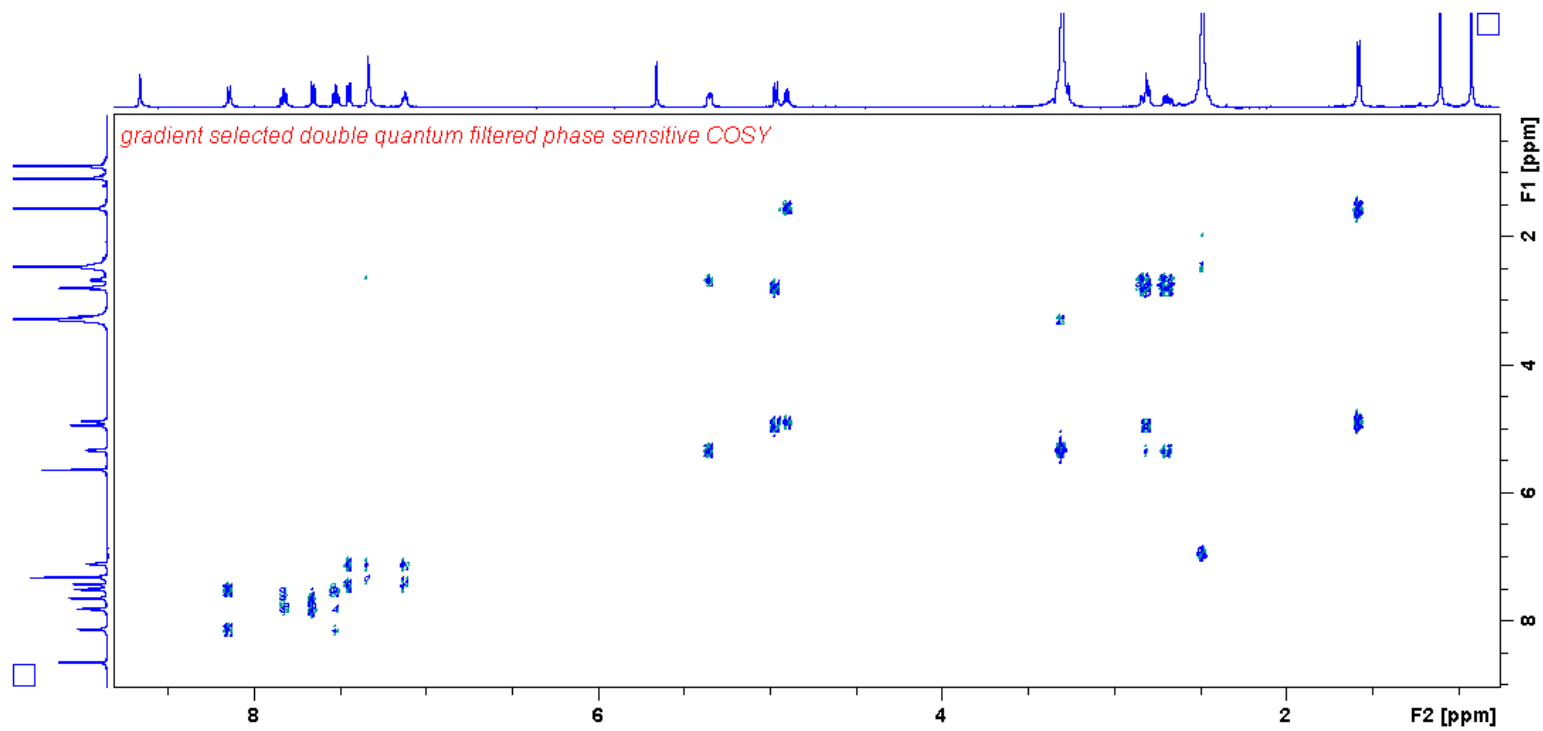
Figure S9.4 ^{1}H-^{1}H COSY NMR spectrum of **18**.
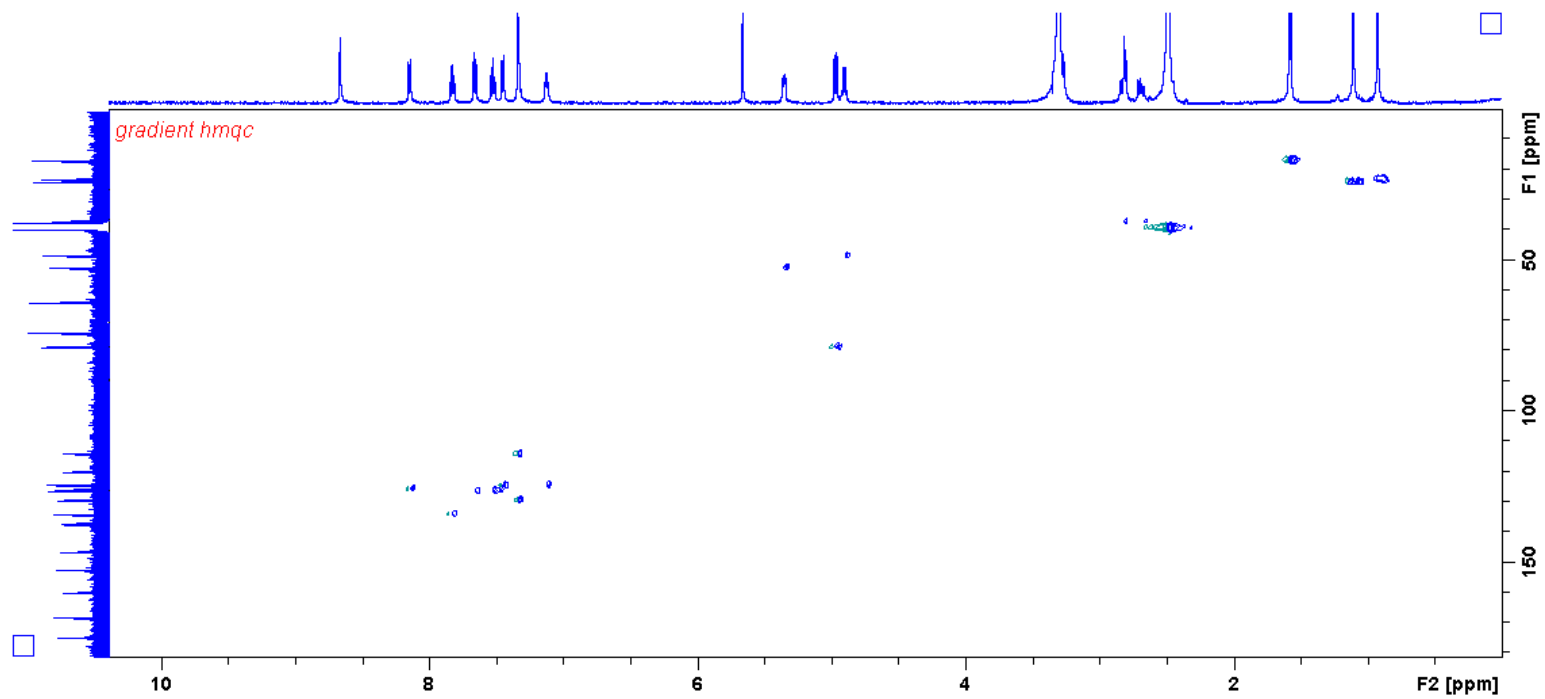
Figure S9.5 ^{1}H-^{13}C HMQC NMR spectrum of **18**.
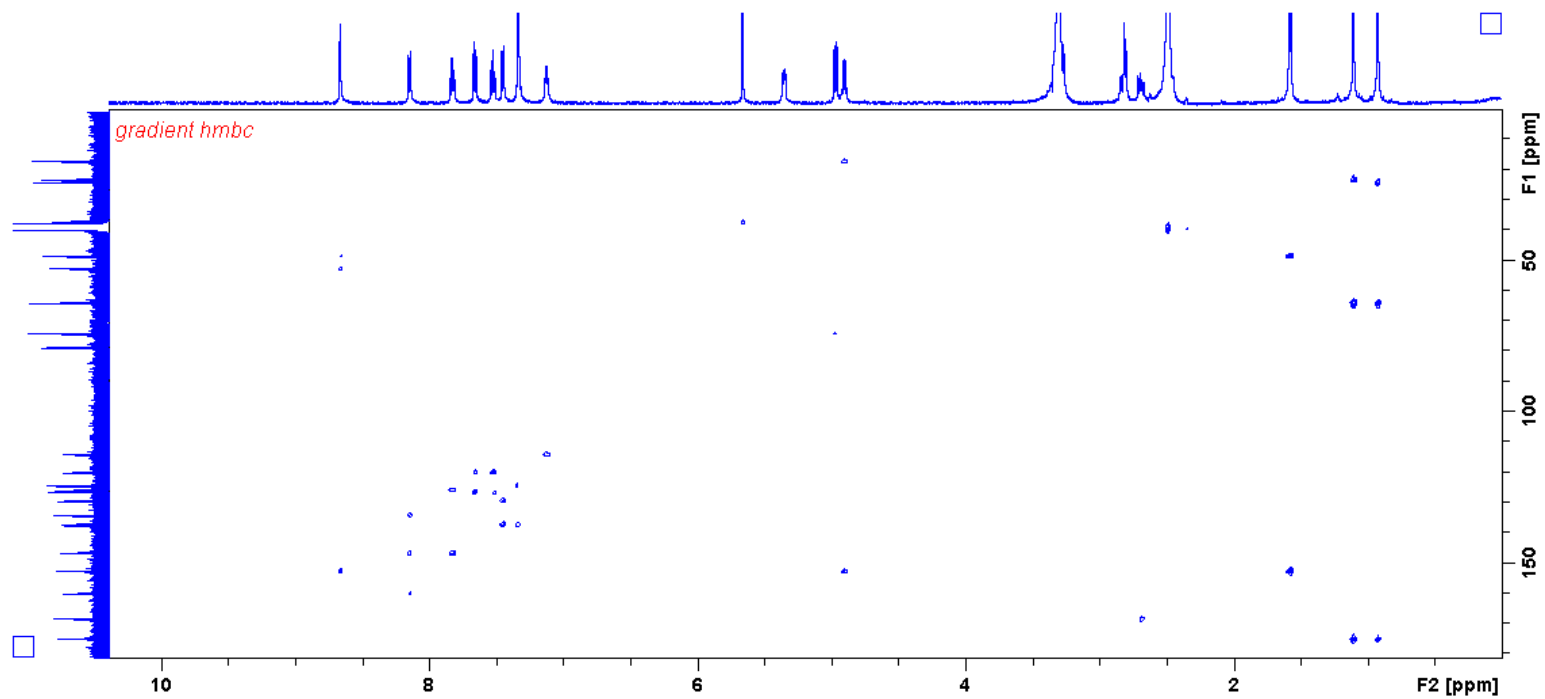
Figure S9.6 ^{1}H-^{13}C HMBC NMR spectrum of **18**.
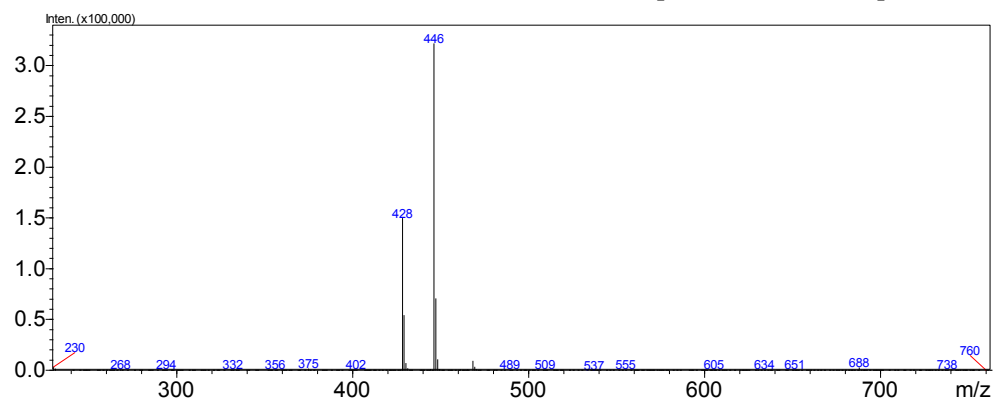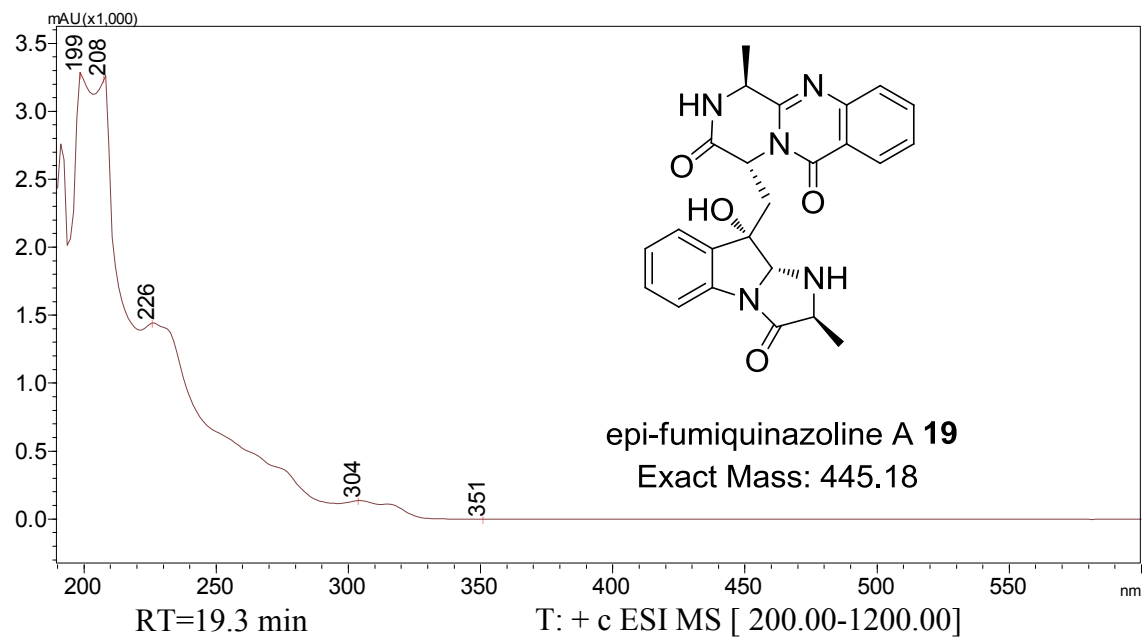
Figure S10.1 UV spectrum and MS measured during LC-MS for **19**.

epi-fumiquinazoline A **19**
Exact Mass: 445.18

RT=19.3 min

T: + c ESI MS [200.00-1200.00]
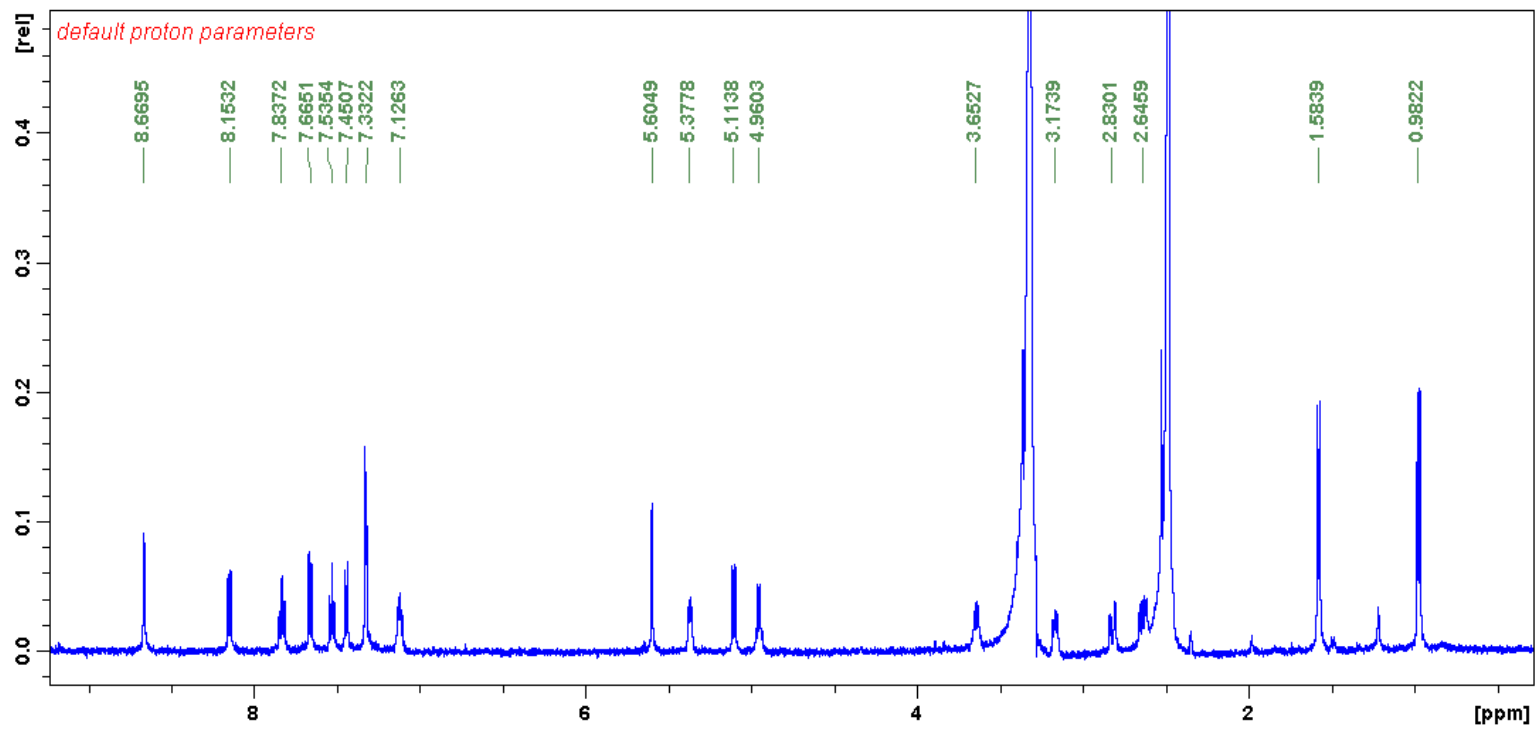
Figure S10.2 ^{1}H NMR spectrum of **19**. Measured in DMSO-d_6 at 500MHz.
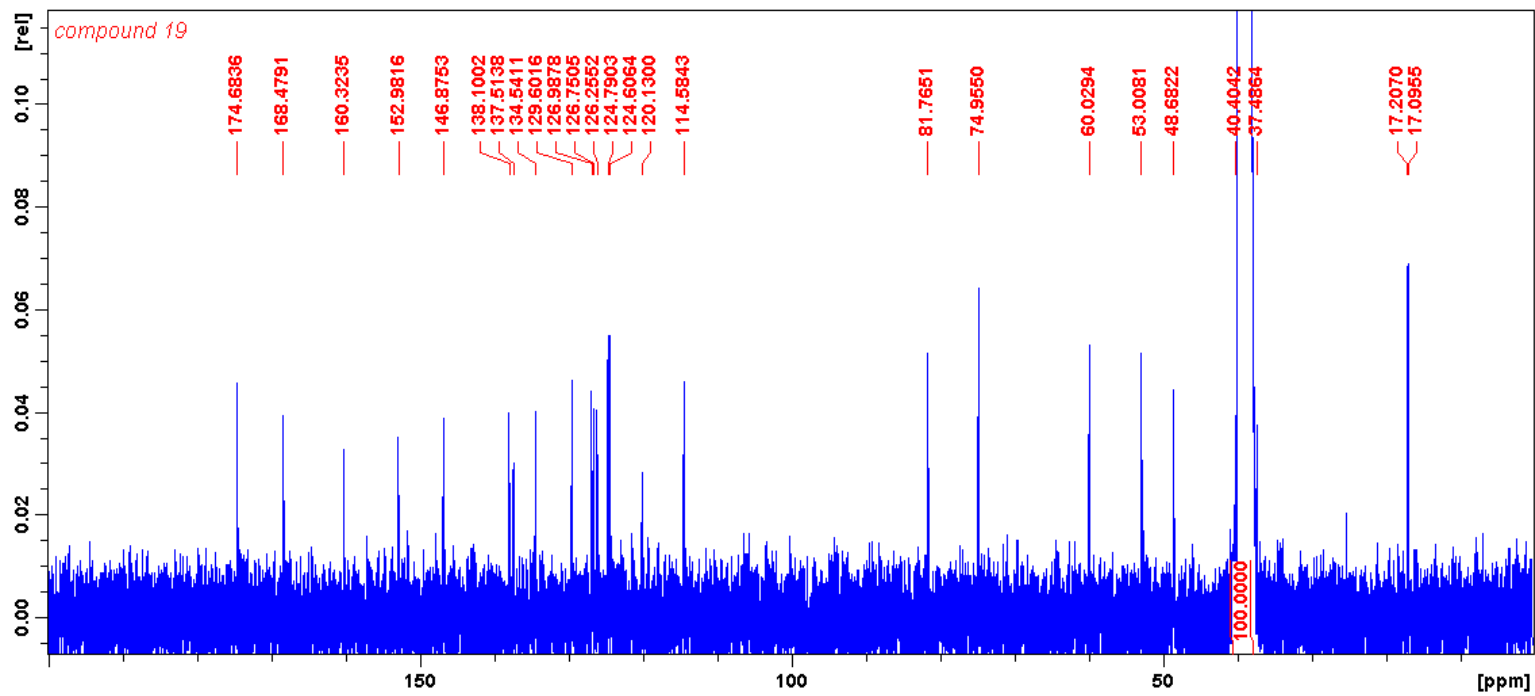
Figure S10.3 ^{13}C NMR spectrum of **19**. Measured in DMSO-d_6 at 125MHz.
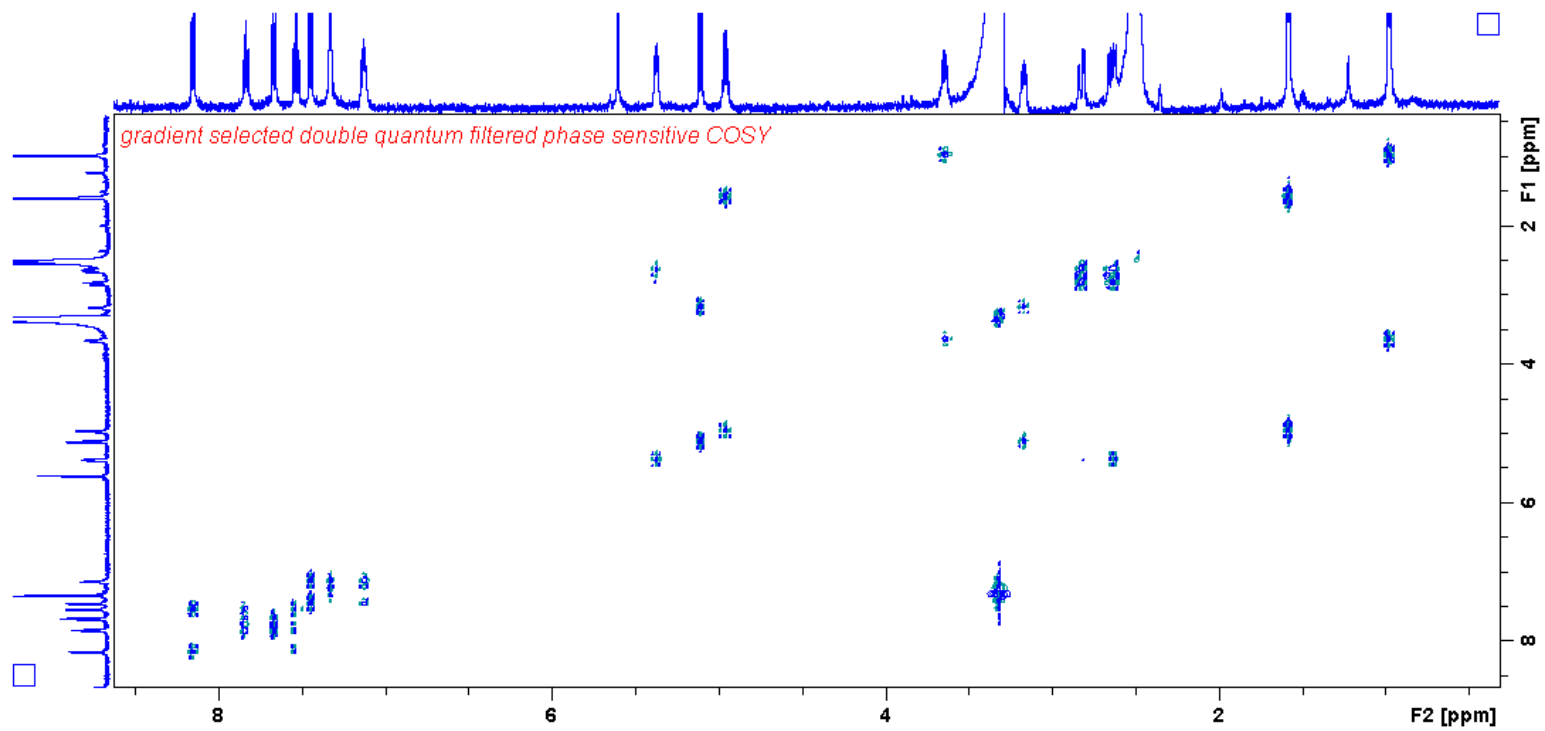
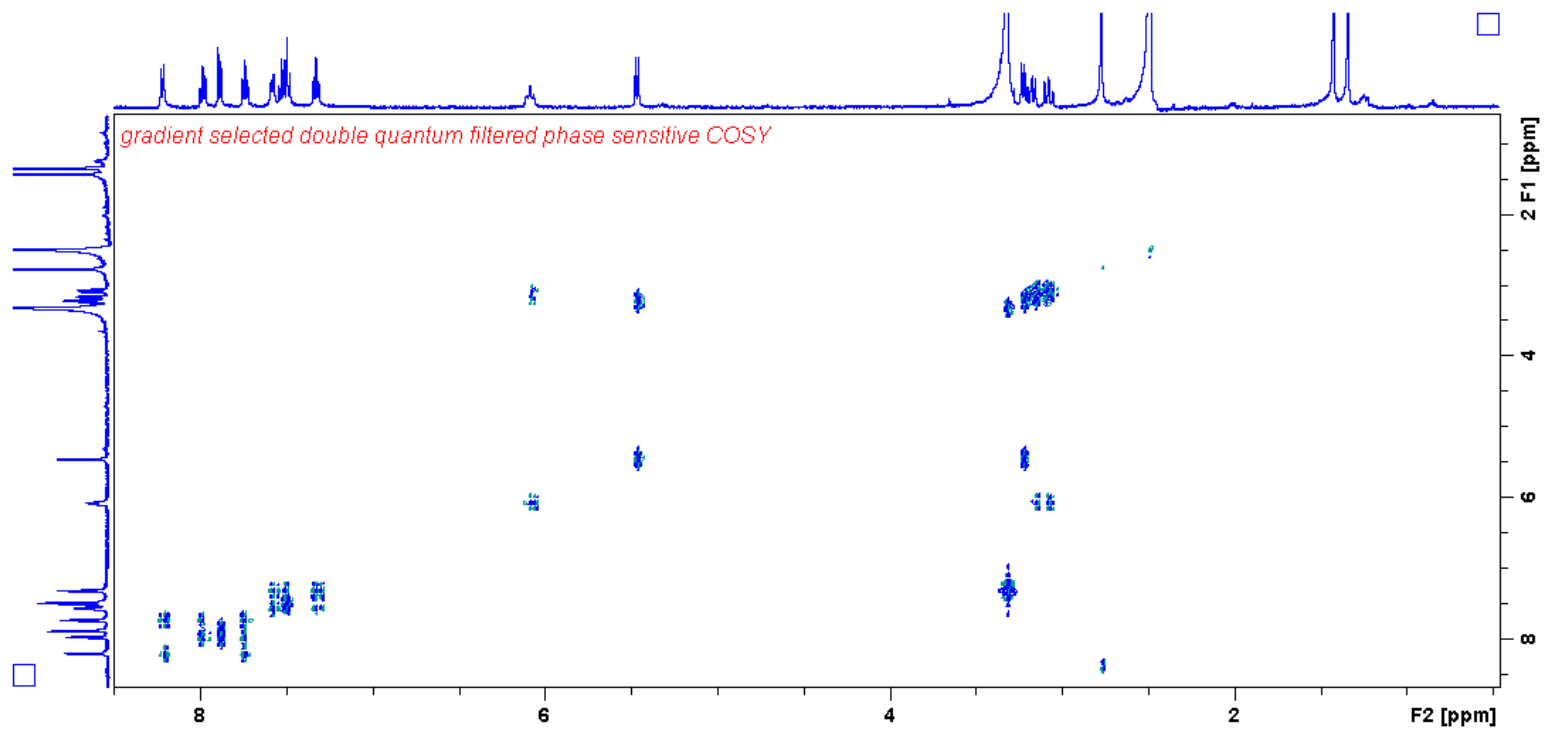**Figure S10.4** ^{1}H-^{1}H COSY NMR spectrum of **19**.
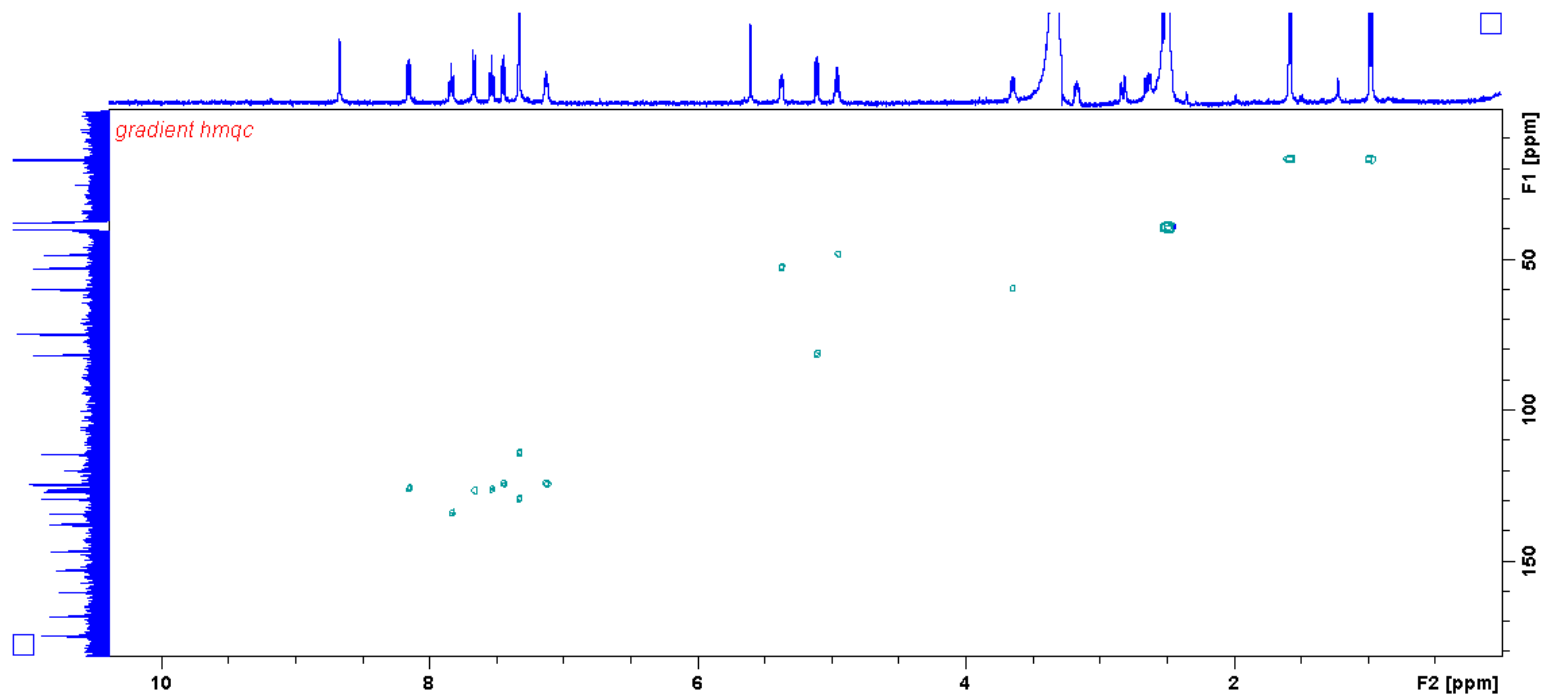
Figure S10.5 ^{1}H-^{13}C HMQC NMR spectrum of **19**.
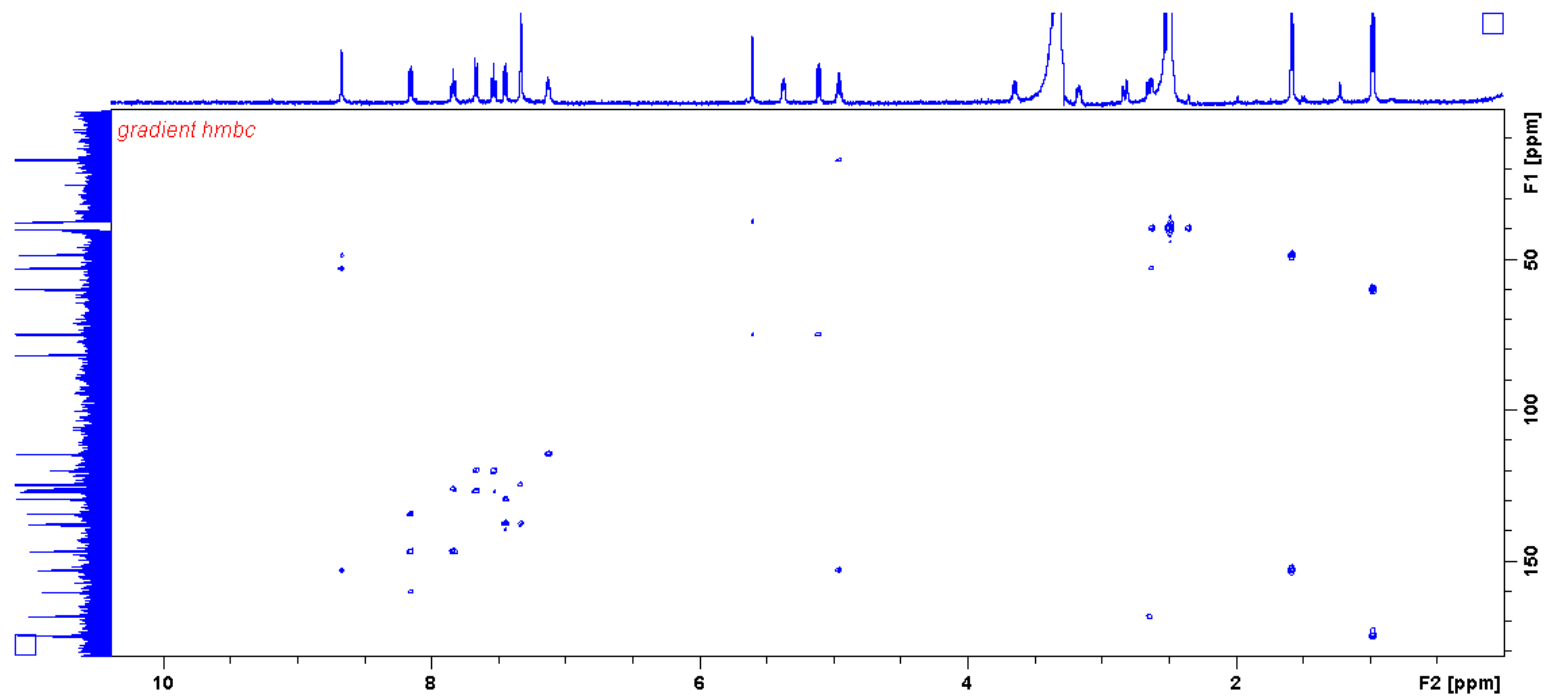
Figure S10.6 ^{1}H-^{13}C HMBC NMR spectrum of **19**.

Nortryptoquialanine **20**

Exact Mass: 504.16

RT=26.1 min

T: + c ESI MS [200.00-1200.00]
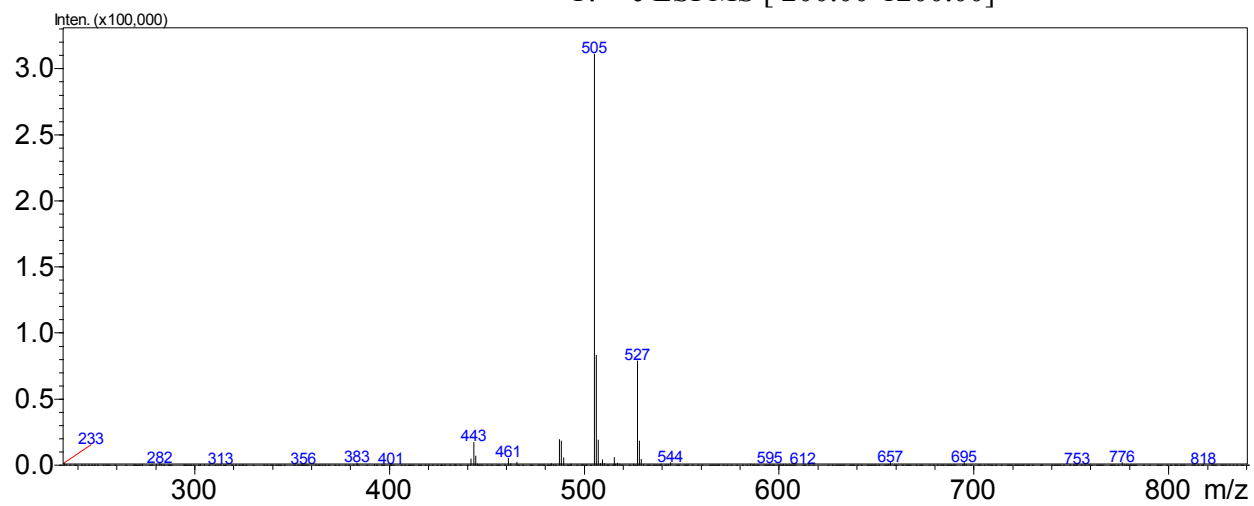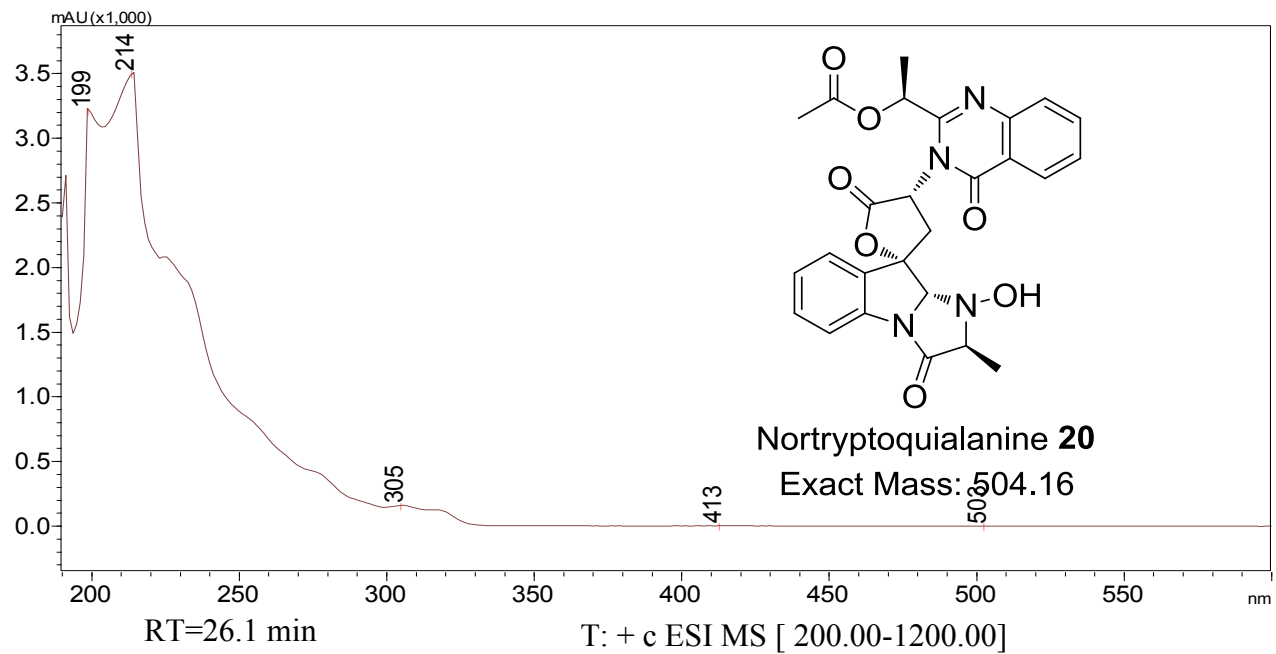
Figure S11.1 UV spectrum and MS measured during LC-MS for **20**.
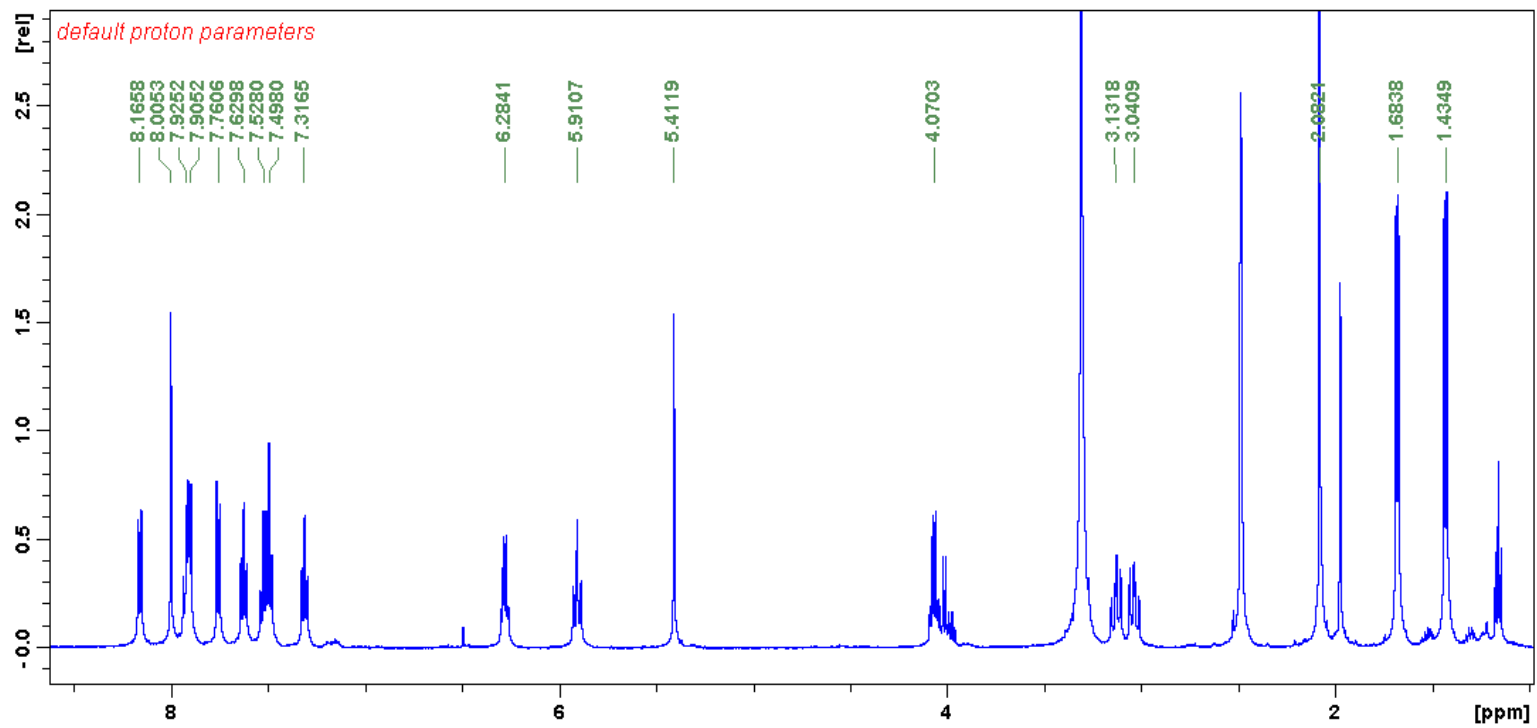
Figure S11.2 ^{1}H NMR spectrum of **20**. Measured in DMSO-d_6 at 500MHz.

Figure S11.3 ^{13}C NMR spectrum of **20**. Measured in DMSO-d_6 at 125MH.
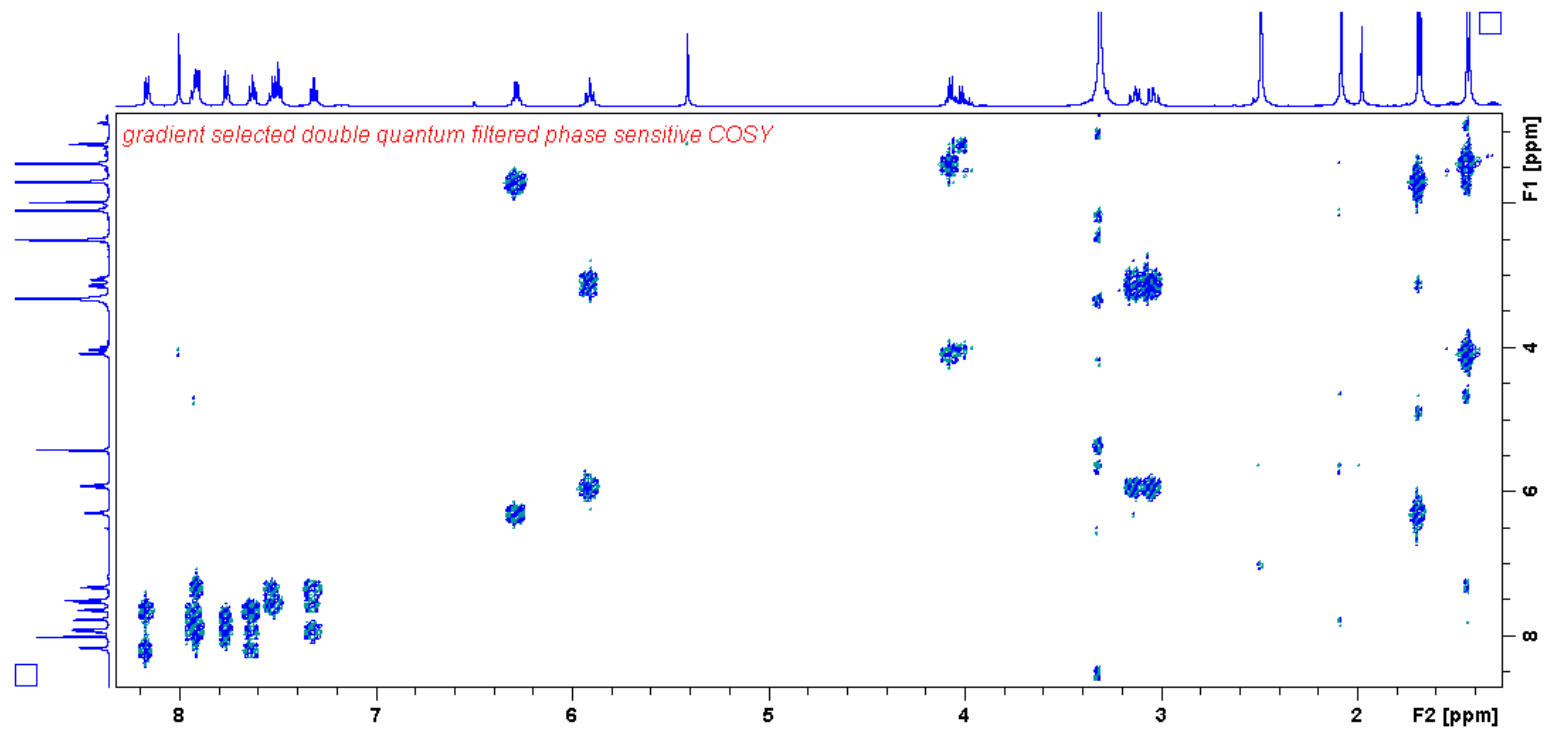
Figure S11.4 ^{1}H-^{1}H COSY NMR spectrum of **20**.
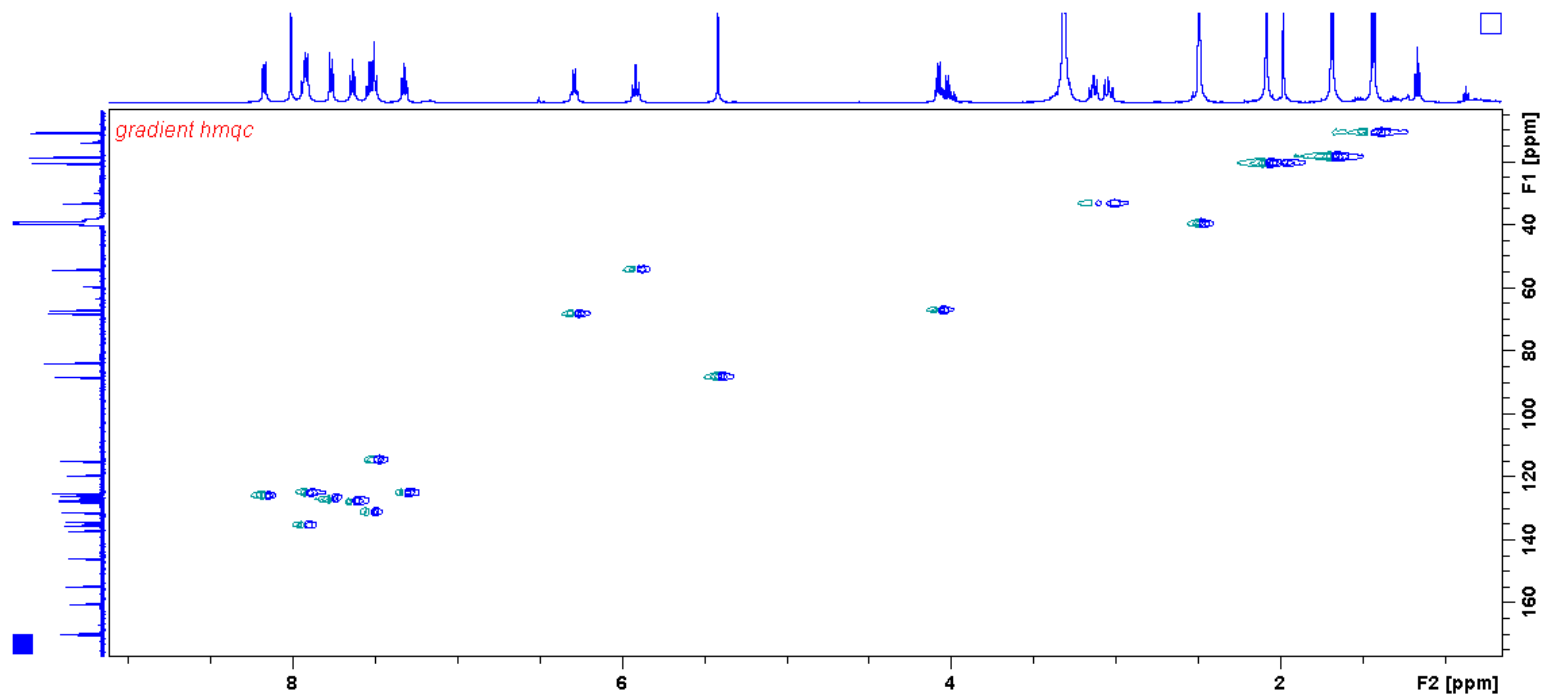
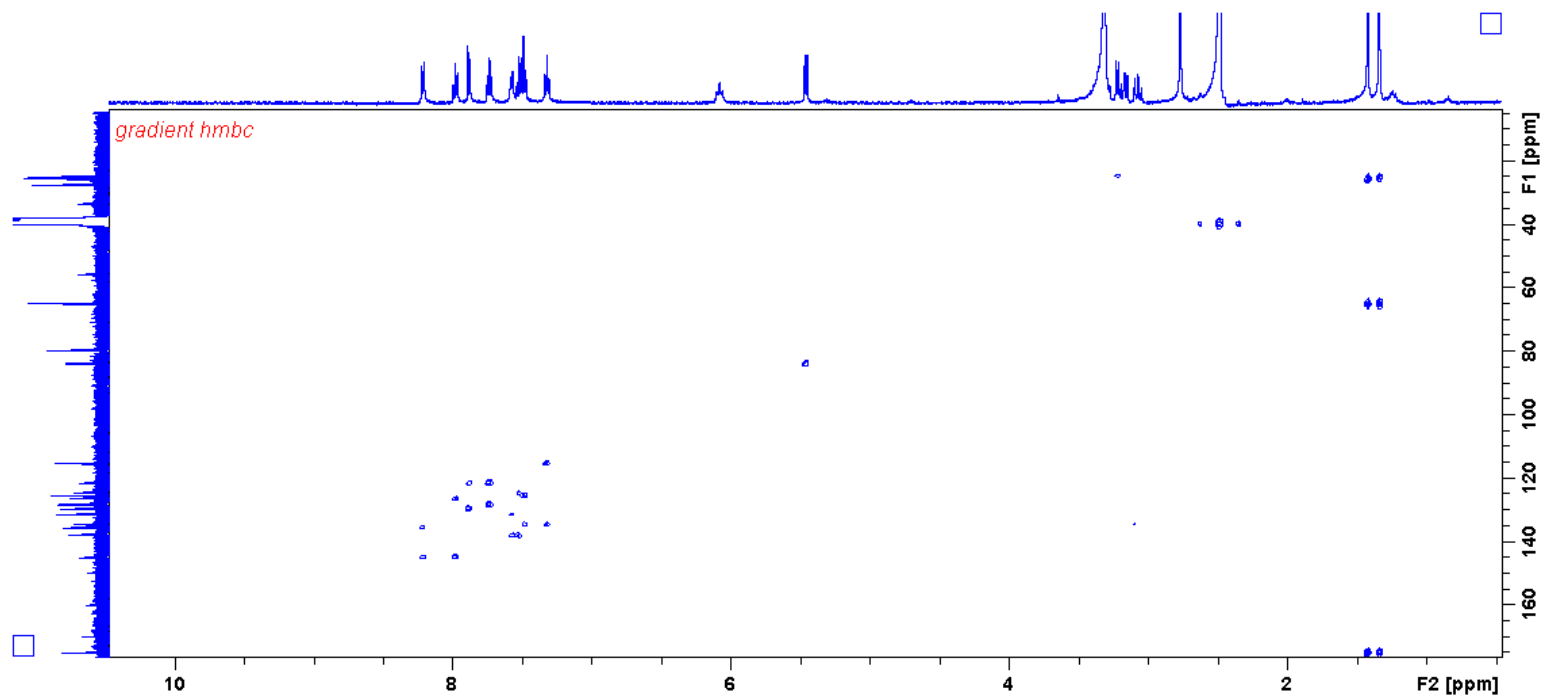**Figure S11.5** ^{1}H-^{13}C HMQC NMR spectrum of **20**.
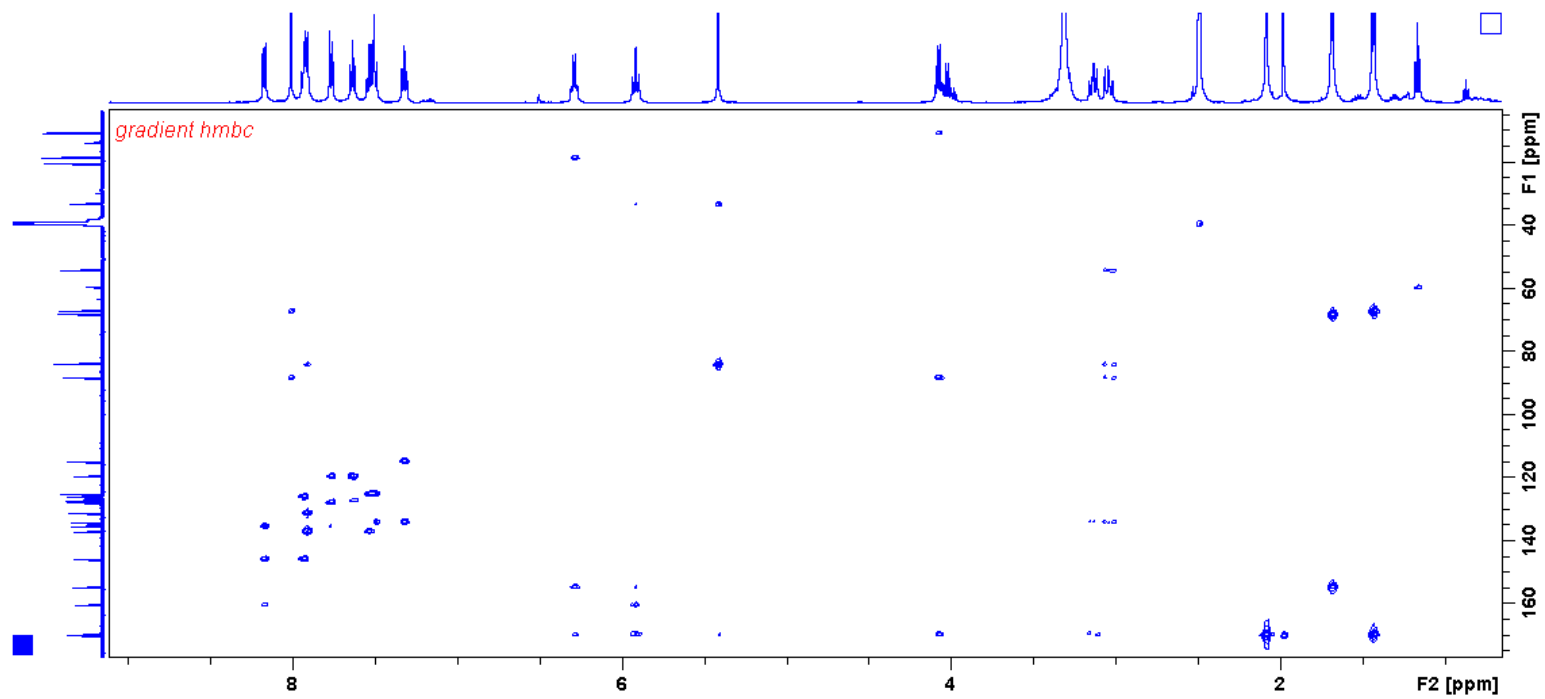
Figure S11.6 ^{1}H-^{13}C HMBC NMR spectrum of **20**.
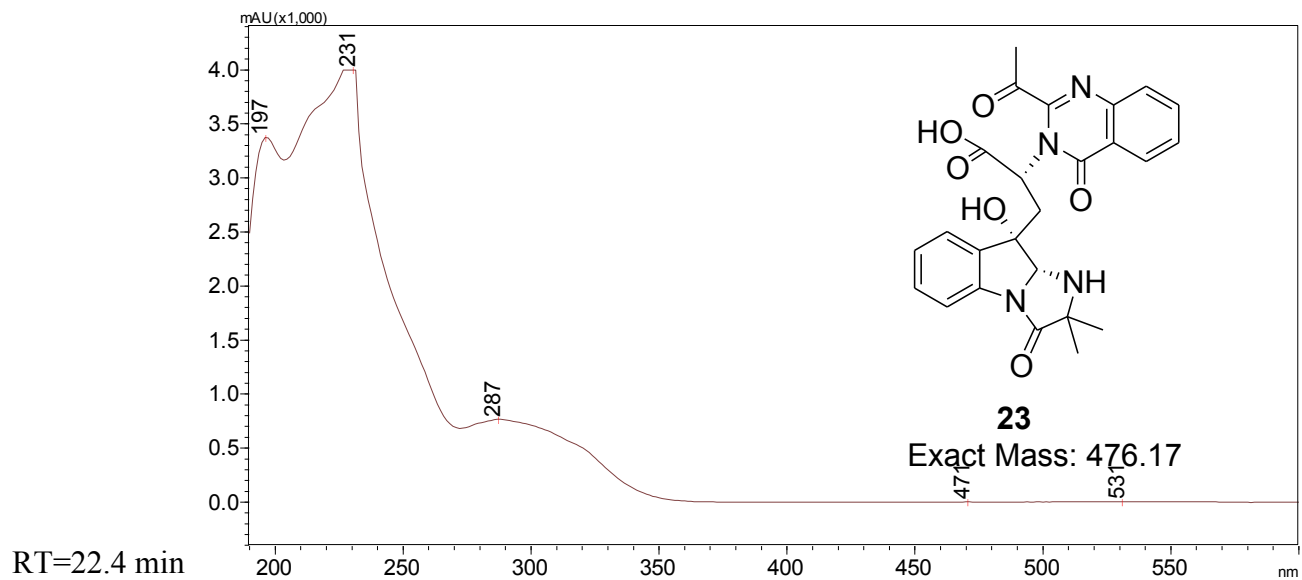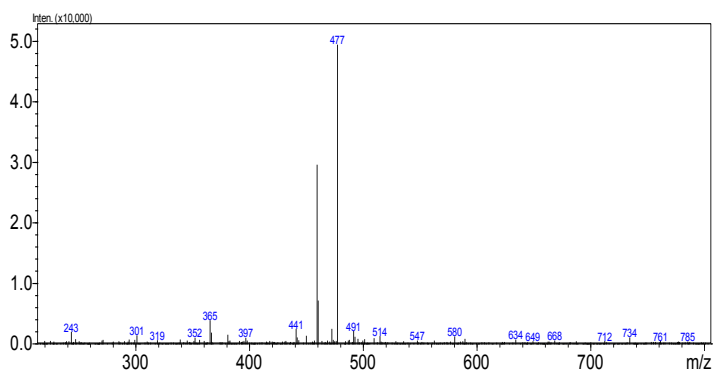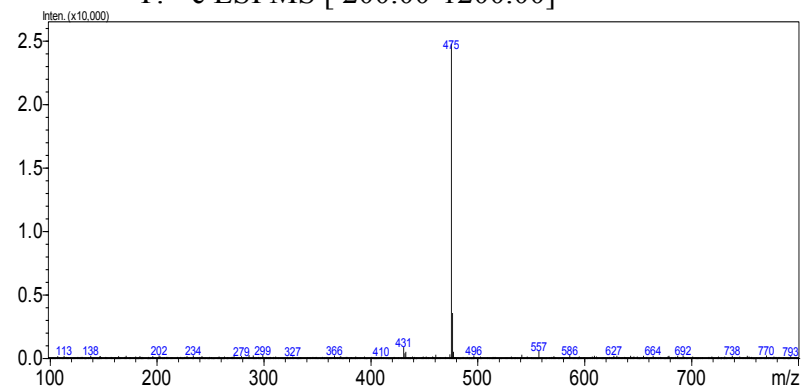
Figure S12 UV spectrum and MS measured during LC-MS for **23**.
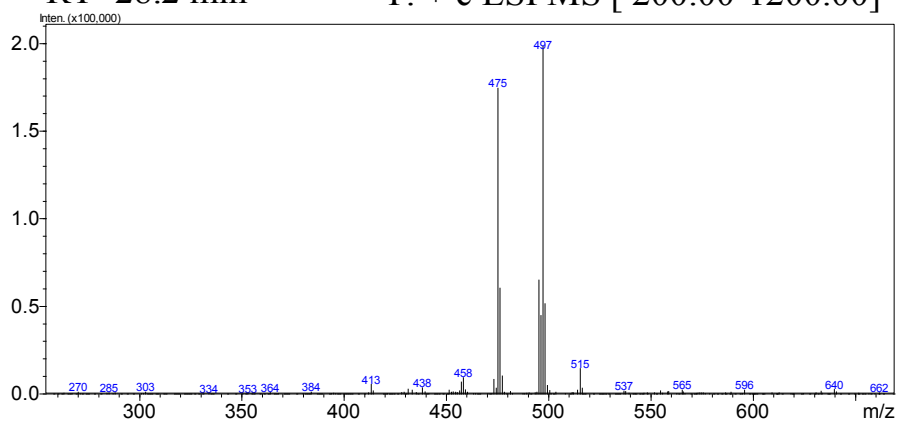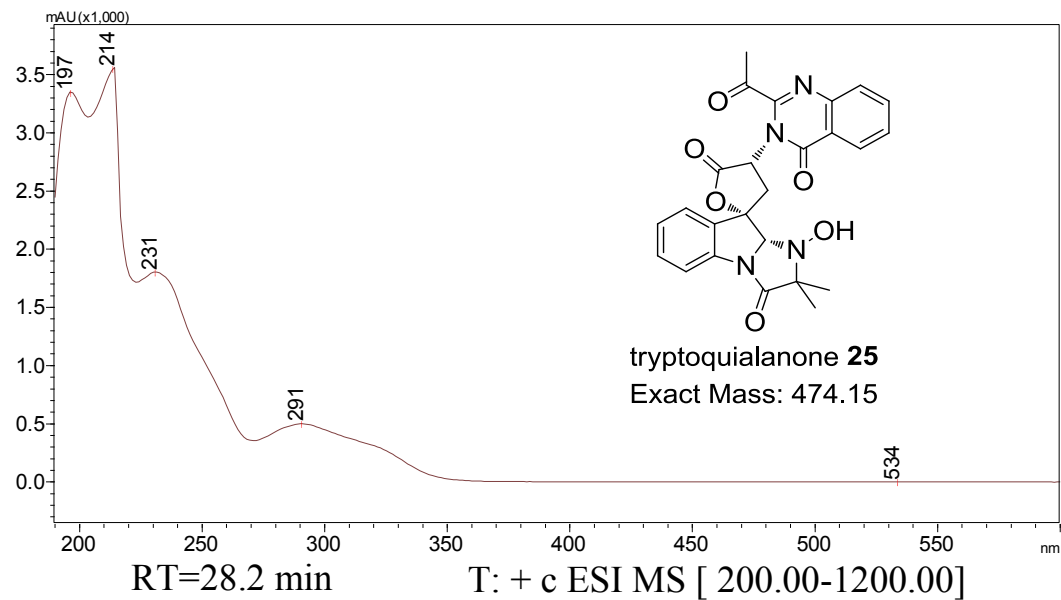
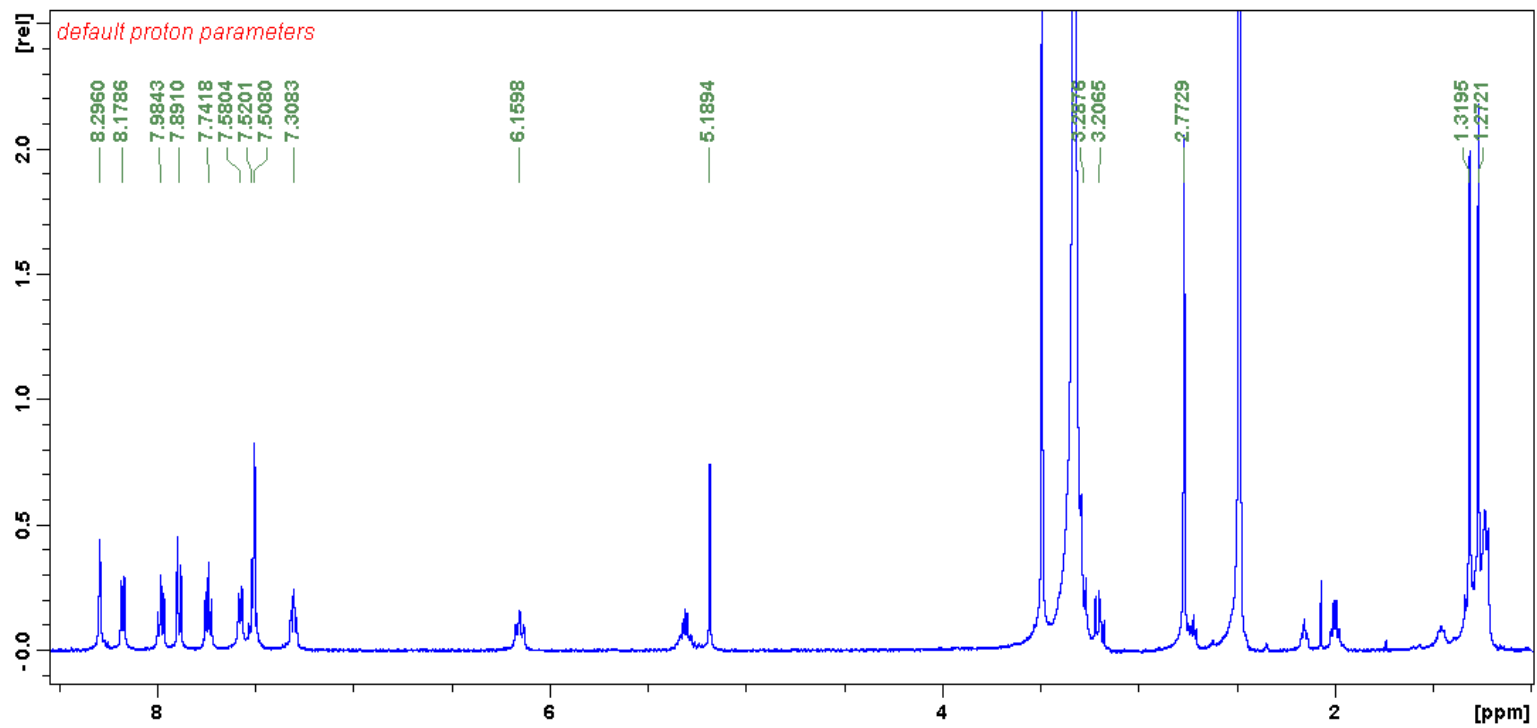**Figure S13.1** UV spectrum and MS measured during LC-MS for **24**.

Figure S13.2 ^{1}H NMR spectrum of **24**. Measured in DMSO-d_6 at 500MHz.

Figure S13.3 ^{13}C NMR spectrum of **24**. Measured in DMSO-d_6 at 125MHz.

Figure S13.4 ^{1}H-^{1}H COSY NMR spectrum of **24**.
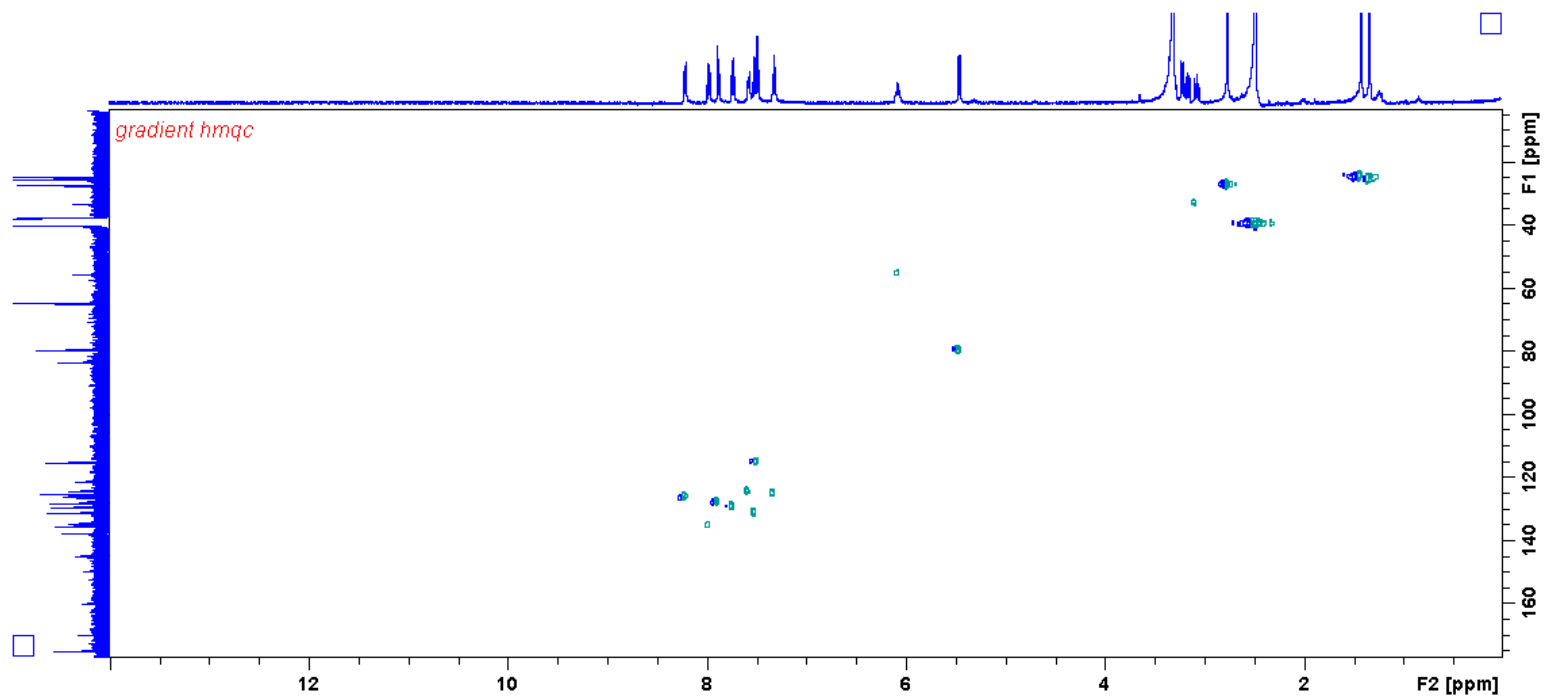
Figure S13.5 ^{1}H-^{13}C HMQC NMR spectrum of **24**.

Figure S13.6 ^{1}H-^{13}C HMBC NMR spectrum of **24**.

Figure S14.1 UV spectrum and MS measured during LC-MS for **25**.

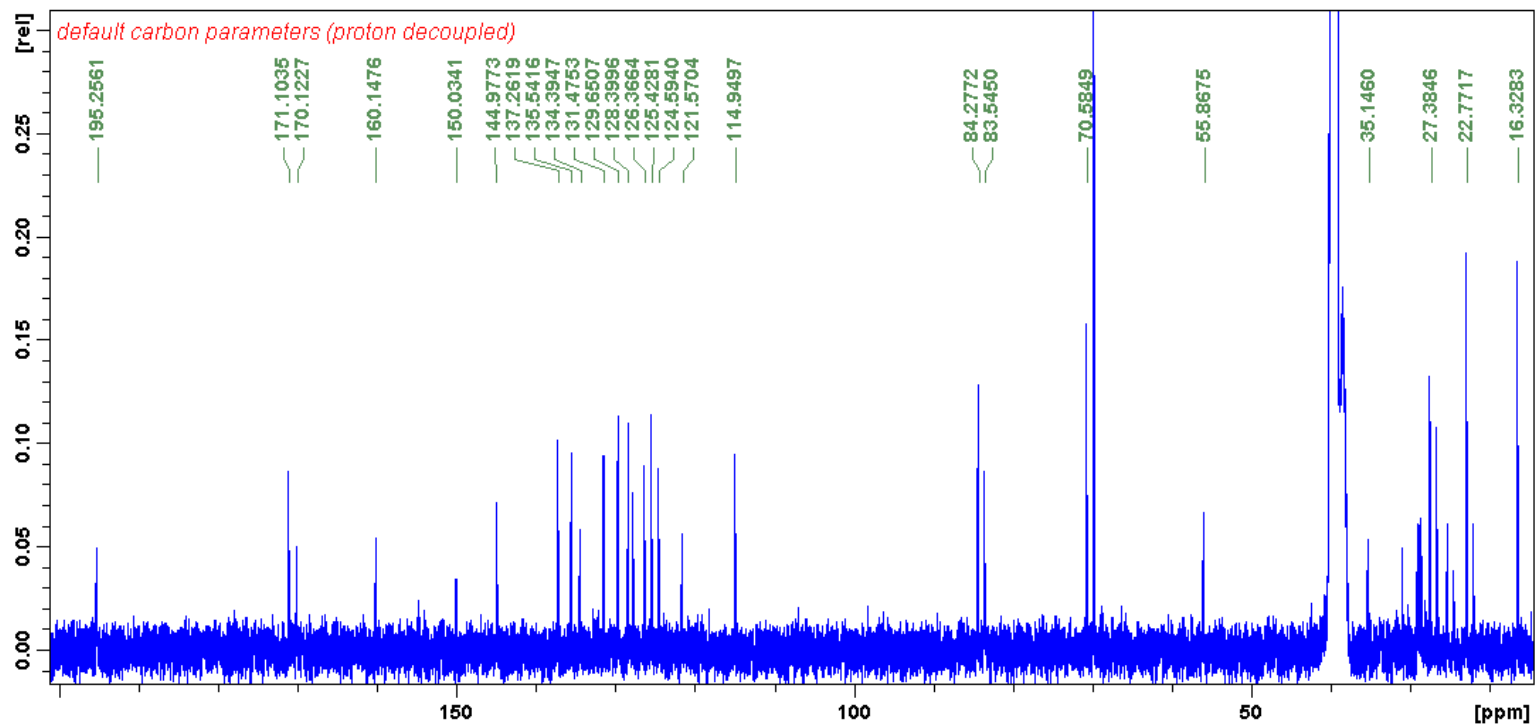**Figure S14.2** ^{1}H NMR spectrum of **25**. Measured in DMSO-d_6 at 500MHz.

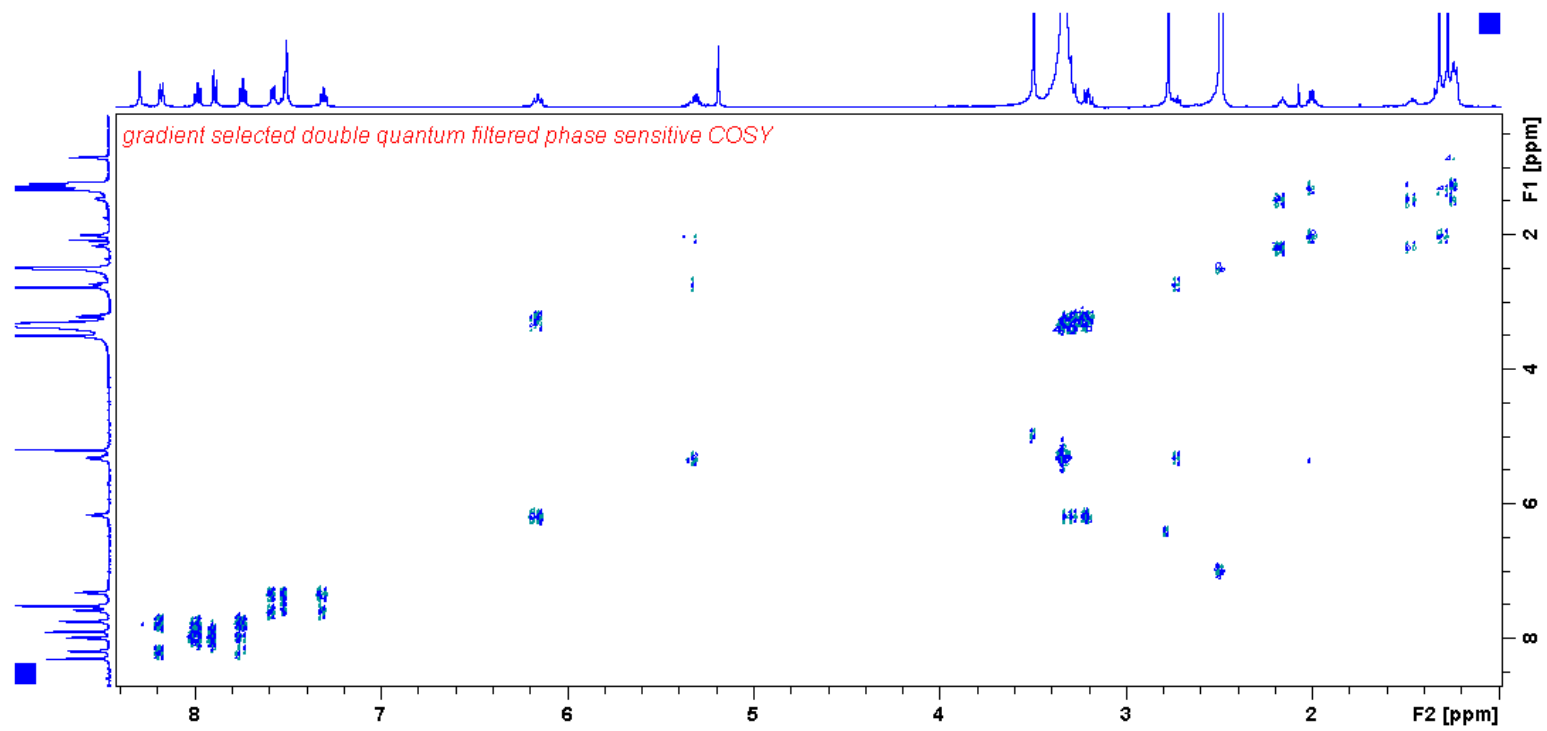**Figure S14.3** ^{13}C NMR spectrum of **25**. Measured in DMSO-d_6 at 125MHz.

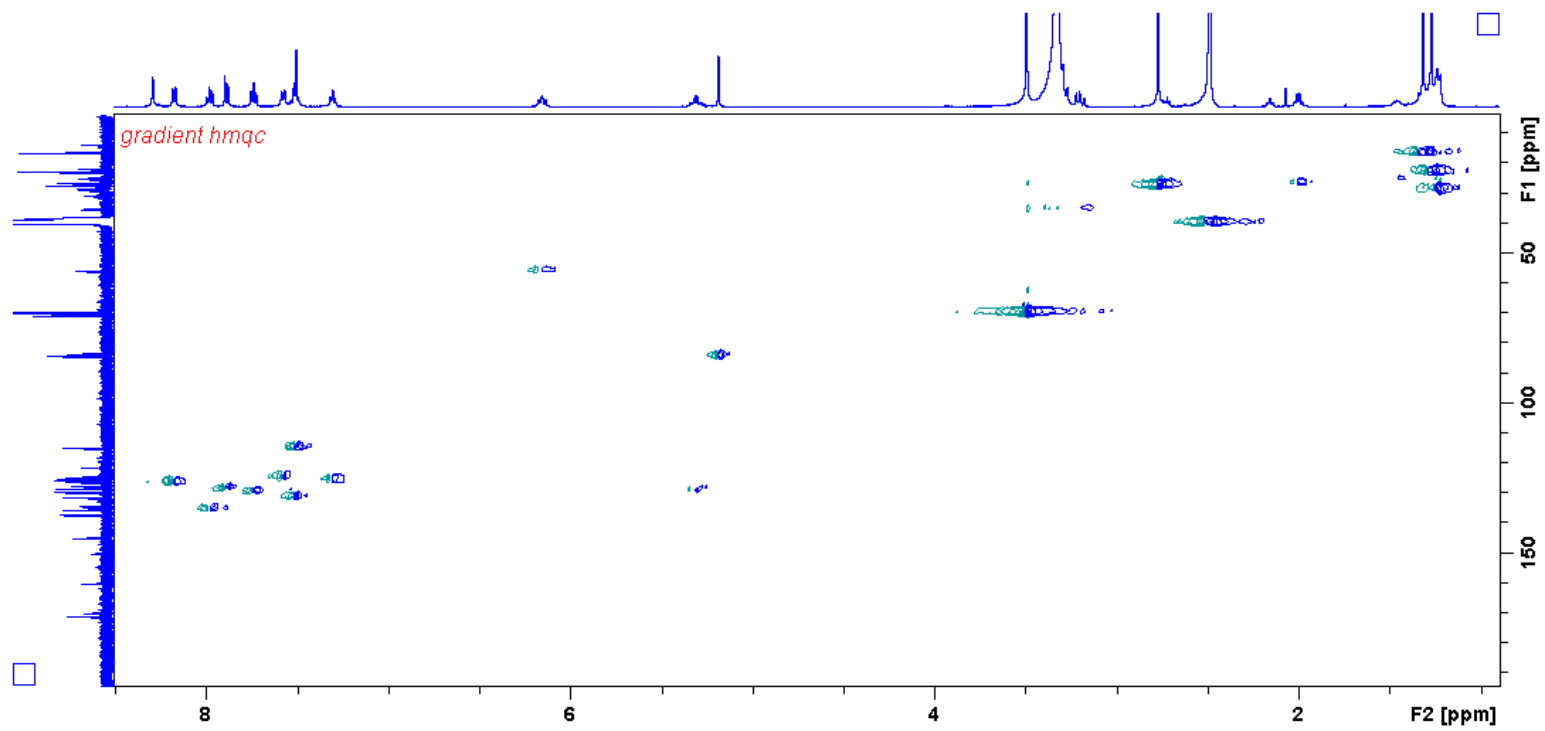**Figure S14.4** ^{1}H-^{1}H COSY NMR spectrum of **25**.

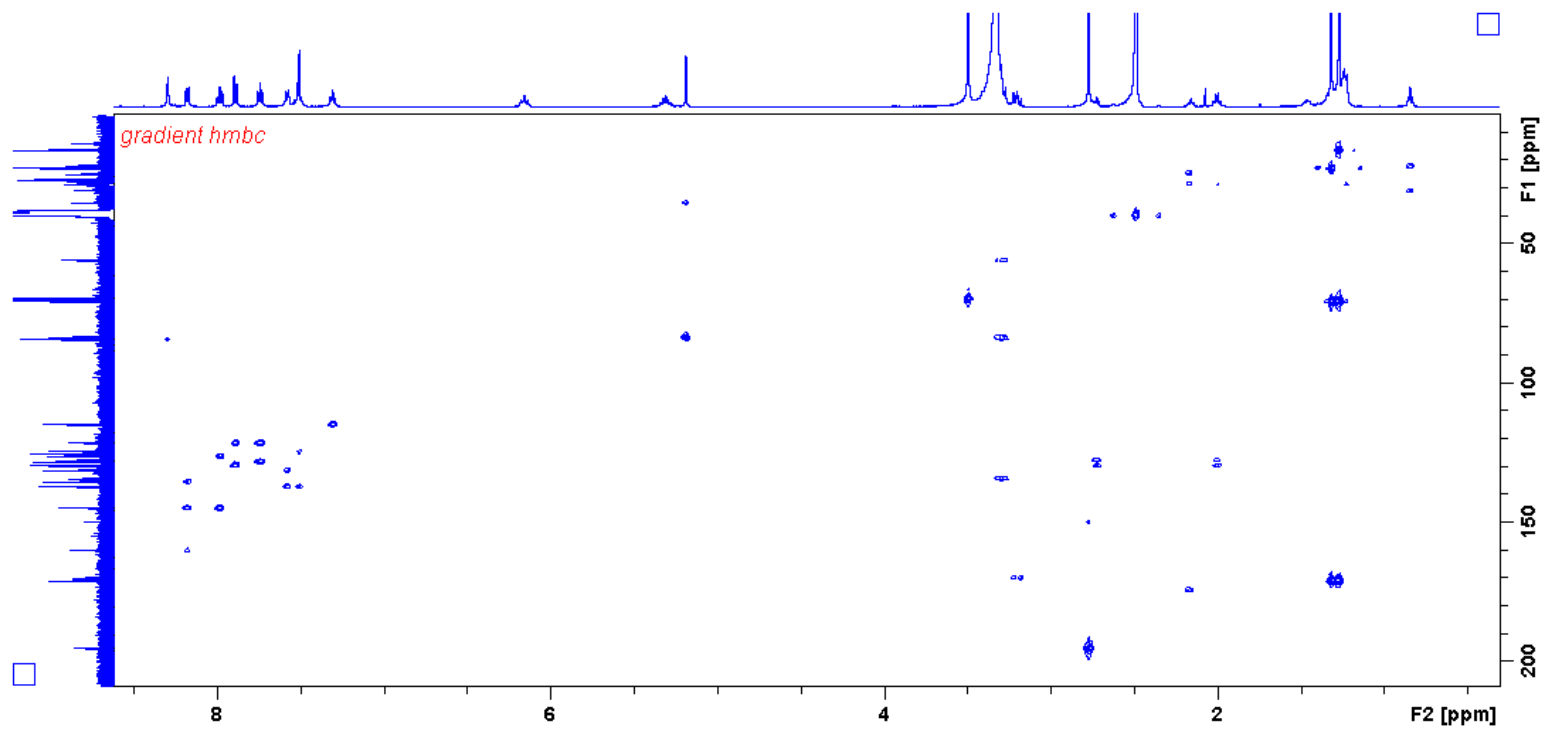**Figure S14.5** ^{1}H-^{13}C HMQC NMR spectrum of **25**.

Figure S14.6 ^{1}H-^{13}C HMBC NMR spectrum of **25**.

26
Exact Mass: 476.17

RT=25.2 min

T: + c ESI MS [200.00-1200.00]
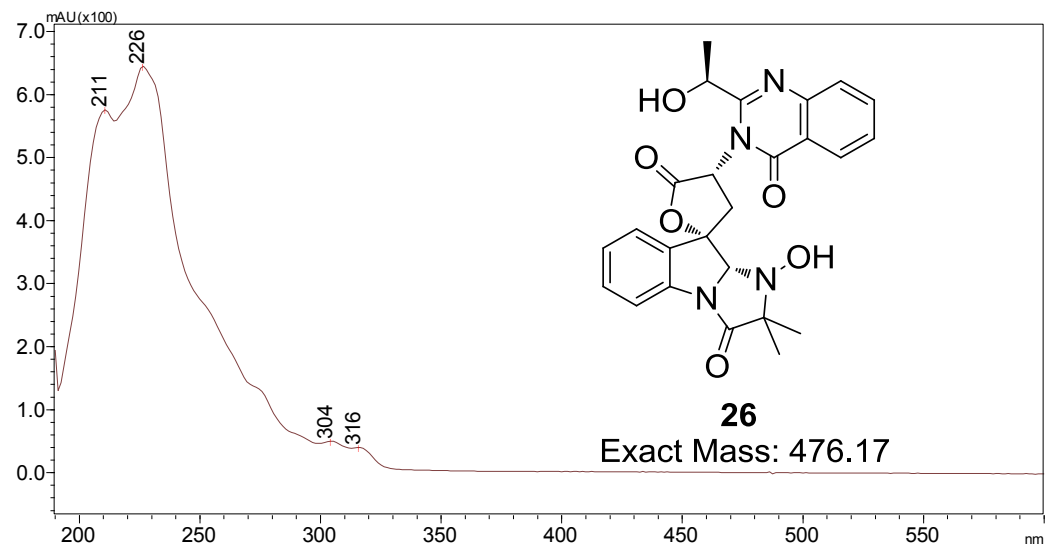
Figure S15 UV spectrum and MS measured during LC-MS for **26**.

RT=26.3 min   T: + c ESI MS [200.00-1200.00]
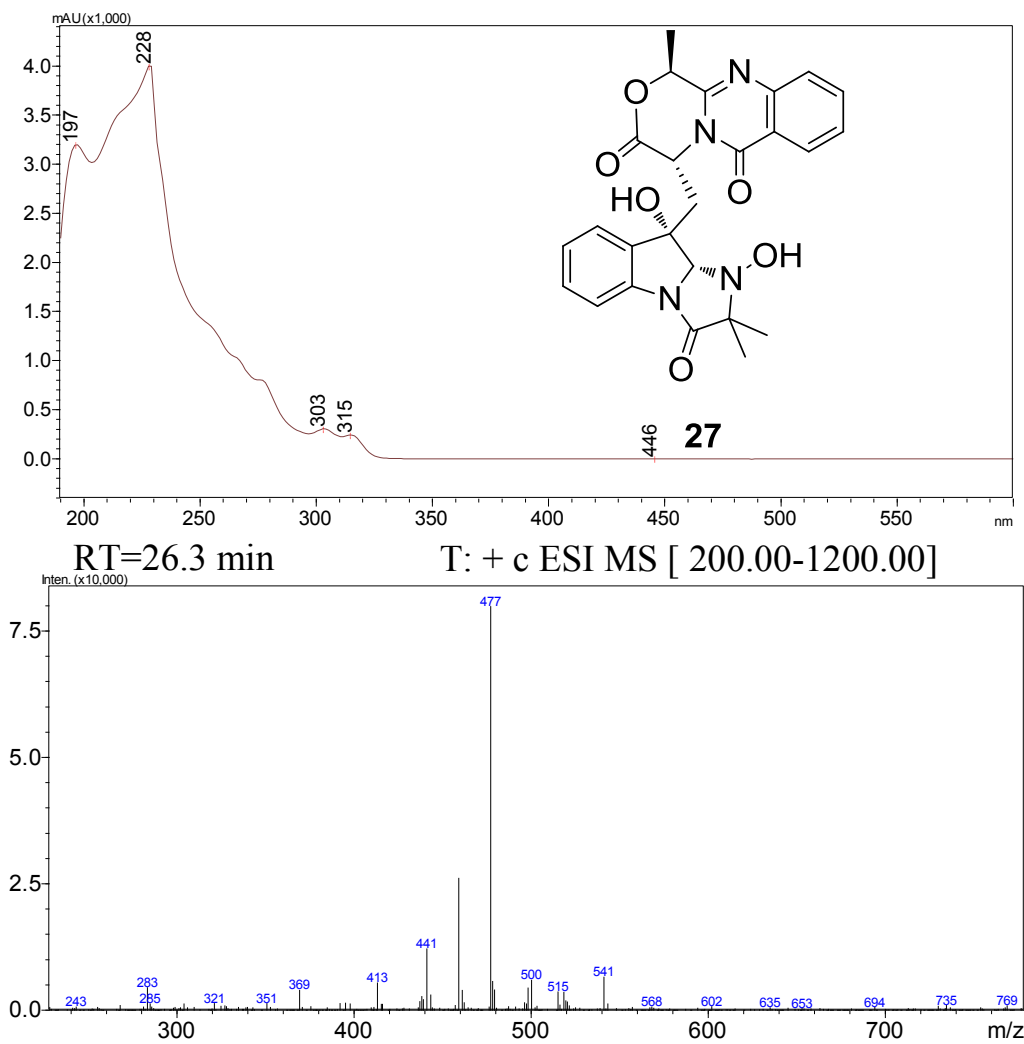
Figure S16 UV spectrum and MS measured during LC-MS for **27**.

208

deoxytryptoquialanine **28**
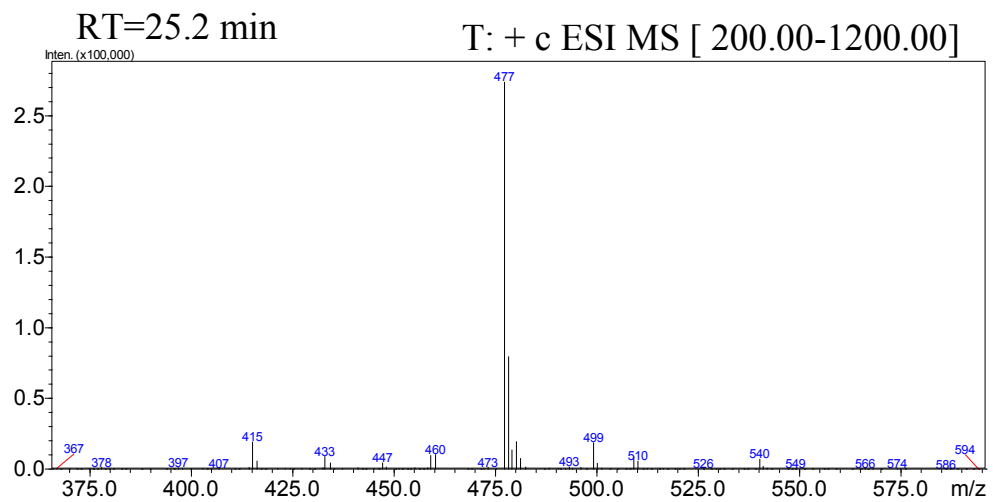Exact Mass: 502.19

RT=24.7 min

T: + c ESI MS [200.00-1200.00]
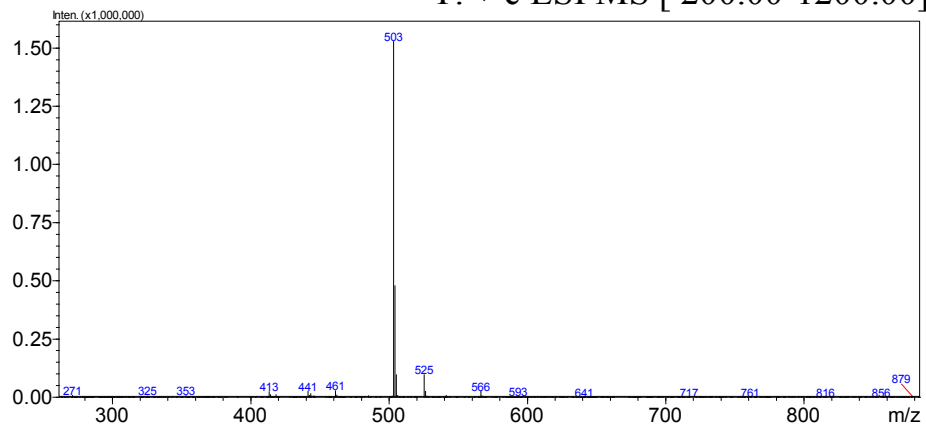
Figure S17.1 UV spectrum and MS measured during LC-MS for **28**.
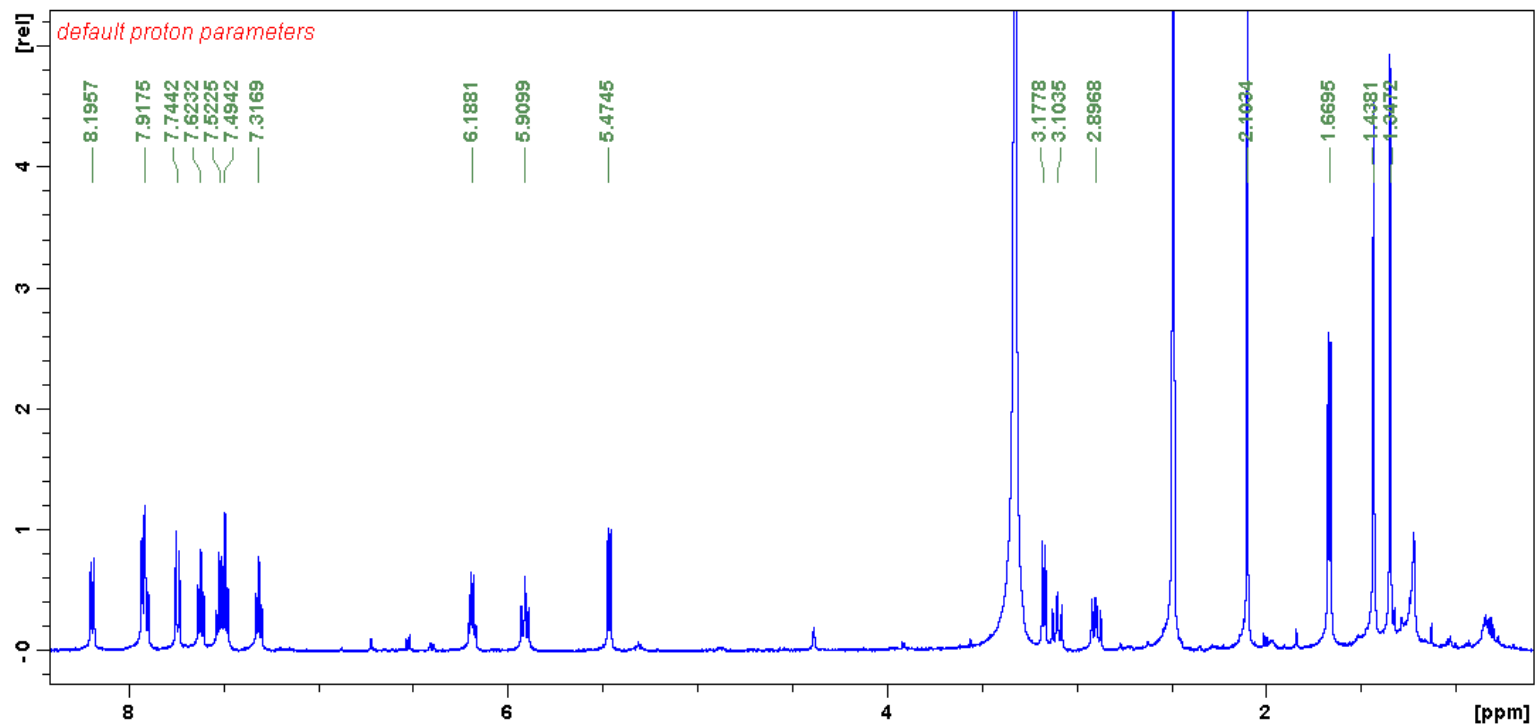
Figure S17.2 ^{1}H NMR spectrum of **28**. Measured in DMSO-d_6 at 500MHz.
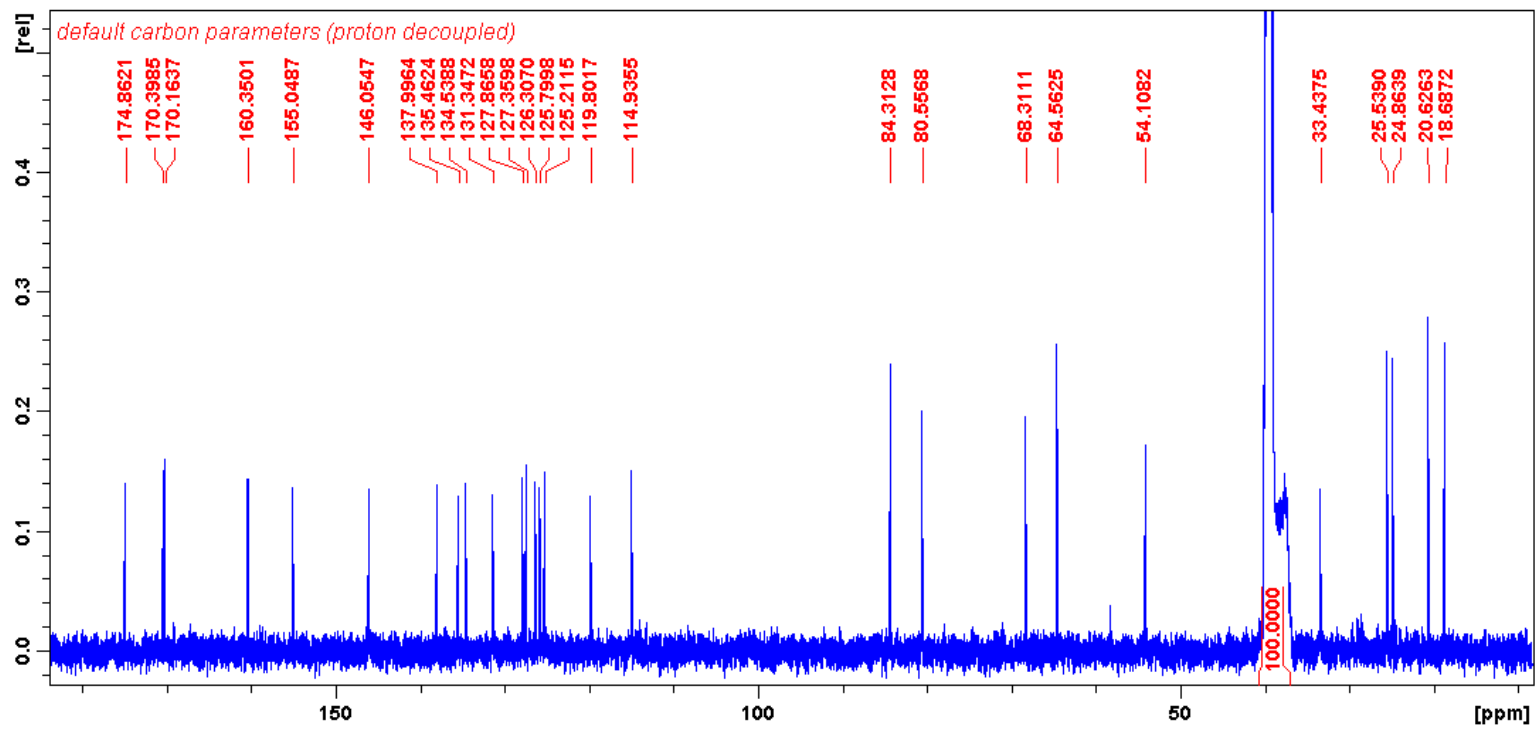
Figure S17.3 ^{13}C NMR spectrum of **28**. Measured in DMSO-d_6 at 125MHz.
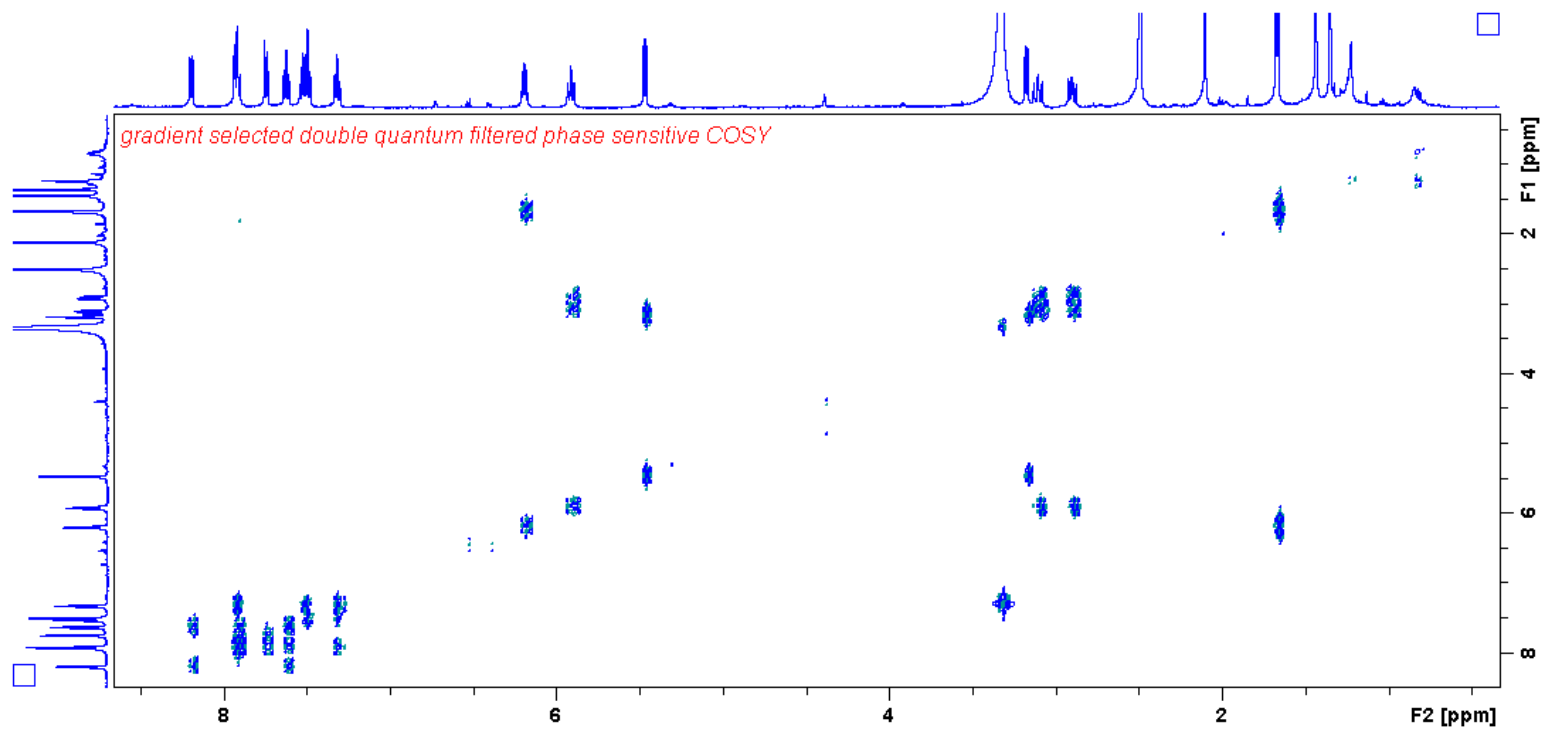
Figure S17.4 ^{1}H-^{1}H COSY NMR spectrum of **28**.
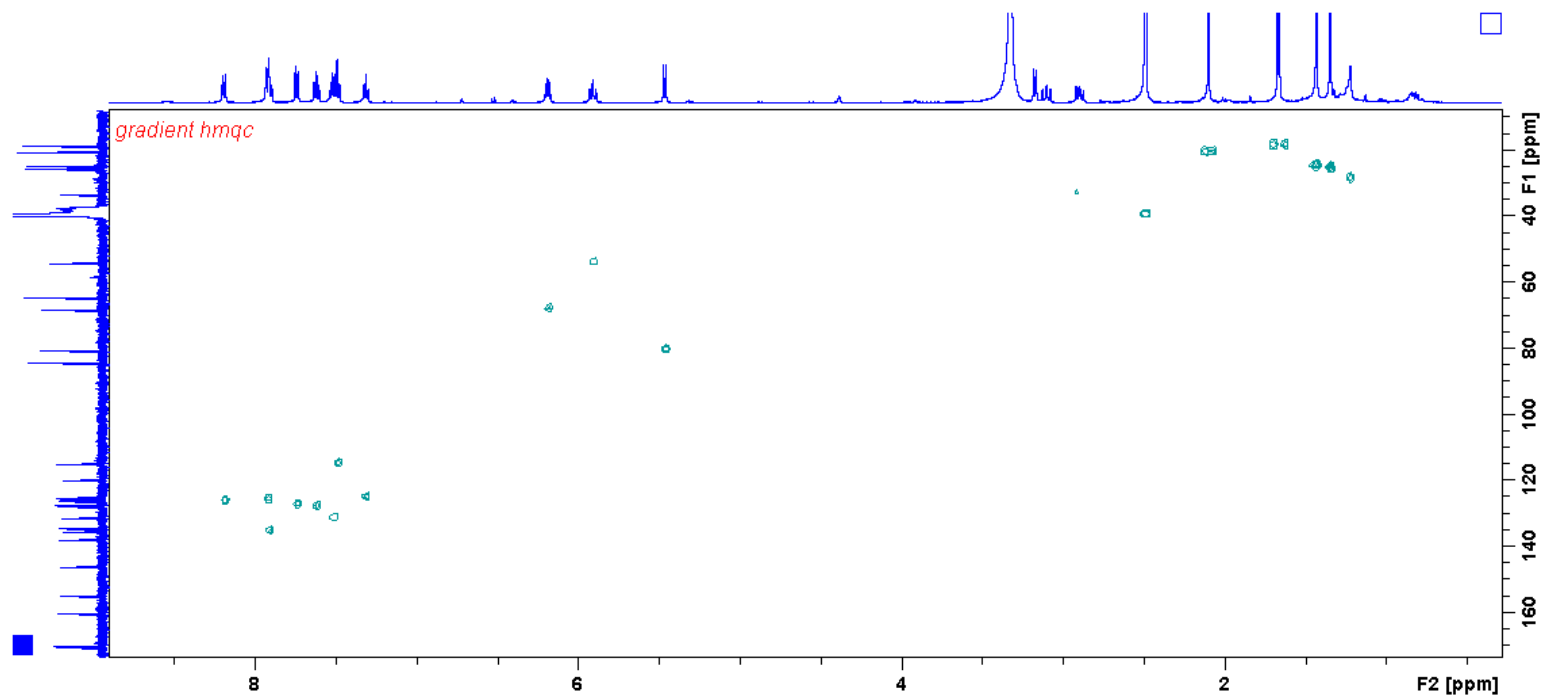
Figure S17.5 ^{1}H-^{13}C MQC NMR spectrum of **28**.
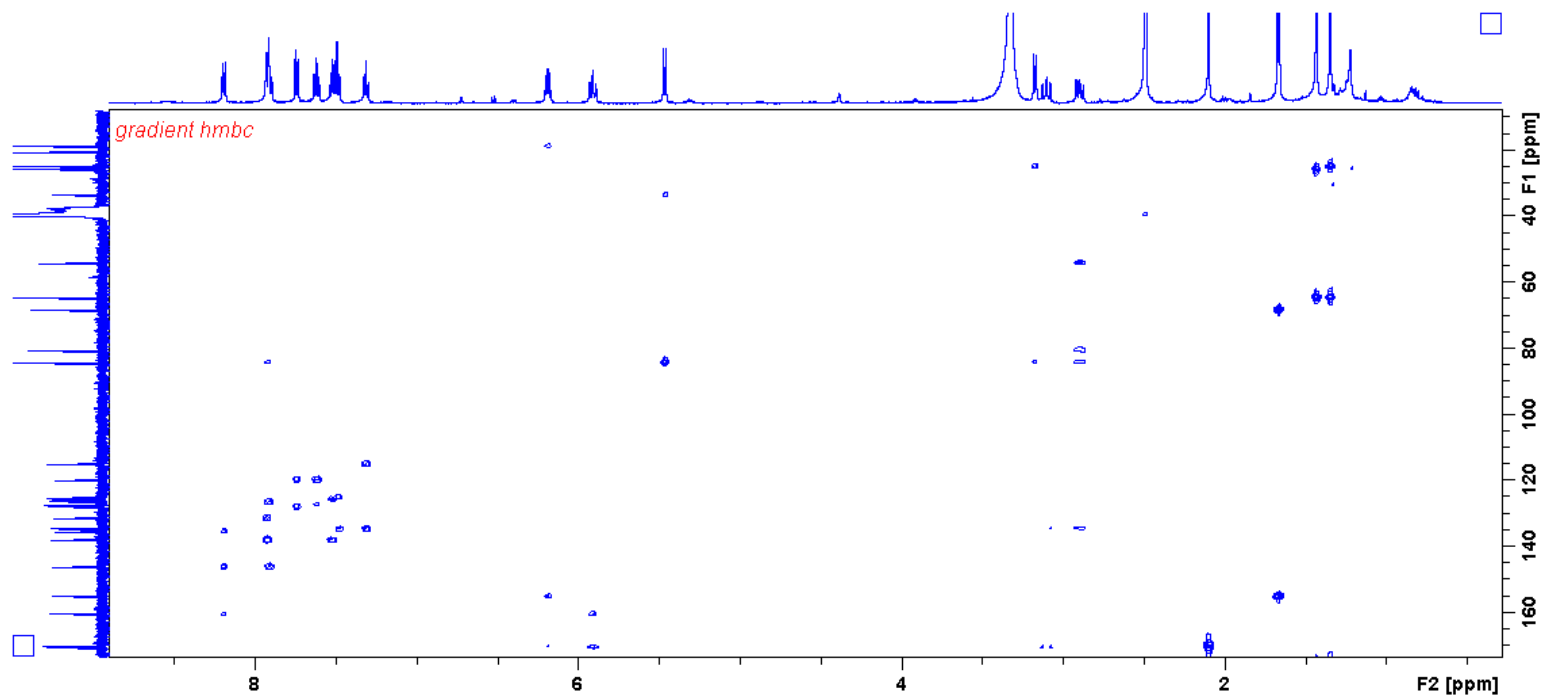
Figure S17.6 ^{1}H-^{13}C HMBC NMR spectrum of **28**.
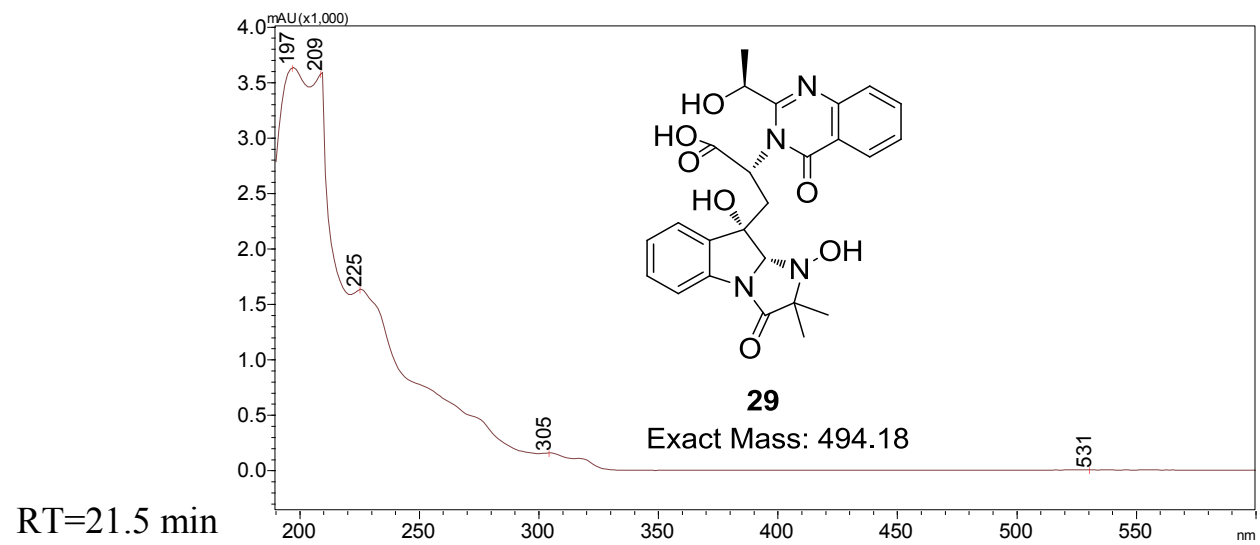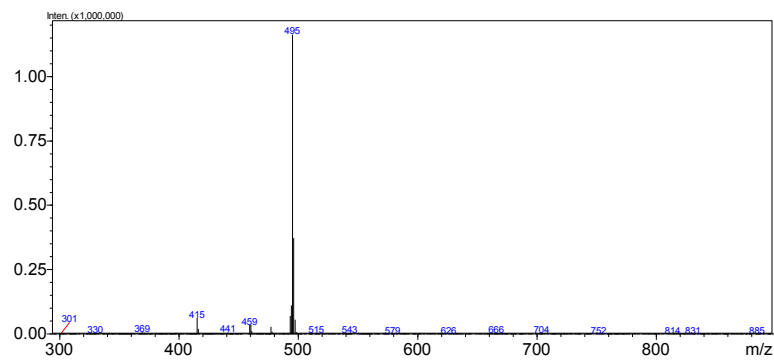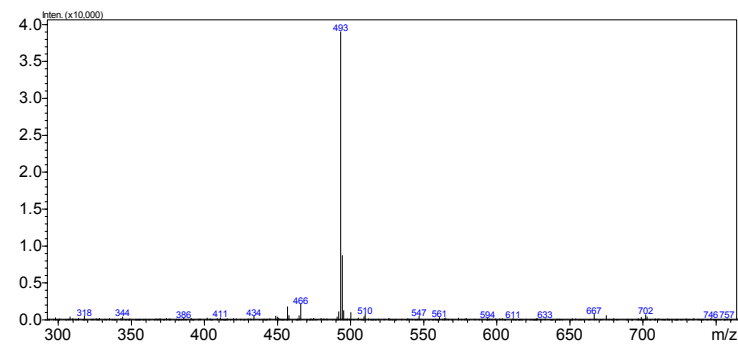
Figure S18 UV spectrum and MS measured during LC-MS for **29**.

30
Exact Mass: 748.28
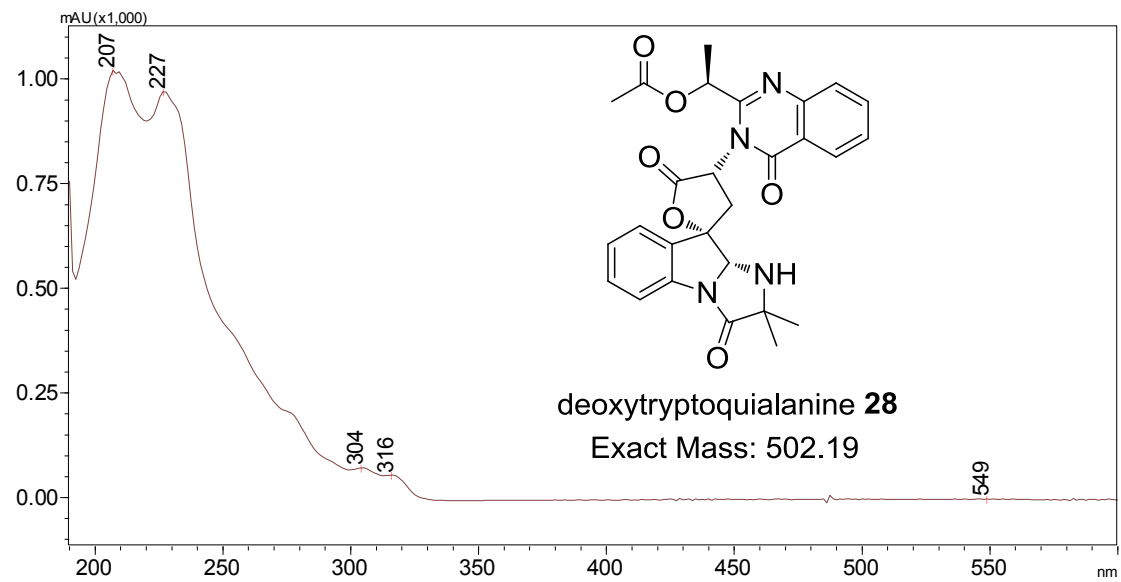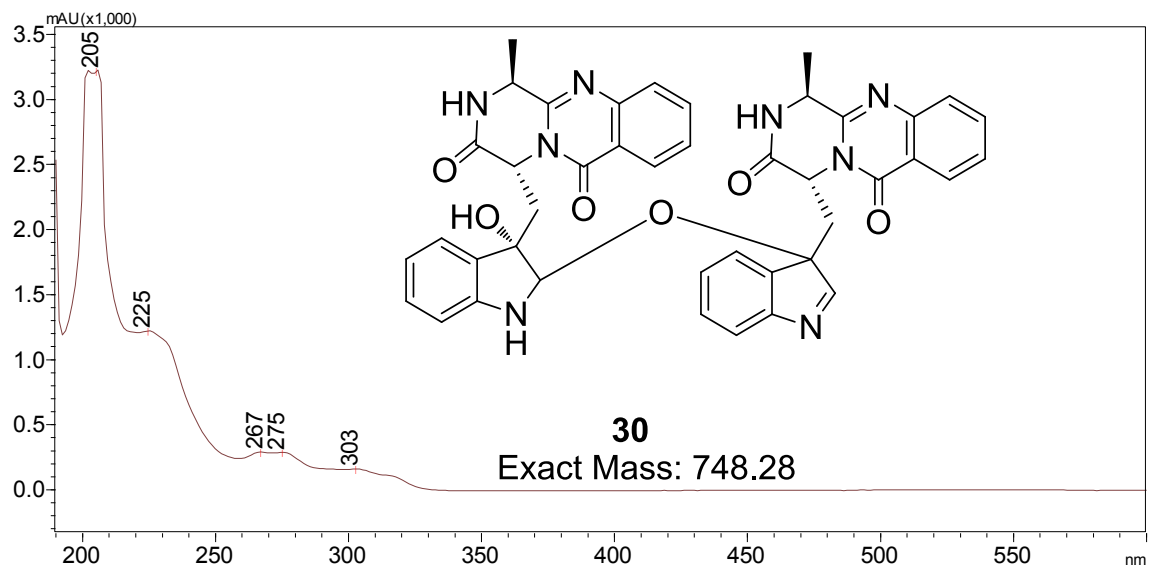
RT=25.8 min

T: + c ESI MS [200.00-1200.00]
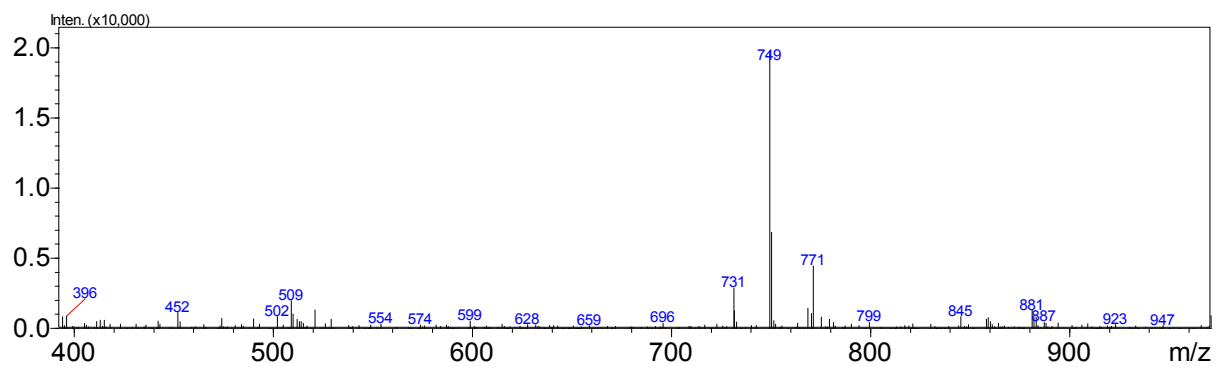
Figure S19 UV spectrum and MS measured during LC-MS for **30**.

4.2 Supporting Information for Section 2.2

Molecular cloning

E. coli XL1-Blue (Stratagene) and *E. coli* TOPO10 (Invitrogen) were used for DNA manipulations following standard techniques. DNA restriction enzymes were purchased from New England Biolabs and used as recommended by the manufacturer. PCR was performed using Phusion® DNA Polymerase (New England Biolabs), Platinum Pfx DNA polymerase (Invitrogen) and GoTaq polymerase (Promega, Madison, WI). The constructs of pCR-Blunt vector (Invitrogen) containing desired PCR products were confirmed by DNA sequencing (Retrogen, CA). *Saccharomyces cerevisiae* strain BJ5464-NpgA (*MATα ura3-52 his3-Δ200 leu2-Δ1 trp1 pep4::HIS3 prb1Δ1.6R can1 GAL*) was used as the yeast expression host[1, 2].

Protein expression and purification

2.1. Expression and purification of proteins from *E. coli*

The expression plasmids were transformed into *E. coli* BL21 (DE3) strain through electroporation for protein expression. The cells were grown at 37 °C in 0.5 L Luria–Bertani (LB) medium with 35 mg l^{-1} kanamycin to an OD600 of 0.4~0.6 and then isopropylthio-β-D-galactoside (IPTG) was added to a final concentration of 0.1 mM to induce protein expression for 16 hours at 16 °C. The cells were harvested by centrifugation (3750 rpm, 15 mins, 4 °C), resuspended in 30 ml lysis buffer (50 mM Tris-HCl, 2 mM EDTA, 2 mM DTT, 500 mM NaCl, 5 mM imidazole, pH 7.9) and lysed through sonication on ice. Cellular lysate was centrifuged at 14,000 rpm for 30 min at 4 °C to remove cellular debris. Ni-NTA agarose resin was added to the soluble fraction and the solution was incubated at 4°C for at least 2 hours. The protein-resin mixture was loaded into a gravity flow column and proteins were eluted with an increasing gradients of imidazole in buffer A (50 mM Tris-HCl, 500 mM NaCl, pH 7.9).

2.2. Expression and purification of proteins from *S. cerevisiae*

For TqaA, AnaPS, TqaA-ΔC$_T$, TqaA H^{3766}A, TqaA H^{2658}A and AnaPS-E^0 enzymes, which all contain *N*-terminal FLAG tag, the yeast strain BJ5464-NpgA was used as an expression host. Two vacuolar proteases PEP4 and PRB1 were inactivated in this host, which is critical to minimize proteolysis of the large recombinant proteins. The 2 μm expression plasmids were transformed into *S. cerevisiae* BJ5464-NpgA by using S. c. EasyComp^{TM} Transformation Kit (Invitrogen). For 1 L of yeast culture, the cells were grown in YPD media with 1% dextrose for 72 hours at 25 °C. The cells were harvested by centrifugation (2500 g, 20 min, 4 °C), resuspended in 20 ml lysis buffer (50 mM NaH$_2$PO$_4$ pH 8.0, 0.15 M NaCl, 10 mM imidazole) and lysed using sonication on ice. Lysate was centrifuged at 35,000 g for 60 min at 4 °C to remove cellular debris. FLAG-tagged proteins were purified by using ANTI-FLAG® M1 Agarose Affinity Gel (Sigma-Aldrich), following the supplied protocols. The cleared cell lysate was applied onto a gravity flow column with packed ANTI-FLAG Agarose Affinity Gel. After washing steps as standard protocols, the protein was eluted with the FLAG peptide elution buffer (0.5 mg ml^{-1} FLAG peptide, 50 mM Tris-HCl, pH 7.4, 100 mM NaCl). After affinity column, TqaA/AnaPS protein was loaded onto a Superdex 200 (GE Healthcare) column and eluted with a flow rate of 0.5 ml min^{-1} in 50 mM Tris pH 8.0 and 100 mM NaCl buffer. A major peak eluting at approximately over 600 kDa was collected for further analysis.

3. ATP-[^{32}P]PP$_i$ exchange assay for TqaA and AnaPS[3]

This assay was used to monitor the substrate-dependent exchange of the ^{32}P label of [^{32}P]PP$_i$ into ATP from the adenylation reactions catalyzed by the A-domains of TqaA or AnaPS . Reactions (100 μl) contained 2 mM ATP, 2 mM MgCl$_2$, 3 mM Na$_4$[^{32}P]PP$_i$ (0.12 μCi), 0.2 μM enzyme, and 2 mM amino acid substrate in Tris reaction buffer (50 mM Tris-HCl [pH 7.5], 100

mM NaCl, 5 mM TCEP, and 5% glycerol). Reactions were initiated by addition of enzyme, incubated at 25 °C for 1.5 hours, and then quenched by addition of a charcoal solution (1.6% w/v activated charcoal, 100 mM sodium pyrophosphate, 3.5% perchloric acid in water). The charcoal was pelleted by centrifugation and washed with a solution containing 100 mM sodium pyrophosphate and 3.5% perchloric acid, and the charcoal-bound radioactivity detected by liquid scintillation counting.

4. Chemical synthesis of peptidyl-SNAC, peptidyl-*S*-CoA substrates

Ant-D-Trp, Ant-L-Trp, Ant-D-Trp-L-Ala, Ant-L-Trp-L-Ala, Ant-L-*N*-Me-Trp-L-Ala, benzoate (benz)-D-Trp-L-Ala and salicylate (sal)-D-Trp-L-Ala are purchased from RS Synthesis®. The purity of all the peptides are >95% determined by HPLC. The traces provided by the vendors are shown in **Supplemental Figure 37.**

4.1. BOC protection of dipeptides

Ant-Trp (D or L) **4** (50 mg, 0.155 mmol, 1.0 equiv) and BOC anhydride (169 mg, 0.773 mmol, 1.0 equiv) were dissolved in ethanol (10 ml) and heated at 60 °C for 5 hours. The solvent was then removed *in vacuo* and the residue was dissolved in 50% acetonitrile/water and purified by preparative HPLC on a Beckmann Coulter Gold system equipped with a reverse phase C18 column (Phenomenex Luna, 250 x 21.2 mm, 10 micron) with detection at 280 nm. Solvent system A (water plus 0.1% TFA) and B (acetonitrile plus 0.1% TFA) held at 35% B for 1 min and then run over a linear gradient of 35-75% over 30 min, before increasing from 75-95% B over 1 min and a holding at 95% B for 5 min. The column was then equilibrated back to initial conditions by returning to 35% B and holding for 5 min. The peak with retention time of 29.5 min was collected.

^{1}H-NMR (DMSO-d$_6$, 400 MHz) δ ppm: 1.45 (s, 9 H), 3.13 - 3.35 (m, 3 H), 4.55 - 4.69 (m, 1 H), 6.93 - 7.01 (m, 1 H), 7.05 (q, *J*=6.91 Hz, 2 H), 7.21 (s, 1 H), 7.33 (d, *J*=7.83 Hz, 1 H), 7.46 (t, *J*=7.83 Hz, 1 H), 7.59 (d, *J*=7.43 Hz, 1 H), 7.72 (d, *J*=7.83 Hz, 1 H), 8.20 (d, *J*=8.61 Hz, 1 H), 8.91 (d, *J*=7.83 Hz, 1 H), 10.47 (s, 1 H), 10.83 (br. s., 1 H).

^{13}C-NMR (DMSO-d$_6$, 100 MHz) δ: 26.39, 27.93, 53.71, 79.84, 110.25, 111.46, 118.07, 118.32, 118.40, 118.60, 120.98, 121.22, 123.61, 127.08, 128.43, 132.37, 136.12, 139.69, 152.02, 168.46, 173.18

HRMS: *m/z* calculated for C$_{23}$H$_{25}$N$_3$O$_5$: 424.1867 [M+H]$^+$. Found: 424.1869

4.2. Synthesis of dipeptide SNACs (33-SNAC)

Diisopropylethylamine (3.3 µl, 18.9 µmol, 4.0 equiv) was added to a stirred solution of *N*-BOC-Ant-Trp (D or L) (2 mg, 4.73 µmol, 1.0 equiv) and PyBOP (7.4 mg, 14.18 µmol, 3.0 equiv) in DCM (2 ml). To the resulting clear and colorless solution was added *N*-acetylcysteamine (0.75 µl, 7.09 µmol, 1.5 equiv) and the reaction was monitored by LC-MS and stirred until completion (circa 1 hour). The reaction was cooled to 5 °C and TFA (2 ml) was added in order to remove the BOC protecting group. The reaction was again monitored by LC-MS until deprotection was complete (circa 1 hour). The solvent was then removed *in vacuo* and the residue was dissolved in 50% acetonitrile/water and purified by preparative HPLC on a Beckmann Coulter Gold system equipped with a reverse phase C18 column (Phenomenex Luna, 250 x 21.2 mm, 10 micron) with detection at 275 nm. Solvent system A (water plus 0.1% TFA) and B (acetonitrile plus 0.1% TFA) held at 20% B for 1 min and then run over a linear gradient of 20-30% over 10 min, followed by a gradient of 30-40% B over 20 min, before increasing from 40-95% B over 1 min and a holding at 95% B for 5 min. The column was then equilibrated back

to initial conditions by returning to 20% B and holding for 5 min. The peak with retention time of 31 min was collected. HPLC assessments of the purity of D or L **33-SNAC** can be seen in **Figure S24**.

4-SNAC: ^{1}H NMR (400 MHz, DMSO-d_6) ppm: 2.84 - 2.95 (m, 2 H), 3.10 - 3.18 (m, 2 H), 3.19 - 3.24 (m, 1 H), 3.30 (dd, J=14.87, 4.30 Hz, 1 H), 4.72 - 4.82 (m, 1 H), 6.52 (t, J=7.43 Hz, 1 H), 6.67 (d, J=8.22 Hz, 1 H), 6.94 - 7.02 (m, 1 H), 7.03 - 7.10 (m, 1 H), 7.15 (t, J=7.63 Hz, 1 H), 7.19 (s, 1 H), 7.32 (d, J=7.83 Hz, 1 H), 7.54-7.58 (m, 1 H), 7.56-7.60 (m, 1 H), 8.03 (br. s., 1 H), 8.70 (d, J=7.83 Hz, 1 H), 10.82 (br. s., 1 H).

HRMS: *m/z* calculated for $C_{22}H_{24}N_4O_3S$: 425.1642 [M+H]$^+$. Found: 425.1645

4.3. Synthesis of tripeptide (Ant-D-Trp-L-Ala) SNAC (31-SNAC)

Potassium carbonate (28 mg, 0.203 mmol, 4.0 equiv) was added to a stirred solution of Ant-D-Trp-L-Ala **3**(20 mg, 0.051 mmol, 1.0 equiv) and PyBOP (53 mg, 0.152 mmol, 3.0 equiv) in 50% THF/water (4 ml). To the resulting clear and colorless solution was added *N*-acetylcysteamine (11 μl, 0.101 mmol, 2.0 equiv) and the reaction was monitored by LC-MS and stirred for 2 hours. The solvent was then removed *in vacuo* and the residue was dissolved in 50% acetonitrile/water and purified by preparative HPLC on a Beckmann Coulter Gold system equipped with a reverse phase C18 column (Phenomenex Luna, 250 x 21.2 mm, 10 micron) with detection at 275 nm. Solvent system A (water plus 0.1% TFA) and B (acetonitrile plus 0.1% TFA) held at 5% B for 1 min and then run over a linear gradient of 5-40% over 20 min, followed by a gradient of 40-60% B over 10 min, before increasing from 60-95% B over 2 min and a holding at 95% B for 5 min. The column was then equilibrated back to initial conditions by

returning to 5% B and holding for 5 min. The peak with retention time of 27.5 min was collected. HPLC assessment of the purity of **31-SNAC** can be seen in **Figures 15B** and **Figure S30**.

3-SNAC: ^{1}H-NMR (600 MHz, MeCN-d_3) ppm: 1.27 (d, J=7.04 Hz, 3 H), 1.84 (s, 3 H), 2.91 (t, J=6.75 Hz, 2 H), 3.20-3.28 (m, 2H), 3.26-3.32 (m, 1H), 3.41 (dd, J=14.67, 5.28 Hz, 1 H), 4.50 (quin, J=7.34 Hz, 1 H), 4.78 - 4.87 (m, 1 H), 6.60 - 6.72 (m, 1 H), 6.73 - 6.82 (m, 1 H), 7.02 - 7.08 (m, 1 H), 7.13 (t, J=7.63 Hz, 1 H), 7.19 (s, 1 H), 7.21 - 7.25 (m, 1 H), 7.26-7.30 (m, 2 H), 7.31 - 7.36 (m, 1 H), 7.39 (d, J=8.22 Hz, 1 H), 7.66 (d, J=7.63 Hz, 1 H), 9.25 (br. s., 2 H).

HRMS: m/z calculated for $C_{25}H_{29}N_5O_4S$: 496.2013 [M+H]$^+$. Found: 496.2013

4.4. Synthesis of tripeptide-*S*-CoAs: Ant-D-Trp-L-Ala-*S*-CoA (31-*S*-CoA), Ant-L-Trp-L-Ala-*S*-CoA (*epi*-31-*S*-CoA), Ant-L-*N*-Me-Trp-L-Ala-*S*-CoA, Benz-D-Trp-L-Ala-*S*-CoA and Sal-D-Trp-L-Ala-*S*-CoA

Potassium carbonate (7 mg, 50.93 µmol, 4.0 equiv) was added to a stirred solution of Ant-D-Trp-L-Ala **3,** Ant-L-Trp-L-Ala *epi*-**3,** Ant-L-*N*-Me-Trp-L-Ala*,* Benz-D-Trp-L-Ala and Sal-D-Trp-L-Ala (12.73 µmol, 1.0 equiv) and PyBOP (19 mg, 38.20 µmol, 3.0 equiv) in 50% THF/water (2 ml). To the resulting clear and colorless solution was added coenzyme A trilithium salt (10 mg, 12.73 µmol, 1.0 equiv) and the reaction was monitored by LC-MS and stirred for 2 hours. The solvent was then removed *in vacuo* and the residue was dissolved in 50% acetonitrile/water and purified by preparative HPLC on a Beckmann Coulter Gold system equipped with a reverse phase C18 column (Phenomenex Luna, 250 x 21.2 mm, 10 micron) with detection at 275 nm. Solvent system A (water plus 0.1% TFA) and B (acetonitrile plus 0.1% TFA) held at 5% B for 1 min and then run over a linear gradient of 5-30% over 20 min, followed by a gradient of 30-45% B over 10 min, before increasing from 45-95% B over 2 min and a holding at 95% B for 5 min. The

column was then equilibrated back to initial conditions by returning to 5% B and holding for 5 min. The peak with the desired product was collected. HPLC assessment of the purity of the various tripeptidyl-CoA samples are shown in **Figures S34**.

3-*S*-CoA: ^{1}H-NMR (DMSO-d$_6$, 600 MHz) δ ppm: 0.73 (s, 3 H), 0.94 (s, 3 H), 1.23 (d, *J*=7.04 Hz, 3 H), 2.25 (t, *J*=6.75 Hz, 2 H), 2.79-2.90 (m, 2 H), 3.09-3.13 (m, 2H), 3.16-3.22 (m, 1H), 3.22-3.28 (m, 2H), 3.30-3.35 (m, 1H), 3.43-3.56 (m, 1H), 3.74-3.81 (m, 1H), 3.87-3.98 (m, 2H), 4.16 (d, *J*=8.22 Hz, 2H), 4.37 (t, *J*=7.04 Hz, 1H), 4.72 (d, *J*=4.70 Hz, 1H), 4.73-4.78 (m, 1H), 4.78-4.83 (m, 1H), 5.96 (d, *J*=5.28 Hz, 1H), 6.48 (t, *J*=7.34 Hz, 1 H), 6.65 (d, *J*=8.22 Hz, 1 H), 6.95 -7.00 (m, 1 H), 7.05 (t, *J*=7.63 Hz, 1 H), 7.11 (t, *J*=7.63 Hz, 1 H), 7.21 (s, 1 H), 7.30 (d, *J*=8.22 Hz, 1 H), 7.48 (d, *J*=7.63 Hz, 1 H), 7.67 (d, *J*=7.63 Hz, 1 H), 7.75 (t, *J*=5.87 Hz, 1 H), 8.10-8.11 (m, 1 H), 8.11-8.13 (m, 1 H), 8.24 (s, 1 H), 8.48 (s, 1 H), 8.71 (d, *J*=7.63 Hz, 1 H), 10.79 (br. s., 1 H).

HRMS: *m/z* calculated for C$_{42}$H$_{56}$N$_{11}$O$_{19}$P$_3$S: 1144.2760 [M+H]$^+$. Found: 1144.2761

***epi*-31**-*S*-CoA: ^{1}H-NMR (DMSO-d$_6$, 600 MHz) δ ppm: 0.72 (s, 3 H), 0.94 (s, 3 H), 1.33 (d, *J*=7.04 Hz, 3 H), 2.25 (t, *J*=6.75 Hz, 2 H), 2.81-2.91 (m, 2 H), 3.09-3.13 (m, 2H), 3.13-3.21 (m, 1H), 3.22-3.28 (m, 2H), 3.33-3.41 (m, 1H), 3.43-3.56 (m, 1H), 3.74-3.81 (m, 1H), 3.87-3.98 (m, 2H), 4.10-4.19 (m, 2H), 4.46 (quin, *J*=7.19 Hz, 1H), 4.69-4.73 (m, 1H), 4.73-4.78 (m, 1H), 4.78-4.82 (m, 1H), 5.96 (d, J=5.28 Hz, 1H), 6.46 (t, *J*=7.34 Hz, 1 H), 6.63 (d, *J*=8.22 Hz, 1 H), 6.99 (t, *J*=7.34 Hz, 1 H), 7.03 – 7.07 (m, 1 H), 7.09 (t, *J*=7.63 Hz, 1 H), 7.25 (s, 1 H), 7.31 (d, *J*=8.22 Hz, 1 H), 7.46 (d, *J*=7.63 Hz, 1 H), 7.74 (t, *J*=5.58 Hz, 1 H), 7.77 (d, *J*=8.22 Hz, 1 H), 8.12 (t, *J*=5.58 Hz, 1 H), 8.14 (d, *J*=8.22 Hz, 1 H), 8.23 (s, 1 H), 8.47 (s, 1 H), 8.84 (d, *J*=7.04 Hz, 1 H), 10.79 (br. s., 1 H).

HRMS: *m/z* calculated for $C_{42}H_{56}N_{11}O_{19}P_3S$: 1144.2760 [M+H]$^+$. Found: 1144.2759

Ant-L-*N*-Me-Trp-L-Ala-*S*-CoA:

HRMS: *m/z* calculated for $C_{43}H_{58}N_{11}O_{19}P_3S$: 1158.2917 [M+H]$^+$. Found: 1158.2917

Benz-D-Trp-L-Ala-*S*-CoA:

HRMS: *m/z* calculated for $C_{42}H_{55}N_{10}O_{19}P_3S$: 1129.2651 [M+H]$^+$. Found: 1129.2653

Sal-D-Trp-L-Ala-*S*-CoA:

HRMS: *m/z* calculated for $C_{42}H_{55}N_{10}O_{20}P_3S$: 1145.2600 [M+H]$^+$. Found: 1145.2601

4.5. Synthesis of *R*-32 and *S*-32

Isatoic anhydride (1.0 g, 6.13 mmol, 1.0 equiv) and tryptophan (L or D) (1.25 g, 6.13 mmol, 1.0 equiv) were suspended in DMSO (7 ml). The resulting mixture was refluxed for 1.5 hours, before being cooled to room temperature and partitioned between ethyl acetate (15 ml) and water (80 ml). The aqueous phase was then extracted with ethyl acetate (2 x 15 ml). The combined organic fractions were then washed with water (15 ml), 5% NaHCO$_3$ (2 x 15 ml), water (15 ml) and brine (15 ml). The organic layer was then dried over magnesium sulphate and the solvent removed *in vacuo*. The resulting residue was purified by flash column chromatography over silica eluting with 50-100% ethyl acetate/hexane to give the desired product.

A portion of the resulting white *R*-32 or off-white *S*-32 solids were then dissolved in acetonitrile and further purified by preparative HPLC on a Beckmann Coulter Gold system equipped with a reverse phase C18 column (Phenomenex Luna, 250 x 21.2 mm, 10 micron) with detection at 254 nm. Solvent system A (water plus 0.1% TFA) and B (acetonitrile plus 0.1%

TFA) held at 25% B for 1 min and then run over a linear gradient of 25-55% over 20 min, before increasing from 55-95% B over 1 min and a holding at 95% B for 5 min. The column was then equilibrated back to initial conditions by returning to 25% B and holding for 8 min. The peak with retention time of 18 minwas collected.

R-**32** or S-**32**: ^{1}H-NMR (400 MHz, MeOD-d_4) ppm: 3.14 (dd, J=14.87, 9.00 Hz, 1 H), 3.39 (dd, J=14.87, 5.87 Hz, 1 H), 4.09 (dd, J=8.61, 5.87 Hz, 1 H), 6.89 - 6.98 (m, 1 H), 7.05 (t, J=7.43 Hz, 1 H) 7.09 - 7.16 (m, 1 H), 7.13 (s, 1H), 7.22 (t, J=7.83 Hz, 1 H), 7.30 (d, J=7.83 Hz, 1 H), 7.43 (d, J=7.83 Hz, 1 H), 7.48 - 7.55 (m, 1 H), 7.78 (dd, J=7.83, 1.17 Hz, 1 H).

^{13}C-NMR (MeOD-d_4, 100 MHz) δ ppm 25.10, 54.70, 110.71, 112.50, 118.99, 119.98, 122.42, 122.62, 124.95, 125.94, 127.21, 128.52, 131.81, 134.28, 138.19, 138.22 171.10, 173.86.

HRMS: *m/z* calculated for $C_{18}H_{15}N_3O_2$: 306.1237 $[M+H]^+$. Found 306.1234

5. Chiral analysis of *R*-32 and *S*-32 standards and product of AnaPS

200 μM solutions of *R*-**32** and *S*-**32** were prepared by dissolving a known quantity of synthetic stndards in 50% acetonitrile/water. The product of AnaPS and AnaPS-E^0 was assessed by preparing a 50 μl reaction containing 1 μM AnaPS, 3 mM ATP, 2 mM $MgCl_2$, 0.5 mM TCEP, and 1 mM amino acid substrates (Ant and L-Trp) in NaP_i reaction buffer (50 mM NaP_i [pH 7.4], 100 mM NaCl, and 5% glycerol), which was incubated at 25 °C for 22 hours after addition of enzyme. The sample was then quenched by adding an equal volume of MeCN and removing the observed precipitate by centrifugation. 20 μl of each sample (*R*-**32** standard, *S*-**32** standard, product of AnaPS incubation, product of AnaPS mutant incubation) were injected onto a ChiralCel OD-RH cellulose tris column (150 × 4.6 mm) for homochiral HPLC analysis (with diode-array detection). The injected sample was separated at a flow rate of 0.5 ml min^{-1} using a

linear gradient of 25-65% MeCN in water over 30 min, followed by a ramp up to 95% MeCN over 1 min and a hold at 95% MeCN for 2.5 min.

6. HPLC-based time course study of product formation by TqaA and AnaPS

Reactions (300 µL) contained 1 µM enzyme, 3 mM ATP, 2 mM MgCl$_2$, 0.5 mM TCEP, and 1 mM amino acid substrates (Ant, L-Trp, and L-Ala for TqaA or Ant and L-Trp for AnaPS) in NaP$_i$ reaction buffer. Reactions were incubated at 25 °C and 50 µl aliquots were taken at 1, 2.5, 4.5, 6, and 22 hours after addition of enzyme and quenched by adding an equal volume of MeCN. Precipitate was removed by centrifugation and 20 µL samples were injected onto an Alltima C18 column (150 × 4.6 mm) for HPLC analysis (with diode-array detection). The injected sample was separated at a flow rate of 1 mL/min using a linear gradient of 25-55% MeCN in water (including 0.1% TFA) over 20 min, followed by a ramp up to 95% MeCN over 1 min and a hold at 95% MeCN for 5 min. Under these conditions **14** eluted at 16.2 min and *R*-**32** eluted at 11.6 min. Integration of the product peaks (**14** at 276 nm and *R*-**32** at 254 nm) was used to generate a plot of product peak area vs. time in order to approximate enzymatic rate. Initial rate data (obtained as integration area per hour) was converted to µM per hour using standard curves generated from 20 µl injections of **1** or *R*-**32** samples of known concentration.

7. Amino acid loading assay for TqaA and AnaPS

The loading of [^{14}C] labeled amino acids onto the T-domains of TqaA or AnaPS was performed to monitor thiolation activity and estimate percentage of the *holo* proteins (as expressed from *S. cerevisiae* BJ5464-NpgA). For TqaA, reactions contained 2 µM enzyme, 3 mM ATP, 2 mM MgCl$_2$ and 50 µM amino acid substrate (either [carboxy-^{14}C]Ant (0.13 µCi), L-[3-^{14}C]Trp (0.13 µCi), L-[3-^{14}C]Ala (0.14 µCi), or L-[U-^{14}C]Ala (0.32 µCi)) in Tris reaction

buffer. The reaction parameters were identical for AnaPS except that [^{14}C] was omitted from testing. Reactions were incubated at 25 °C and 50 µl of the reaction was quenched with 10% TCA (containing 50 µg BSA) at 15 and 30 min after addition of substrate. Precipitated protein was collected by centrifugation and the protein pellet washed twice with 10% TCA and dissolved in 80% formic acid for liquid scintillation counting. A ratio of nmoles radioactivity counted to nmoles protein was used to calculate % loading based on 1 equivalent of [^{14}C] amino acid labeling 1 equivalent of NRPS protein.

Table S8. Sequences of primers used in this paper

Primer	5′-Sequence-3′
TqaA_P1_F:	ATGGCTAGCGATTATAAGGATGATGATGATAAGACTAGTATGACGGTCCAATCTCTCGAG
TqaA_P1_R:	CCAACAAAGCAGGATGGTCACC
TqaA_P2_F:	CCACAATCTCTCCAACAGCCAG
TqaA_P2_R:	GATCAATTATCACGCCCTTGGG
TqaA_P3_F:	TACATCTGGCTCGACCGGCACA
TqaA_P3_R:	CATCAATGGGAAGGTCACGGCC
TqaA_P4_F:	TTGCTTCGGCTATGTCAAGTCT
TqaA_P4_R:	TCATTTAAATTAGTGATGGTGATGGTGATGCACGTGGAATGTAATATCCAATAATCTTGC
TqaA-ΔC$_T$_P4_R:	TCATTTAAATTAGTGATGGTGATGGTGATGCACGTGATCTTCCAATGTGCTTGTGAC
C$_T$-BspHI-F:	AAAAAA**TCATGA**TGGCAGTTGCCAAAGACAGCACC
C$_T$-EcoRI-R:	TTTTTT**GAATTC**TCAGTGGTGGTGGTGGTGGTGGAATGTAATATCCAATAATCT
AnaPS_P1F:	ATGGCTAGCGATTATAAGGATGATGATGATAAGACTAGTATGACAATCACTACAATGACT
AnaPS_P1R:	TCCGTTGCTCAAGTTCGCAG
AnaPS_P2F:	TGTTGGCTGAACTCGCTTCTG

AnaPS_P2R:	CATCGGGAAGCCAATGTCCC
AnaPS_P3F:	AGGTCTCGGATGTCGAGCAAA
AnaPS_P3R:	TCATTTAAATTAGTGATGGTGATGGTGATGCACGTGAGCAAAACCGCTGGCAAGAT
C_T_H173A-R:	AGAGACTCCATCCCACTGGGAGGCTGAAATCTGGATCGTGAGAAC
C_T_H173A-F:	GTTCTCACGATCCAGATTTCAGCCTCCCAGTGGGATGGAGTCTCT
TqaA S^{3765}A _F:	GTCGTTCTCACGATCCAGATTGCACACTCCCAGTGGGATGGAGTC
TqaA S^{3765}A _R:	GACTCCATCCCACTGGGAGTGTGCAATCTGGATCGTGAGAACGAC
TqaA S^{3765}H _F:	GTCGTTCTCACGATCCAGATTCACCACTCCCAGTGGGATGGAGTC
TqaA S^{3765}H _R:	GACTCCATCCCACTGGGAGTGGTGAATCTGGATCGTGAGAACGAC
AnaPS-E^0_F:	TCCCTTTTCATGATTGGACATGCCCTGGTACTTGATCTTGTCTCG
AnaPS-E^0_R:	CGAGACAAGATCAAGTACCAGGGCATGTCCAATCATGAAAAGGGA
TqaA H^{2658}A _F:	ATGATTCGCCTCGACATGAGTGCCGCAATTATGGACCGAACAACC
TqaAH^{2658}A _R:	GGTTGTTCGGTCCATAATTGCGGCACTCATGTCGAGGCGAATCAT
T_3_BspHI_F:	AAAAAA**TCATGA**TGAAGCAGCTCTCCACCGACGC
T_3_EcoRI_R:	TTTTTT**GAATTC**TCAGTGGTGGTGGTGGTGGTGATCTTTGATTGGGCTGAGGGA

*Nucleotides underlined and in bold represent restriction enzyme recognition sites.

Table S9. Fungal NRPSs domain architecture.

Strain	Broad designation	Domain Architecture	% of NRPSs ending with C_T domain
A. terreus NIH2624	ATEG_08448.1	C-A-T-C-A-T-C_T	66%
	ATEG_04322.1	A-T-C-C-A-C-A-T-C_T	
	ATEG_02944.1	C-A-C-A-T-C-Lanthionine synthetase C-like protein	
	ATEG_03576.1	A-T-C-A-T-C_T	
	ATEG_08427.1	A-T-C-A-T-C_T	
	ATEG_01002.1	A-T-C-C-A-T-C-A-C-A-R	
	ATEG_03528.1	A-T-C-A-T-C_T	
	ATEG_05073.1	A-T-C-A-T-C-A-T-C-T-C-T-C_T	
	ATEG_07488.1	A-T-C-T-C_T	
A. fumigatus Af293	Afu1g10380	A-T-C-C-A-A-T-C-A-T-C-C-T-C-T	62.5%
	Afu1g17200	A-T-C-A-T-C-A-T-C-T-C-T-C_T	
	Afu3g03350	A-T-C-A-T-C_T	
	Afu3g03420	A-T-C-A-T-C_T	
	Afu3g12920	A-T-C-A-T-C-T	
	Afu3g13730	A-T-C_T	
	Afu3g15270	A-C-A-T-C_T	
	Afu5g10120	A-T-R	
	Afu5g12730	A-T-C-A-T-E-C-A-T-C-E-T-C-A-T-C-A-T-C-A-T-E-C_T	
	Afu6g08560	A-T-R	
	Afu6g09610	C-A-T-C_T	
	Afu6g09660	A-T-C-A-T-C-T	
	Afu6g12050	A-T-C_T	
	Afu6g12080	A-T-C-A-T-C-E-A-T-C_T	

229

	Afu8g00170	A-C-A-T-C$_T$	
	Afu8g01640	A-T-R	
A. clavatus NRRL 1	ACLA_017890	A-T-C-A-T-E-C-A-T-C$_T$	90%
	ACLA_095980	A-T-C-A-T-E-C-A-T-C$_T$	
	ACLA_076770	A-T-C-A-T-C-A-T-C$_T$	
	ACLA_093780	A-T-C-A-T-C-A-T-E-C-A-T-C-A-T-C$_T$	
	ACLA_098420	A-T-C-A-T-E-C-A-T-E-C-A-T-E-C-A-T-C-A-T-C$_T$	
	ACLA_025160	A-T-E-C-A-T-C-A-T-C-A-T-E-C-T-C-T	
	ACLA_059530	A-T-E-C-A-T-C-A-T-C-A-T-C$_T$	
	ACLA_061000	A-T-C-A-T-C$_T$	
	ACLA_017900	A-T-C$_T$	
	ACLA_044390	C-A-T-C-A-T-C$_T$	
	CHGG_00041	A-T-R	
C. globosum CBS 148.51	CHGG_02251	A-T-C-A-T-C-A-T-C-T-C-T-C$_T$	75%
	CHGG_02283	T-C-C-A-T-C$_T$	
	CHGT_03491	C-A-T-C-A-T-C$_T$	
	CHGG_04477	T-C-T-C-A-T-C-A-T-C$_T$	
	CHGG_06052	C-A-T-R	
	CHGG_06789	C- A- T- E-C- A-T-C- A-T-E-C- A-T-C-A-T-E-C-T-C$_T$	
	CHGG_08540	A-T-C-A-T-C-A-T-E-C-A-T-E-C-A-T-E-T-C$_T$	
	CHGG_09475	A-T-C-A-T-C$_T$	
	CHGG_09543	A-T-C-A-T-C-T-C-A-T-C-T-C-T-C$_T$	
	CHGG_10057	A-T-C-A-T-C-A-T-C-A-T-E-C-A-T-E-C-A-T-E-C-A-T-C$_T$	
	CHGG_10135	A-T-R	

Figure S20. Scheme of *in vivo* homologous recombination for assembly of *tqaA* gene into the 2-*μ*m vector for *S. cerevisiae* expression. The sizes of the overlapping regions between two flanking pieces are within 30 ~ 40 bps.
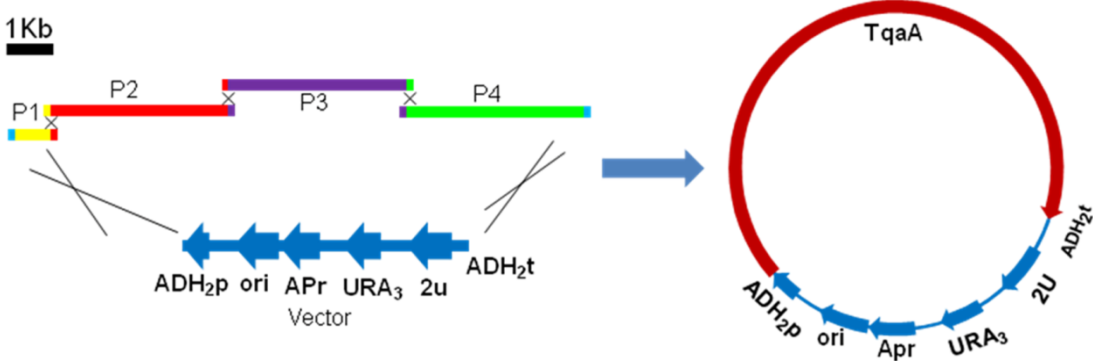
Figure S21. Gel filtration chromatography for TqaA and AnaPS. TqaA/AnaPS purified from ANTI-FLAG® column was loaded onto a Superdex 200 (GE Healthcare) column and eluted with a flow rate of 0.5 ml/min in 50 mM Tris pH 8.0 and 100 mM NaCl. Fractions 3 and 4 were collected for further assays.
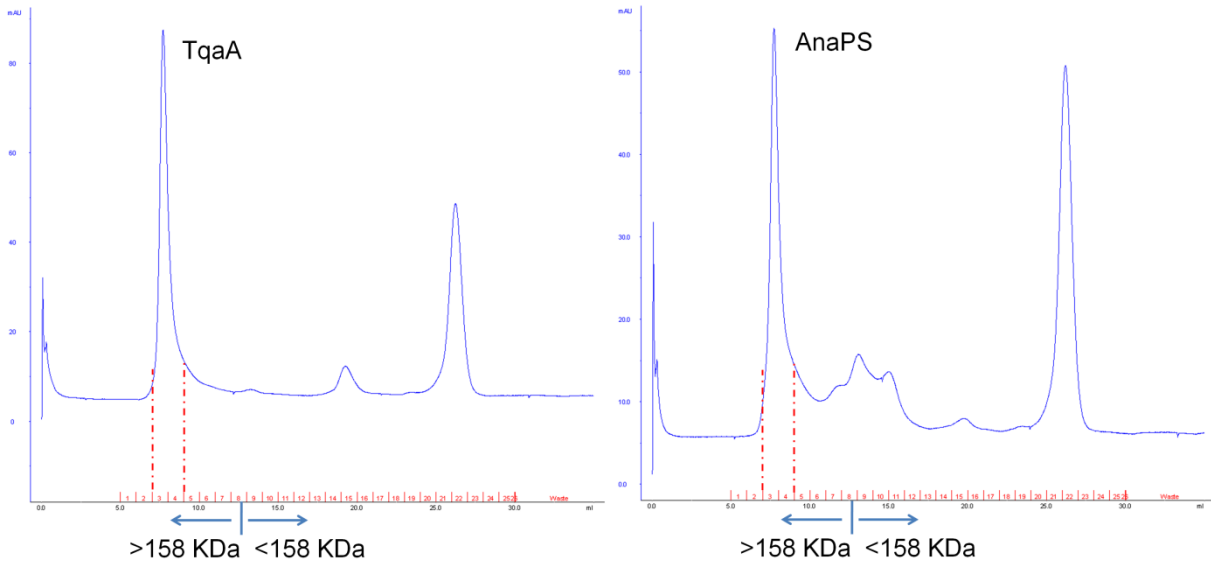
Figure S22. SDS-PAGE of the heterologously expressed proteins in this study. TqaA (450kDa), AnaPS (264kDa), TqaA-ΔC_T (400 kDa), TqaA H^{3766}A (450 kDa), TqaA H^{2658}A (450 kDa) and AnaPS-E^0 (264 kDa) were expressed from BJ5464-NpgA and purified by using ANTI-FLAG® M1 Agarose Affinity Gel; C_T (54 kDa) and T_3 (11 kDa) were expressed from *E. coli* BL21 (DE3) with C-terminal His-tag and purified by using Ni-NTA agarose affinity resin.
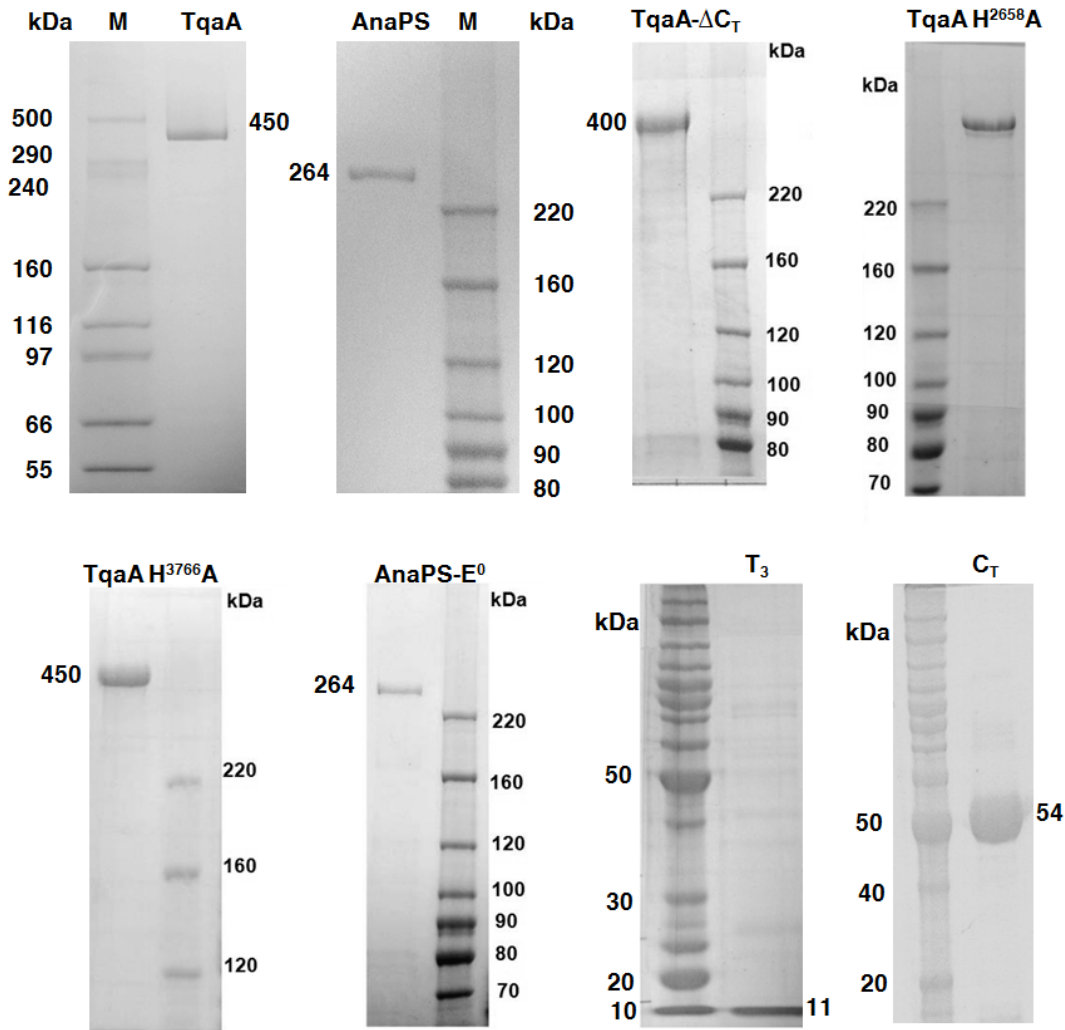
Figure S23. HPLC-based time course study of TqaA and AnaPS in the turnover of **1** and *R*-**2**, respectively. Assays were performed using (**a**) freshly purified TqaA; (**b**) stored (-80 ºC) and thawed TqaA; (**c**) AnaPS (both freshly purified and storage AnaPS displayed nearly the same activity). Each data point was repeated three times. Only data points within the linear range were used to calculate the initial product formation rates. Data represent mean values ± s.d.
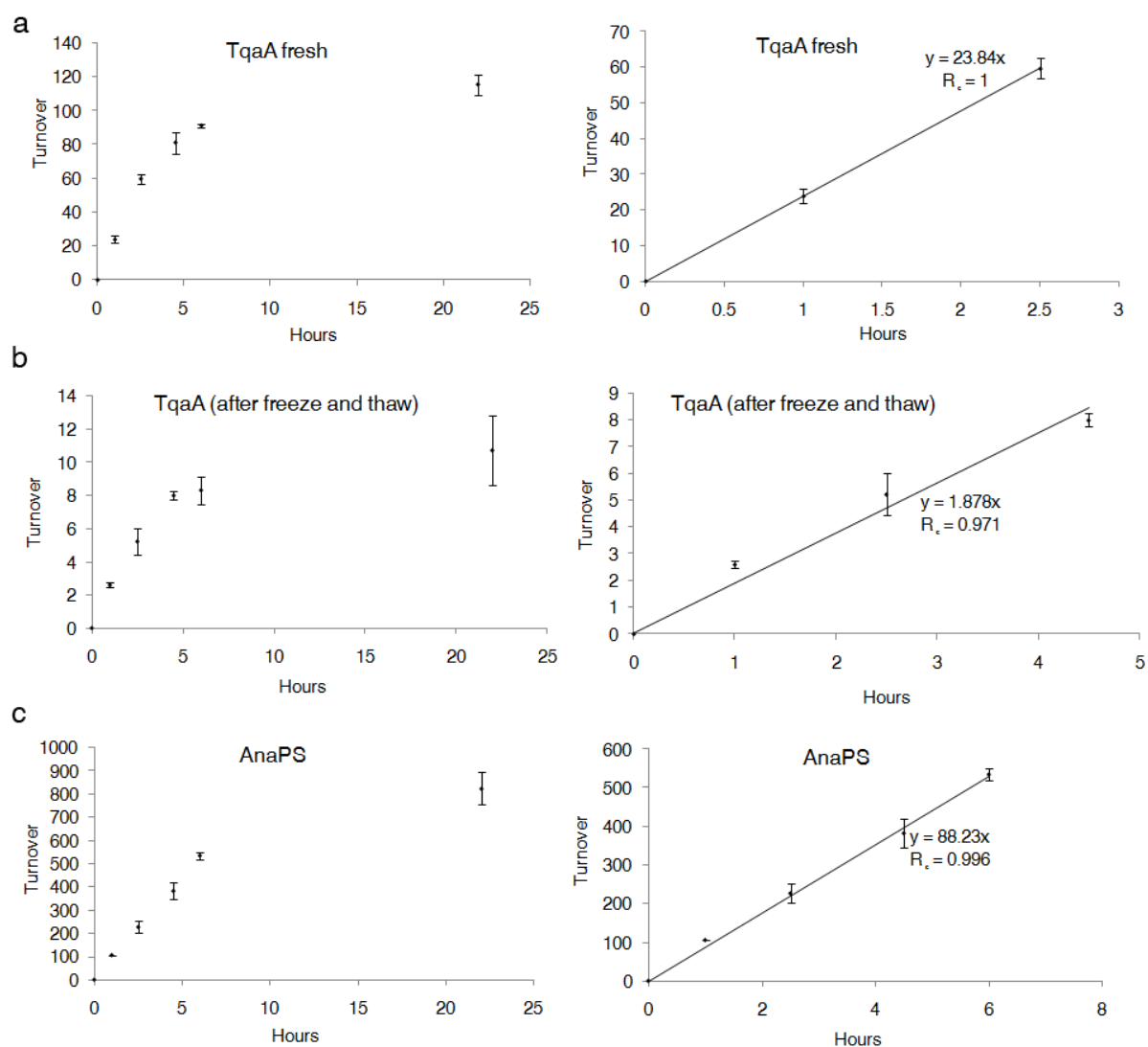
Figure S24. Comparison of the reaction rates of SNACs either with or without enzyme. SNAC reactions (400 µl) contained 0.5 mM TCEP and (**a**) 1 mM D-**33** SNAC without enzyme; (**b**) 1 mM D-**33** SNAC with 1 µM AnaPS; (**c**) 1 mM L-**33** SNAC without enzyme; (**d**) 1 mM **31**-SNAC without enzymes. **31**-SNAC does not undergo spontaneous cyclization to afford **14**, but can undergo hydrolysis to release the free acid **31**. HPLC traces are shown with λ = 272 nm.
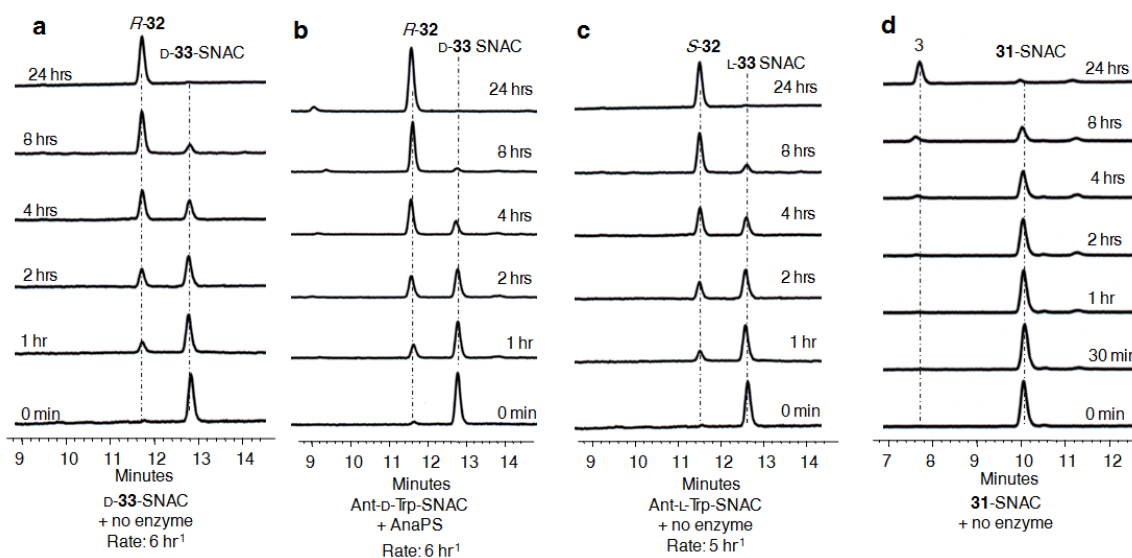
Figure S25. Characterization of TqaA and AnaPS T domain loading assay with [^{14}C] labeled amino acid.
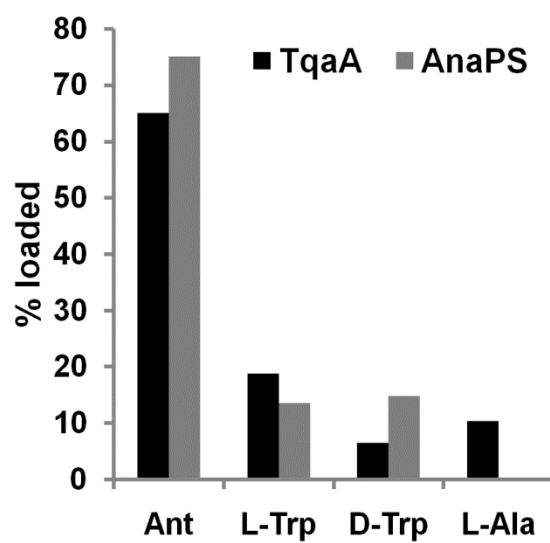
Figure S26. *R*-**32** is not produced by TqaA under *in vitro* assay conditions. HPLC traces (λ = 272 nm) shown here are organic extracts of the following *in vitro* assays: i) 2 mM of Ant, L-Trp, L-Ala and 10 µM TqaA; ii) 2 mM of Ant, L-Trp, L-Ala and 10 µM TqaA C_3^0 (containing the inactivating $H^{2658}A$ mutation); iii) 2 mM of Ant, L-Trp and 10 µM TqaA (no L-Ala added). The inability to produce *R*-**32** in ii and iii suggests C_3 may structurally protect the dipeptide from being cyclized into *R*-**32** spontaneously.
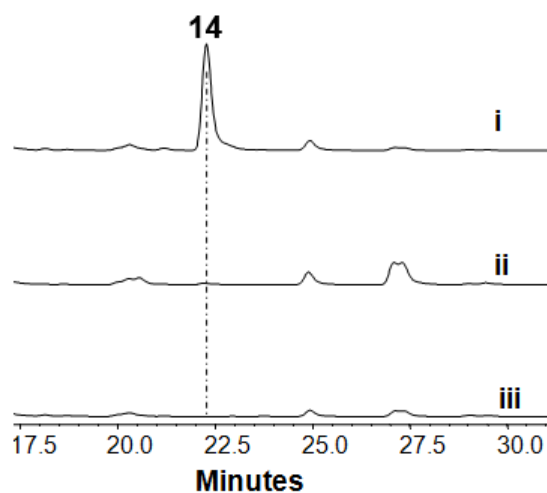
Figure S27. (a) Neighbor-joining phylogenetic analyses of C domains and C_T domain in SimA, Aba1 and TqaA NRPSs; **(b)** Sequences alignment of TqaA C domains and C_T domain. Sequence alignment and tree construction were performed using ClustalW[4].
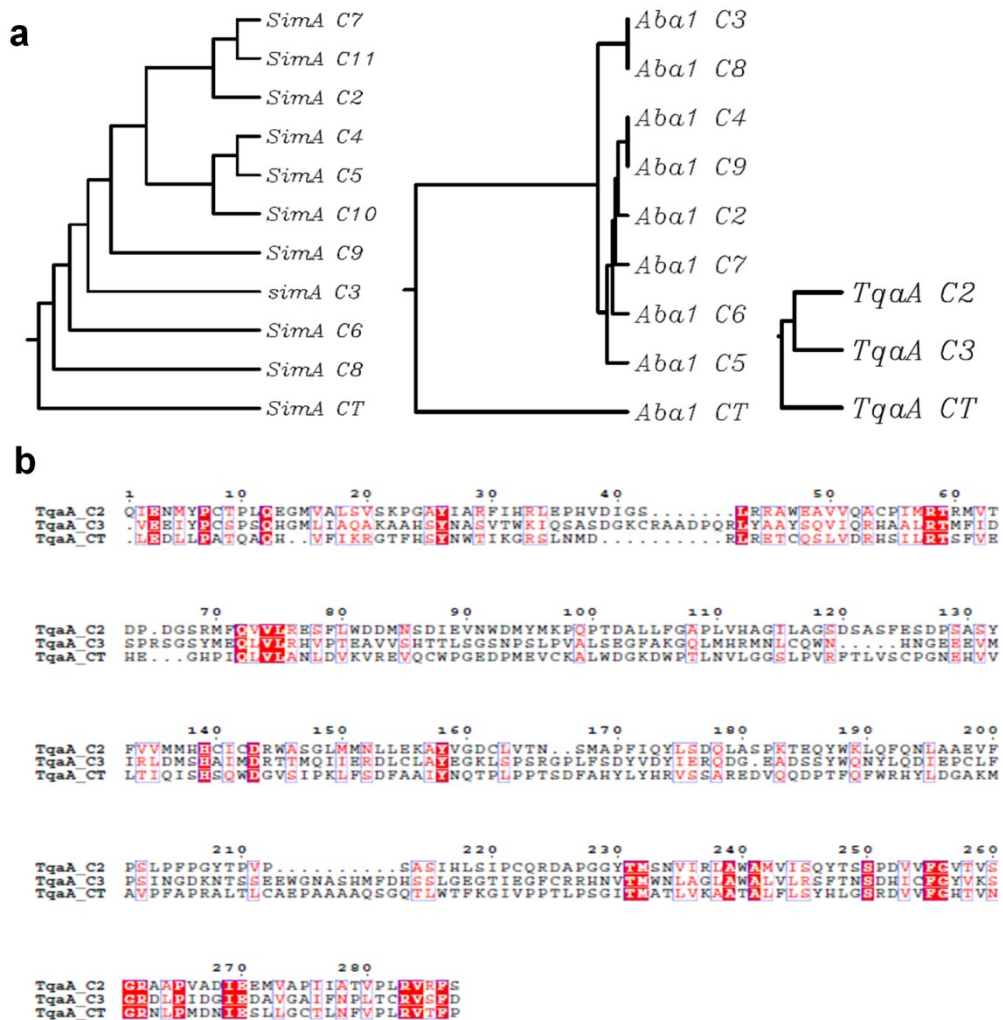
Figure S28. Neighbor-joining phylogenetic tree of select fungal C-domains and ACV synthetase terminal TE domain. Sequence alignment and tree construction were performed using MEGA [241].
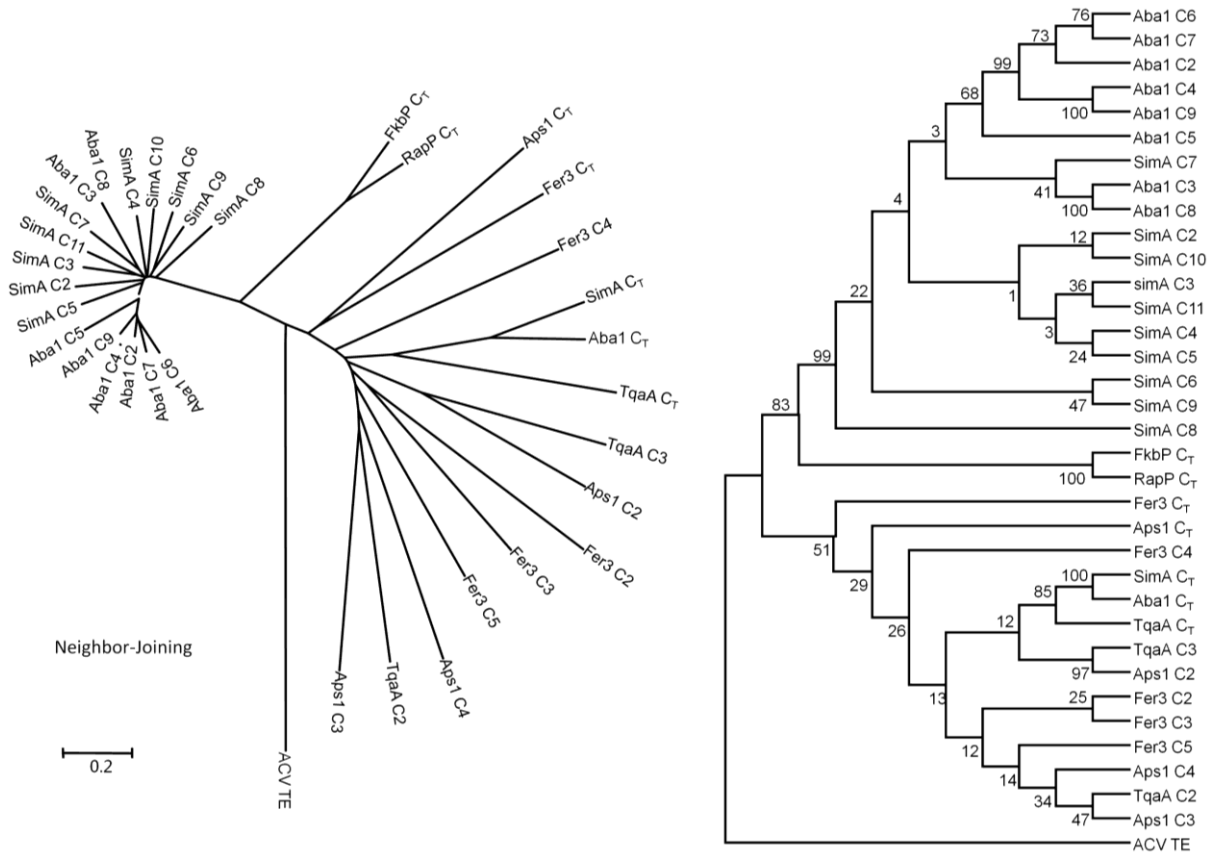
Figure S29. Sequence alignment of the active sites of C_T domains with other fungal NRPS C domains (using the canonical NRPS C domains from the cyclosporine synthetase SimA). All of the chain-extending C domains in SimA containing the "**HHXXXDXXS/T**" motif. In contrast, TqaA Aba1 and SimA C_T domains have another amino acid (**S** or **N**) in place of the first **H**. Sequence alignment was performed using ClustalW [242].

```
SimA_C8   HHMFSDGWSVDILRQELGQFYSAALRGRDPLS
SimA_C10  HHIISDGWSVDIFQQELAQFYSVAVRGHDPLS
SimA_C11  HHIISDGWSTEVLQRELGQFYLAAKSGKAPLS
SimA_C9   HHIISDGWSVDILRQELGQLYSNAS......S
SimA_CT   SHSLYDGLSLEHIVNALHALYSDKH......L
Aba1_CT   SHALYDGLSFEHIIQSLHALYLDIT......L
TqaA_CT   SHSQWDGVSIPKLFSDFAAIYNQTP......L
Aps1_CT   NHAAYDAWSLGMMLRSIGQYYANPRD....DS
Fer3_CT   HHTSYDAWSMRLMADELMQLYHNIDQG...KL
```

Figure S30. Analysis of TqaA C_T active site mutants by *in vivo* product formation and *in vitro* assay. The assays show TqaA H[3766] is essential for cyclization of **1**, while S[3765] in not. Traces (i-iv) shown here are HPLC analyses (λ = 272 nm) of metabolites extracted from 3-day cultures of i) BJ5464-NpgA expressing TqaA; ii) BJ5464-NpgA expressing TqaA H[3766]A; iii) BJ5464-NpgA expressing TqaA S[3765]A; and iv) BJ5464-NpgA expressing TqaA S[3765]H. v) TqaA H[3766]A was purified and subjected to *in vitro* assay. HPLC analysis of the organic extract from the reactions containing: 2 mM of each amino acid building block and 10 µM TqaA H[3766]A.
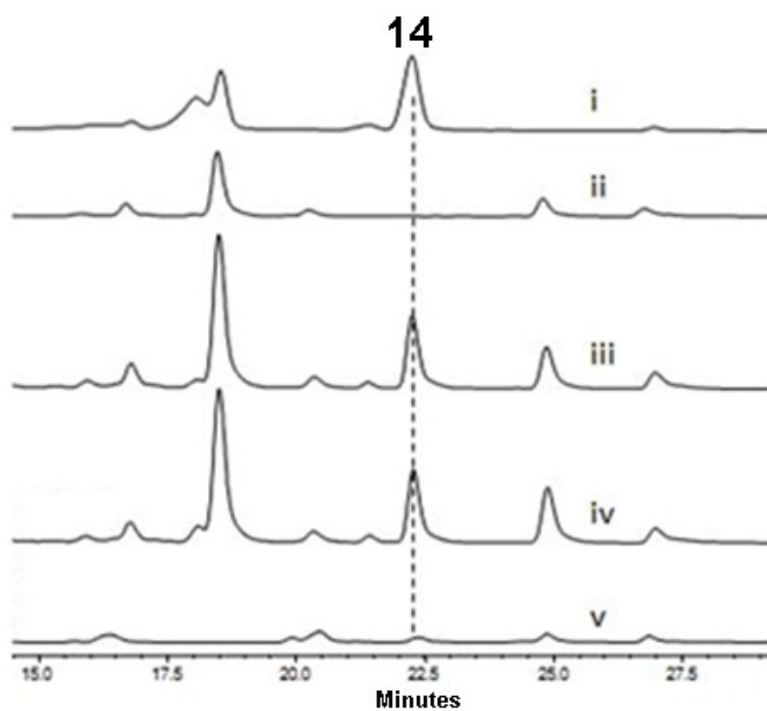

241

Figure S31. Assay of TqaA C_T domain using **31** attached to small molecule thioester carriers. The assays show that C_T cannot cyclize **31**-SNAC or **31**-*S*-CoA to product **14**. 20 μM C_T was used in all assays. HPLC traces (λ = 272 nm) shown here are extracts from *in vitro* reaction containing ii) 200 μM **31**-SNAC and iii) 200 μM **31**-*S*-CoA. Trace i and iv are the standards of **31**-SNAC and **14**, respectively. Hydrolysis of **31** can be detected in both reactions.
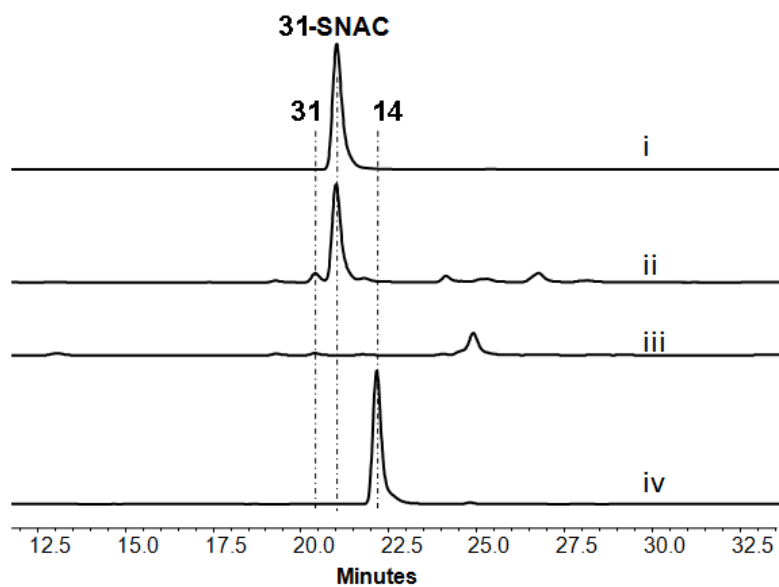
Figure S32. MALDI-TOF mass analyses of the *apo* T_3 and **31**-*S*-T_3. **31**-*S*-T_3 was prepared from an *in vitro* reaction containing *apo* T_3, Sfp and **31**-*S*-CoA. *Apo* T_3 has a $[M+H]^+$ mass value of 11197 and **31**-*S*-T_3 has a $[M+H]^+$ mass value of 11913.
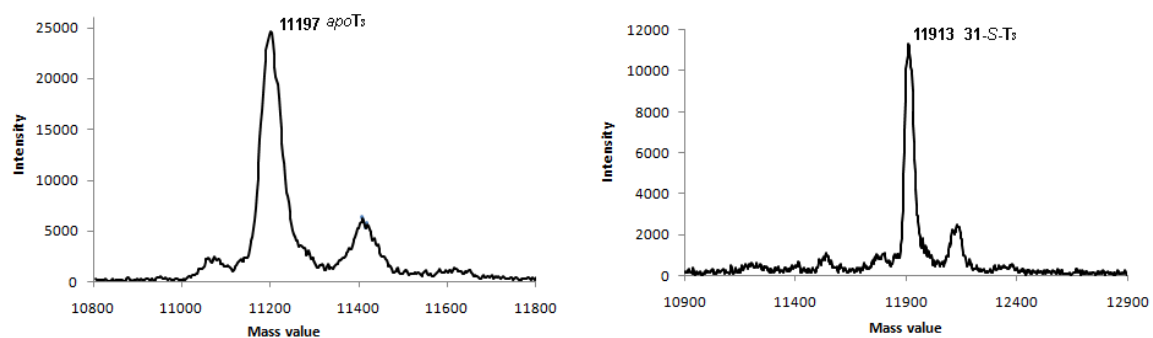

243

Figure S33. Kinetic properties of the C_T domain in cyclizing **14** from **31**-*S*-T$_3$. Initial turnover rates (V/E$_o$) as a function of **31**-*S*-T$_3$ concentration were measured. The concentration of C_T domain was kept at 2 μM. The data were fitted to the Michaelis-Menten equation using nonlinear least squares regression. Each data point was repeated three times. Data represent mean values ± s.d.
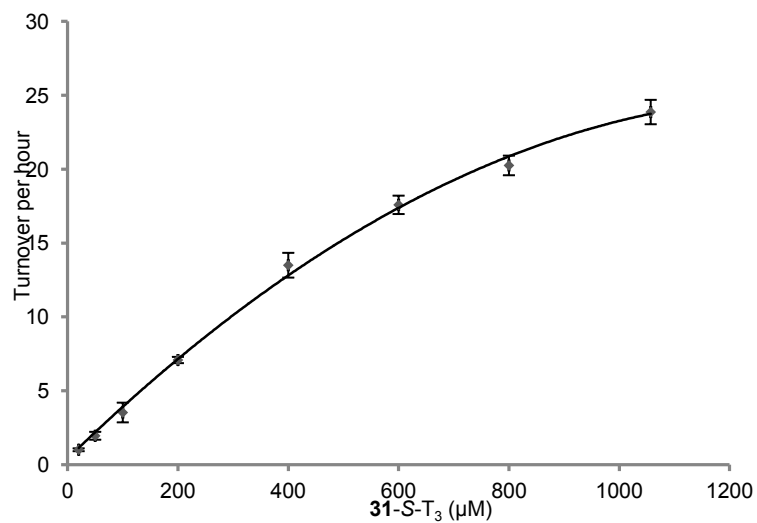
Figure S34. C_T can interact with T_{ApdA} to produce **14** but not with T_{LovF}. In the *in vitro* assays, 200 μM of **31**-*S*-CoA was used as the substrate. All assays contained 20 μM C_T. Traces shown are HPLC traces (λ = 272 nm) of *in vitro* reactions i) with addition of 200 μM T_{ApdA}, C_T was able to cyclize and produce **14** with a relatively low yield, while significant amount of the hydrolysis product **31** was observed; and ii) with addition of 200 μM T_{LovF}, only hydrolysis of **31**-*S*-CoA to **31** free acid can be observed. Trace iii is that of **14** standard.
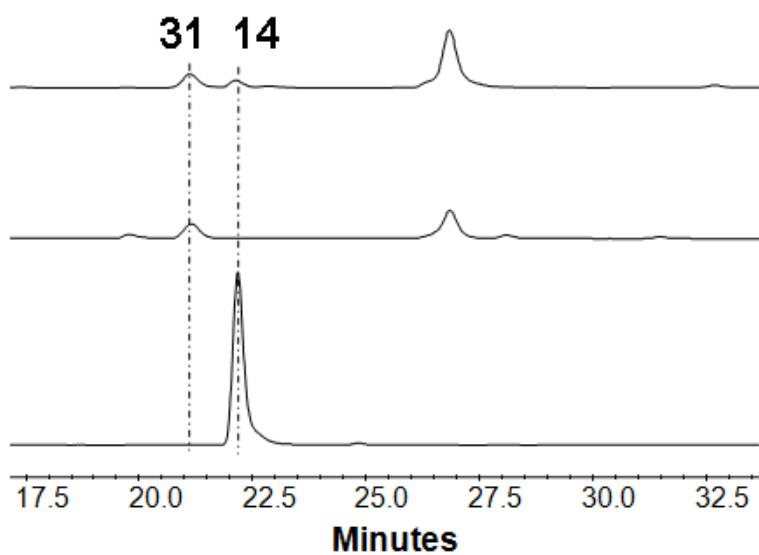
Figure S35. HPLC traces and 2D UV-vis spectra of the tripeptidyl-CoA analogues. High resolution mass (HRMS) analyses showed the measured masses of the synthetic compounds were consistent with the caculated masses. (Sal: Salicylate; Benz: benzoate; abrine: *N*-methyl-tryptophan)

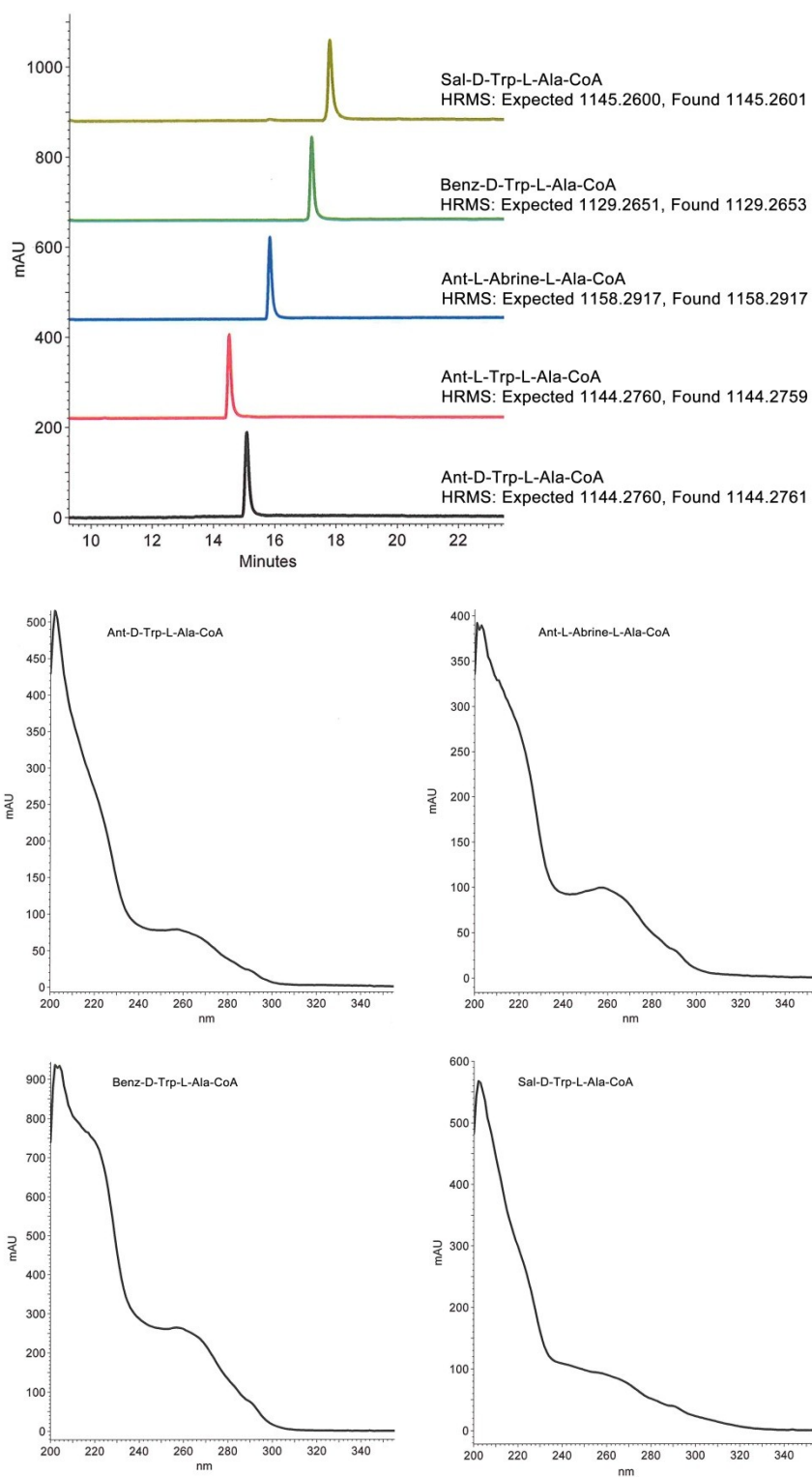

246

Figure S36. High resolution mass analysis of the products from TqaA C_T domain reactions: *N*-methyl-10 membered macrolactam, quaternary cationic product, salicylate-D-Trp-L-Ala free acid and benzoate-D-Trp-L-Ala free acid.
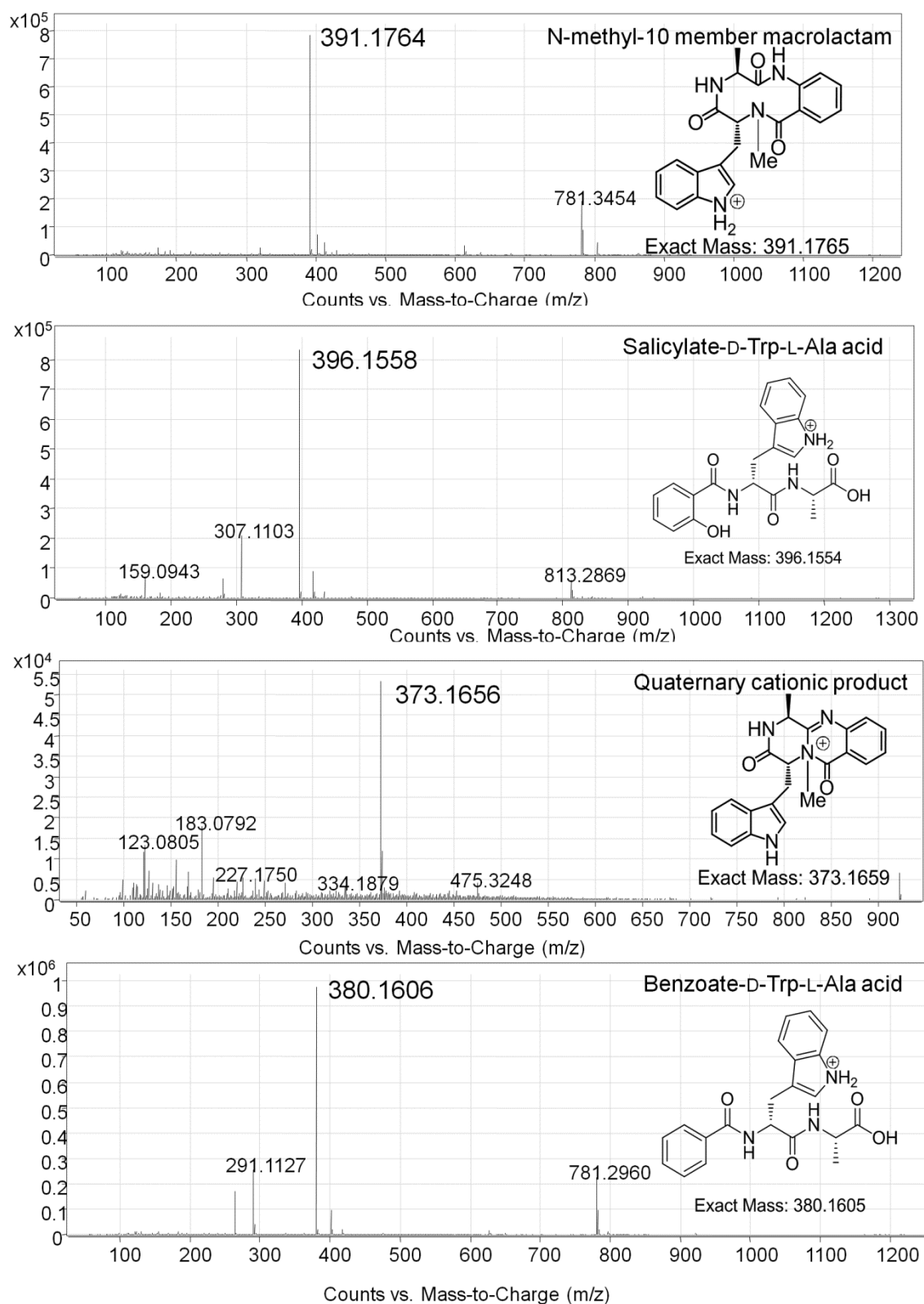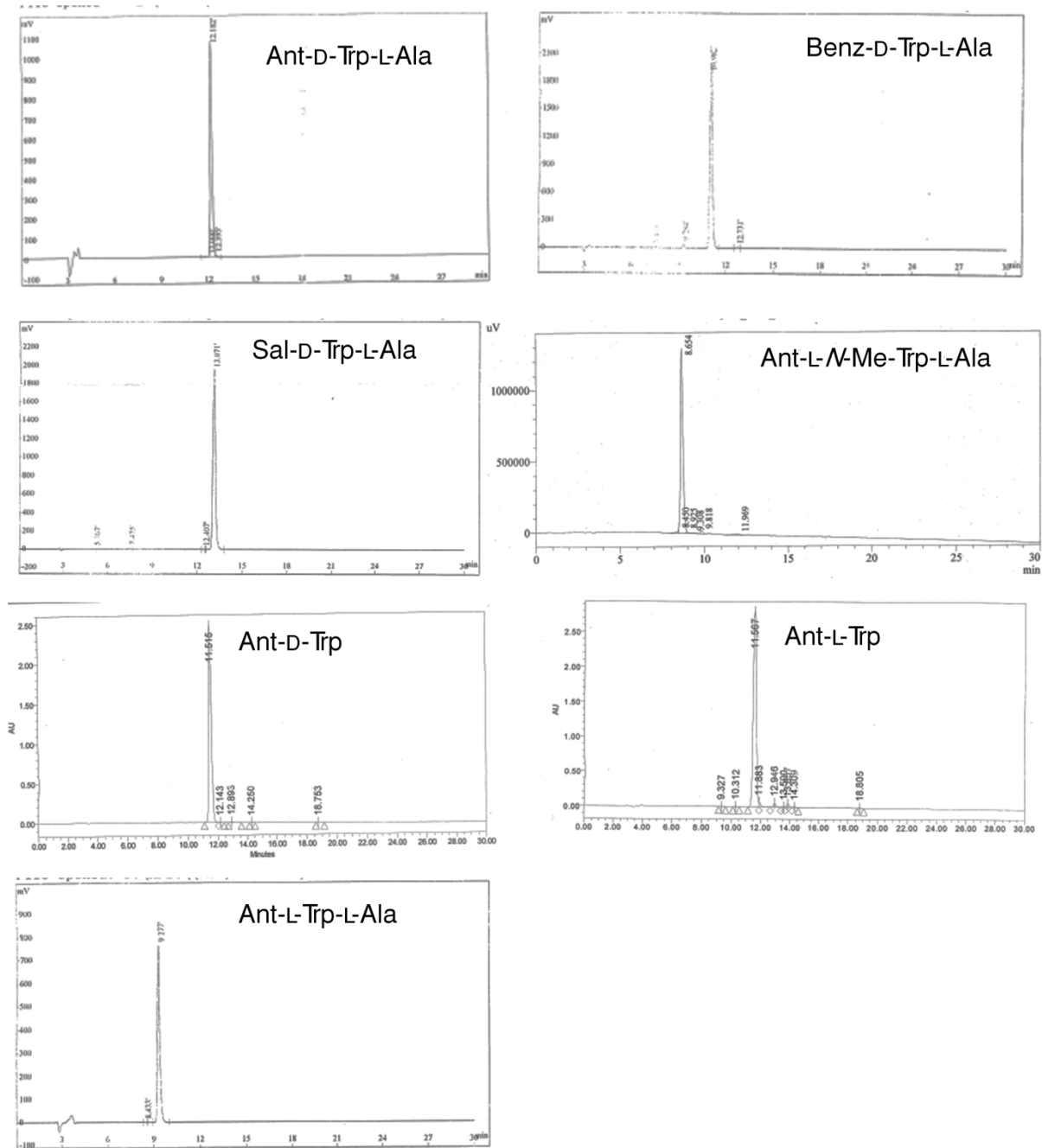

247

Figure S37. HPLC assessments of the purity of the peptides as supplied by the vendor (vendor information: RS Synthesis, LLC. PO Box 70301, Louisville, KY 40270, Phone: 502-614-5920, Fax: 801-780-2235)

4.3   Supporting Information for Section 2.4

Table S10. List of primers used during this study

Primer name	Sequence (5'-3')
AspA_P1_F	atggctagcgattataaggatgatgatgataagactagtatgggttcctacaacgccaa
AspA_P1_R	caccattccgtcagcccatg
AspA_P2_F	catggctcaatagtagggtcc
AspA_P2_R	gccattttcagccgatctaa
AspA_P3_F	gtgtggaaactgaagacaaggtc
AspA_P3_R	cagatcaacacgggtgatga
AspA_P4_F	gacgaagccgttggaccctt
AsPA_P4_R	ttcgctttgtccacaatcca
AspA_P5_F	ttgcgatgggatgtaaactg
AsPA_P5_R	tggtcgacatgctgtagcca
AspA_P6_F	atgtctcatgcacaatatga
AsPA_P6_R	tcatttaaattagtgatggtgatggtgatgcacgtgttgatatccattcaatgcat
AspA_C_T_NcoI_F	aaaaaaccatggtgggcagccatcatgactcgtca
AspA_C_T_EcoRI_R	tttttttgaattctcagtggtggtggtggtggtgttgatatccattcaatgcatg
AsPA_$C_2A_2T_2C_T$_F	atcaactatcaactattaactatatcgtaataccatatgcaggatatctttcactgcatg
AspA_T_2_NcoI_F	aaaaaaccatggtgaactacatatcaaaccaaagg
AspA_P1_NdeI_F	atcaactatcaactattaactatatcgtaataccatatgggttcctacaacgccaa
AsPA_$A_1T_1C_2$_R	tcatttaaattagtgatggtgatggtgatgcacgtgtcgctcagtacatagattgag

Figure S38. Extracted ion mass chromatograms from (a) asperlicin profile from wild type (b) profile from *aspA* (bimodular NRPS) deletion mutant (c) profile of *aspA* deletion mutant after feeding with synthetic asperlicin C (d) profile of *aspA* deletion mutant after feeding with synthetic asperlicin D (e) profile of *aspB* (monooxygenase) deletion mutant.

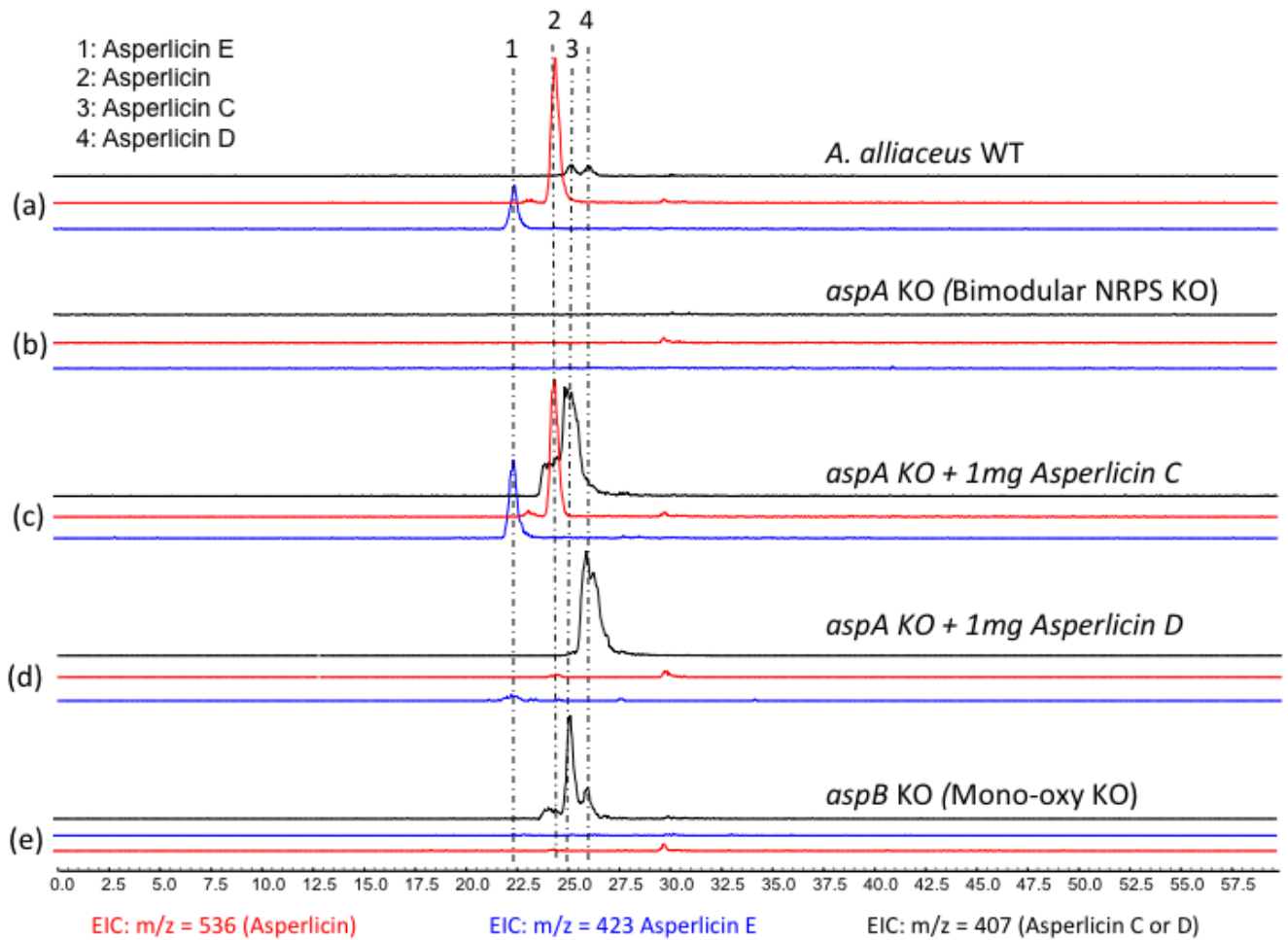**Figure S39.** (a) Comparison of asperlicin profiles of wildtype *A. alliaceus*, *aspB* mutant and *aspC* mutant (b) schematic of proposed competition for AspB-generated epoxy intermediate between valyl-AspC to give asperlicin and the intramolecular collapse to asperlicin E.
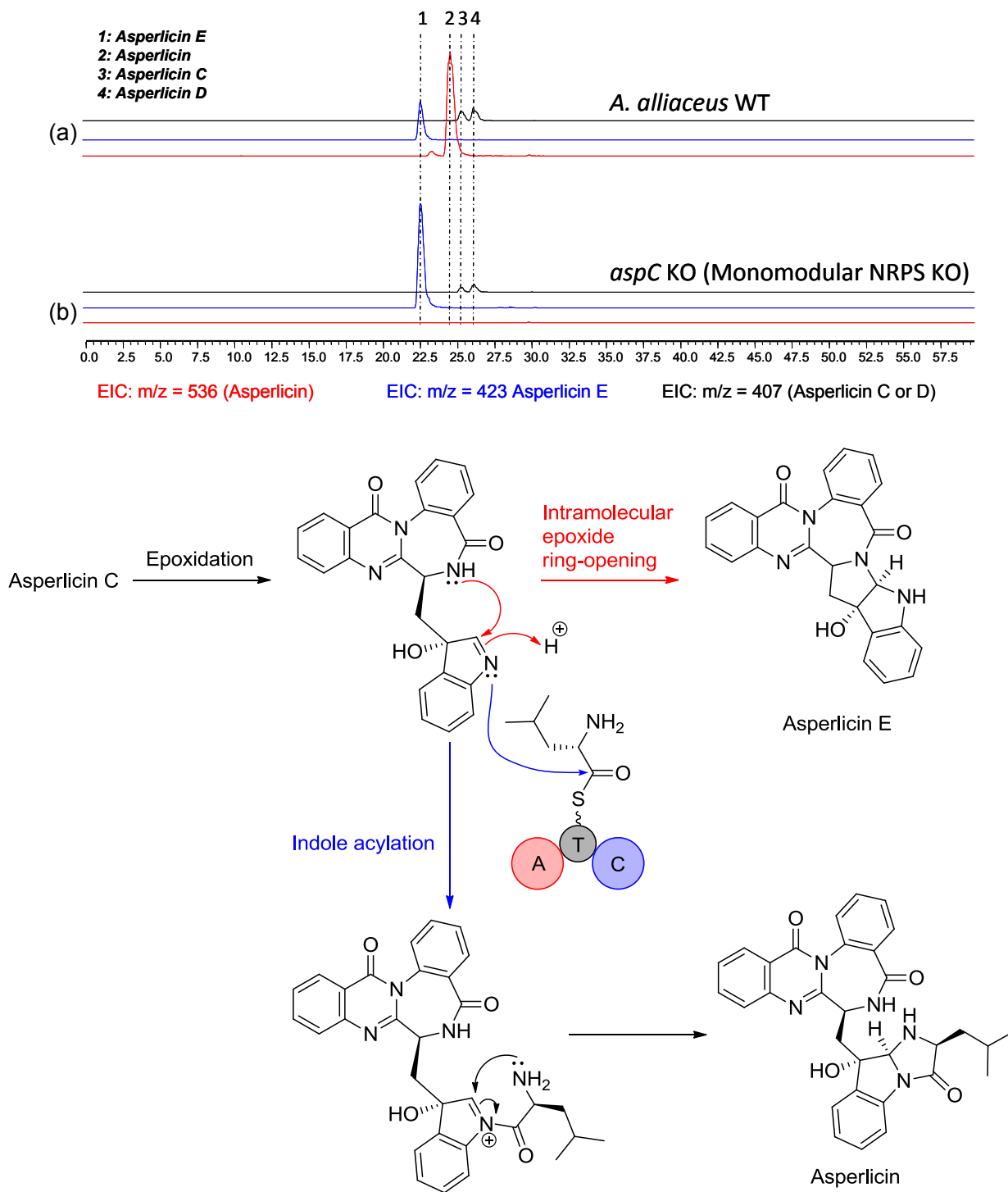
Figure S40. Michaelis-Menten plots of ATP-[^{32}P]PPi exchange assay for AspA with (**A**) Ant, (**B**) L-Trp, and (**C**) benzoic acid as substrates.
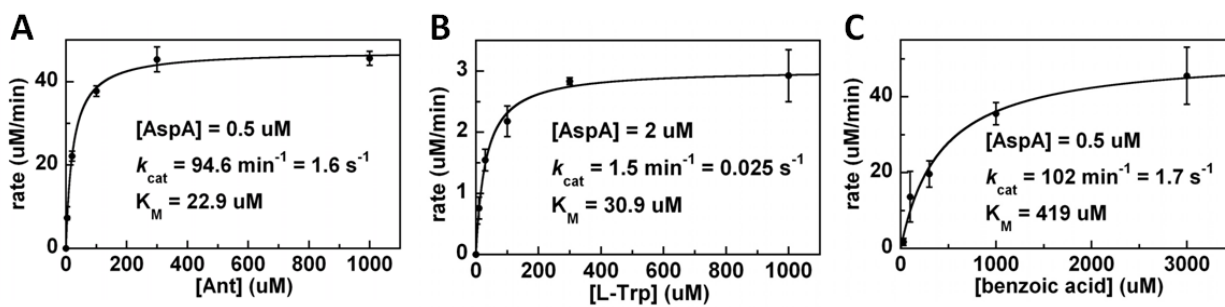
Figure S41. SDS-PAGE gels of the heterologously expressed proteins in this study. AspA_ A_1-T_1-C_2 and AspA_C_2-A_2-T_2-C_T were expressed from BJ5464-NpgA. AspA_C_T and AspA_T_2-C_T were expressed from BL21 (DE3). All proteins were purified by C-terminal His$_6$ tag.
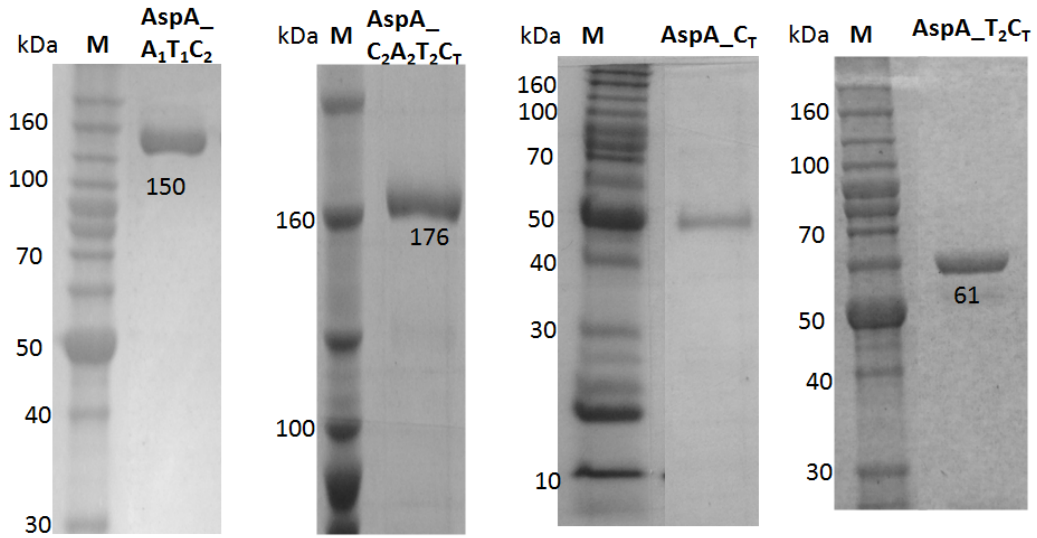
Figure S42. Extracted ion mass chromatograms [M+H]$^+$ =407 from AspA *in vitro* reactions. 1 mM of Ant, L-Trp, 3 mM ATP and 5 mM MgCl$_2$ were used in the assays. Trace i) No AspA reaction control; and trace ii) with addition of 10 μM AspA enzyme.
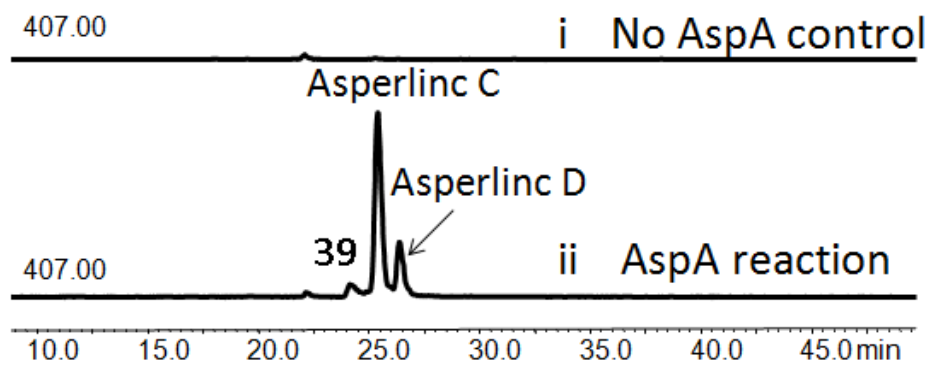
Figure S43. *In vitro* reconstitution of the Asperlicin C, D and **39** production by using dissected AspA_M1 (A_1-T_1-C_2) and AspA_M2 proteins (C_2-A_2-T_2-C_T). 1 mM of Ant, L-Trp, 3 mM ATP and 5 mM MgCl$_2$ were used in the assays. HPCL (280 nm) analyses of trace i) Intact AspA positive control, trace ii) only with 10 μM AspA_M1, trace iii) only with 10 μM AspA_M2 and trace ii) with addition of 10 μM AspA_M1 and AspA_M2 proteins.
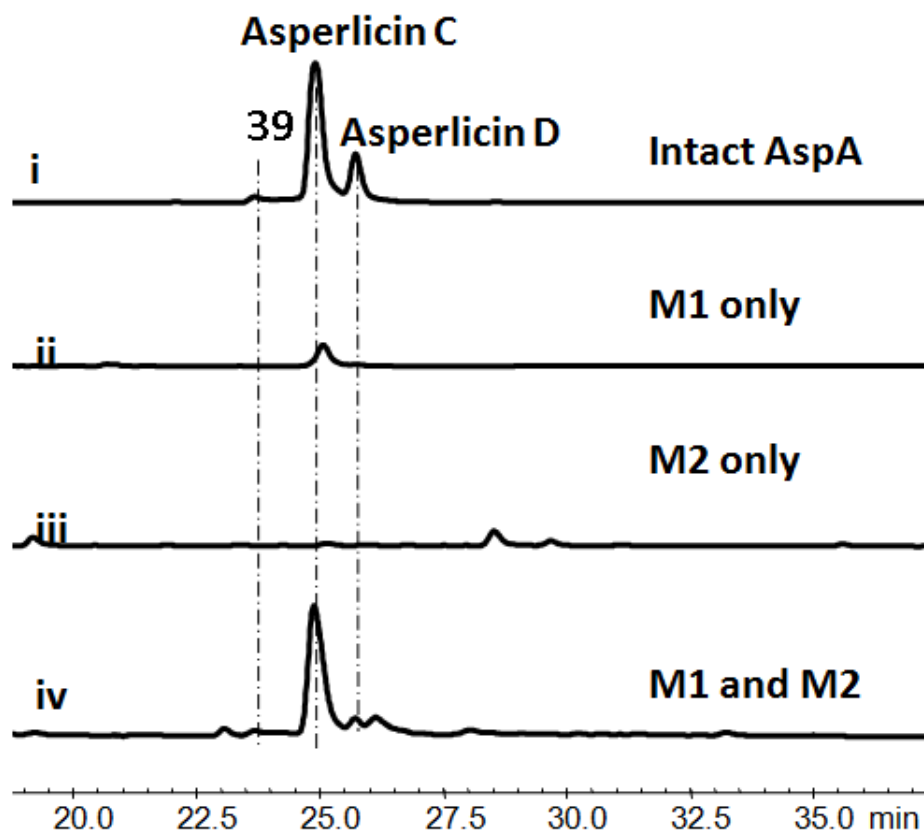
Figure S44. The first module of AspA iteratively utilizes two molecules of Ant. The *holo* form of M1 (50 μM A_1-T_1-C_2) was preloaded with Ant (1 mM Ant, 3 mM ATP) for 1 hour while in parallel the T_2-C_T di-domain was converted to the *holo* (HS-pantetheinyl) form via 20 μM Sfp and 1mM CoASH for 1 hour. The two solutions were mixed on addition of 400 μM Ant-L-Trp-SNAC and incubated overnight before aliquots were analyzed by LC/MS. Extracted ion mass chromatograms $[M+H]^+$ =407 from the assay without T_2C_T di-domain (trace i); from incubation including *holo* T_2C_T (trace ii), and from the full length AspA starting from Ant, L-Trp and ATP (trace iii).
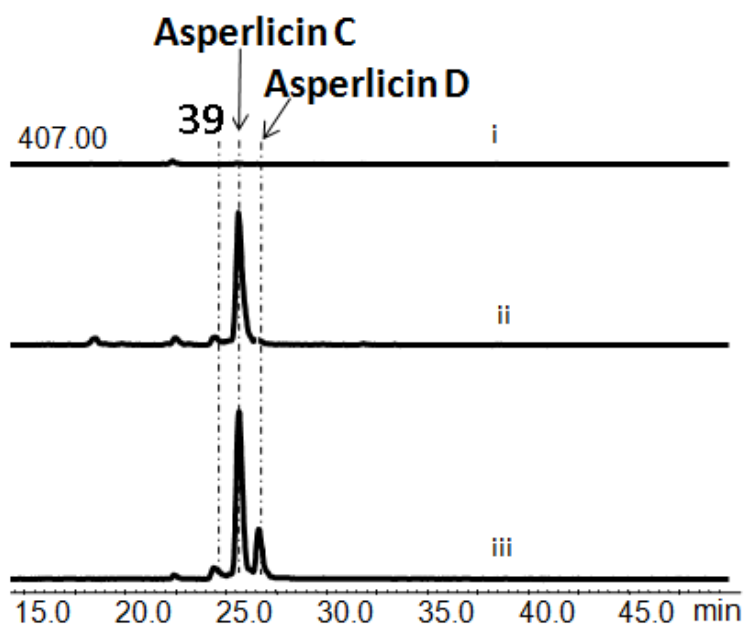
Figure S45. The T_2C_T di-domain fragment of AspA generates Asperlicin C and D from exogenous Ant-Ant- L-Trp-SNAC. Reactions contained 50mM This-HCl buffer, pH 7.5, in 100 μL. Traces (i to iii) contained 20 μM Sfp, 2mM CoASH, 100 μM Ant-Ant-L-Trp-SNAC. 50 μM T_2C_T or C_T was added in traces i and ii, respectively; No enzyme was added in trace iii. The suite of asperlicins C, D and compound 1([M+H]$^+$ = 407) are found in traces i but not ii and iii. Extracted ion mass chromatograms [M+H]$^+$ =407.

257
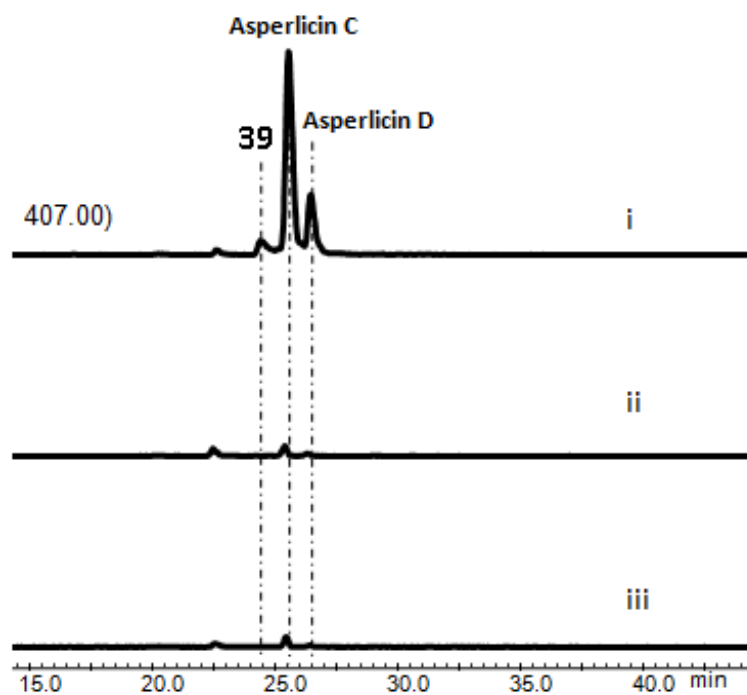
Figure S46. No cyclic Ant-Ant dimer was observed when incubate AspA_A_1-T_1-C_2 with Ant and ATP. HPCL traces (280 nm) are shown i) no enzyme control; ii) with addition of1 mM Ant, 3 mM ATP and 10 μM AspA_A_1-T_1-C_2.
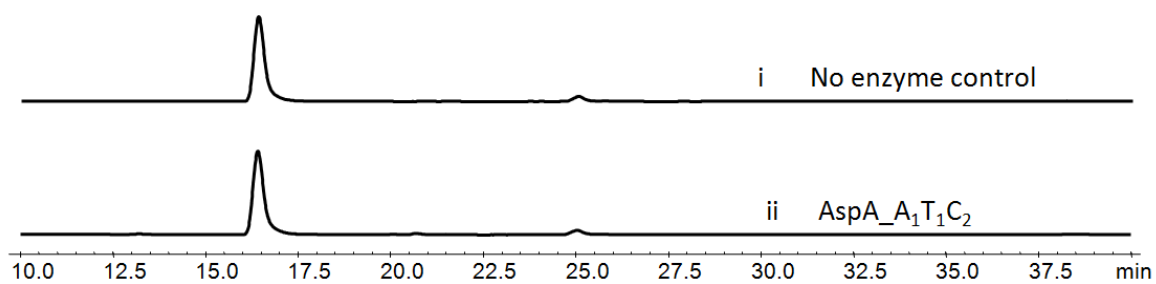
Figure S47. Three alternative mechanisms for the formation of 6,11-macrocycle precursor of asperlicins C and D. The calculations were performed at the PCM(water)/B3LYP/6-31G(d) level [184, 185] using reduced models in which the indol ring from Trp side chain has been substituted by a hydrogen. Relative free energies are in kcal mol[-1] and distances in Angstroms.
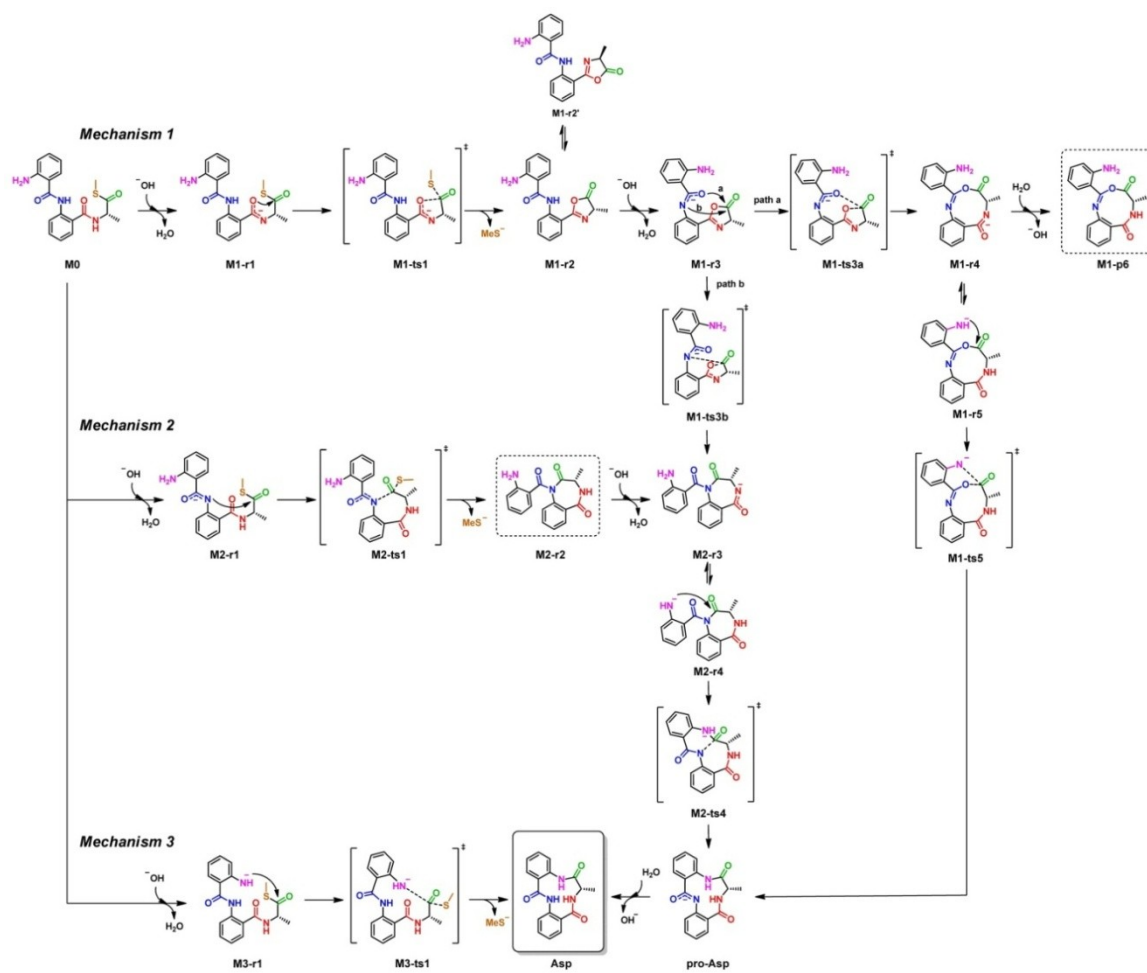
In *mechanism 1*, the $C_4=O$ carbonyl oxygen (imidate form) first attacks the thioester to generate an oxazol-5(4H)-one, followed by attack of the $C_{11}=O$ carbonyl oxygen (path a) and subsequent ring expansion to yield a 9-membered ring, which can then be opened to the 6,11-macrolactam. Another possible reaction pathway (path b) involves the attack of N_1 to the oxazolone to yield a 7-membered ring accessible also from mechanism 2.
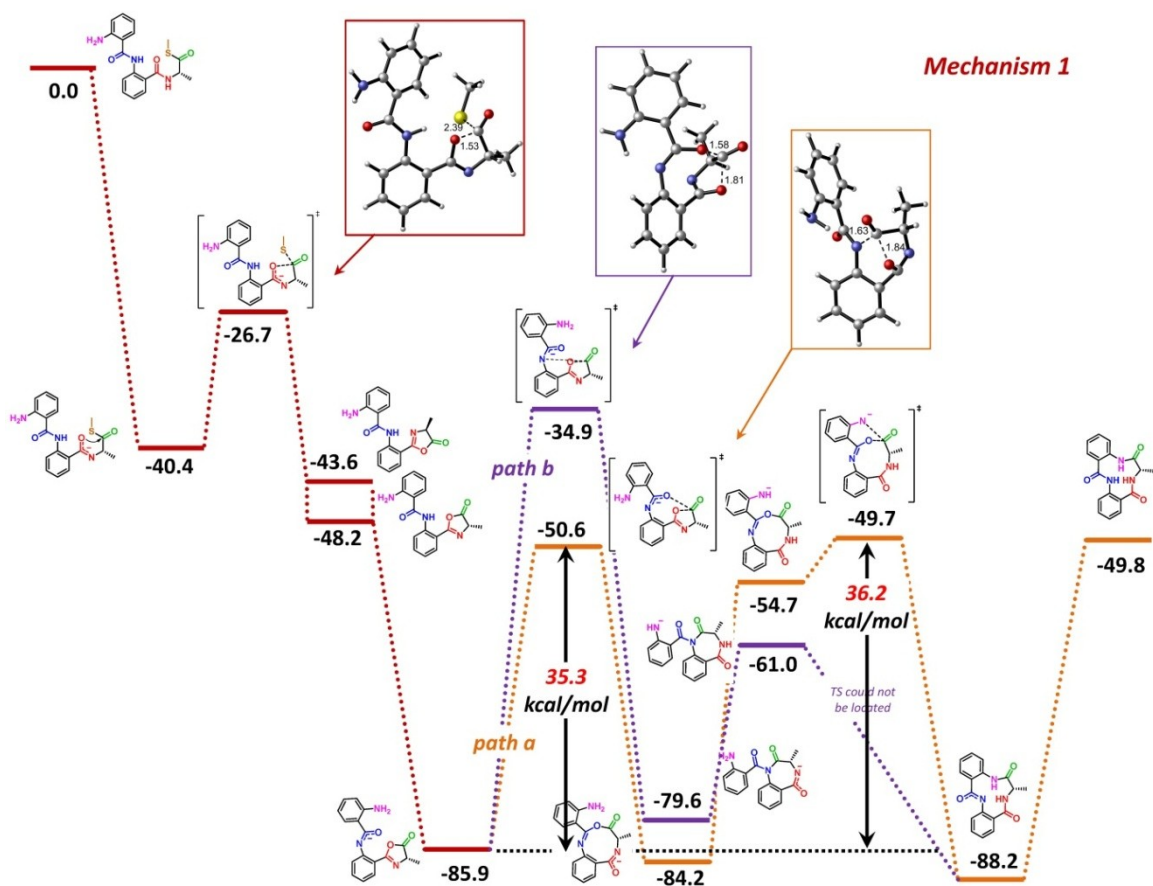
Alternatively in *mechanism 2*, the amide nitrogen bridging the two Ant residues can initiate attack on the thioester to form a 7-membered ring, followed by ring expansion to the 6,11-macrolactam.
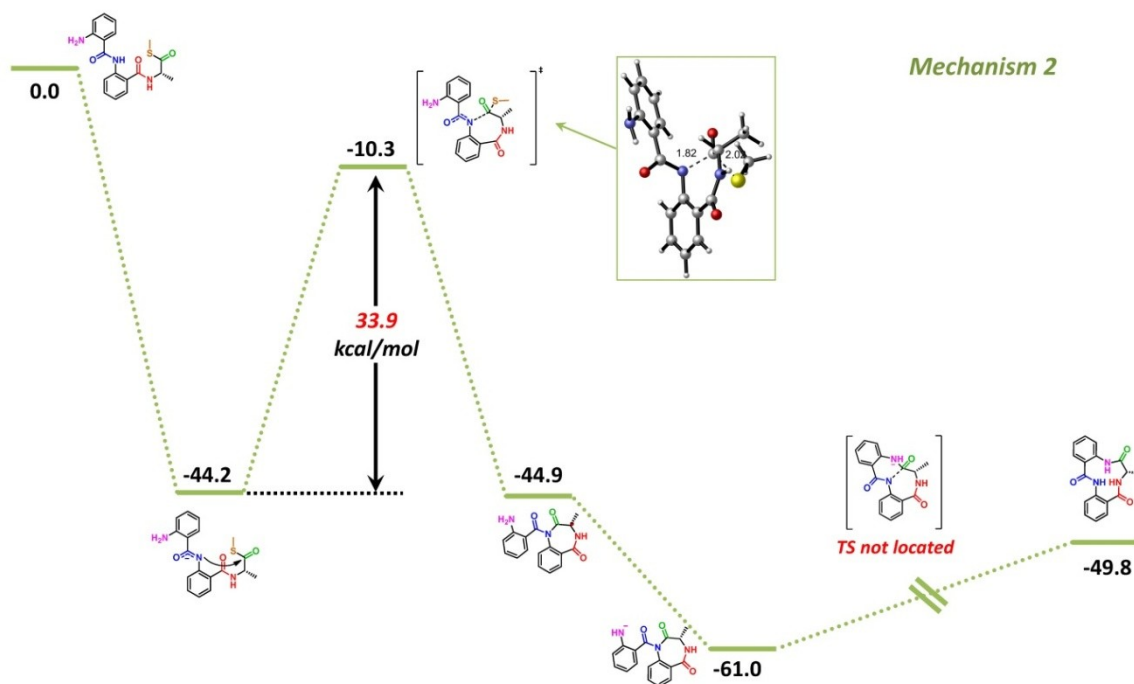
In the originally envisaged *mechanism 3*, the aniline nitrogen serves as the product releasing nucleophile towards the 6,11-macrolactam.
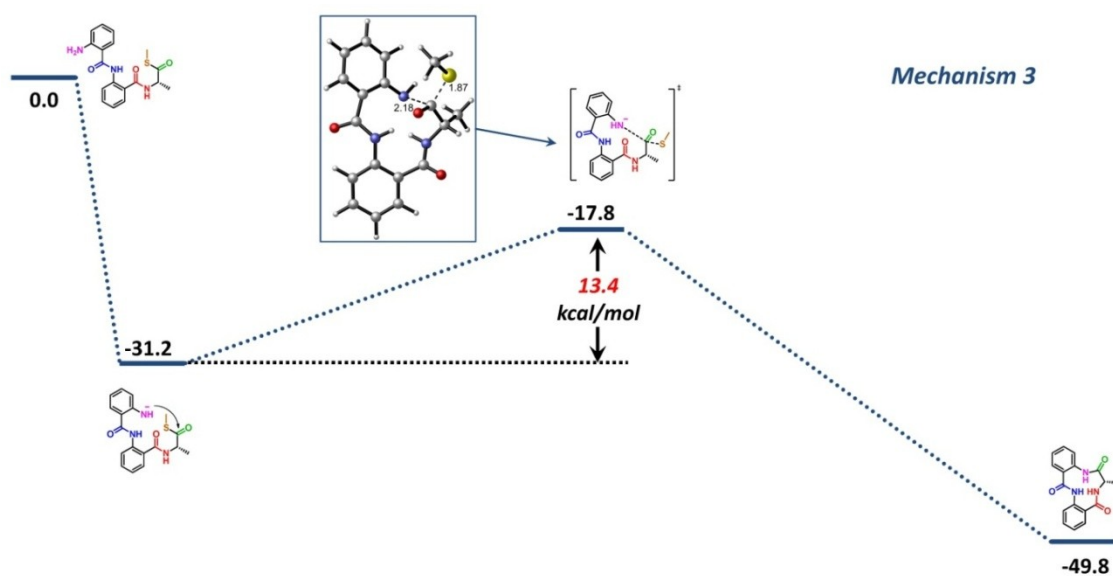
Figure S48. Minimum energy pathways for the nucleophilic addition of neutral aniline to the thioester. All attempts to locate transition structures in these potential energy surfaces were unsuccessful either without explicit solvation (A), in the presence of four water molecules surrounding the NH_2 and SMe (B) or with the assistance of explicit solvation and methylamine as a general base (C). This behavior was observed also when attempting the ring expansion of other intermediate macrocycles by nucleophilic addition of aniline to lactam carbonyls (D). The calculations were performed at the PCM(water)/B3LYP/6-31G(d) level using reduced models in which the indol ring from Trp side chain has been substituted by a hydrogen. Relative free energies are in kcal mol^{-1} and distances in Angstroms.
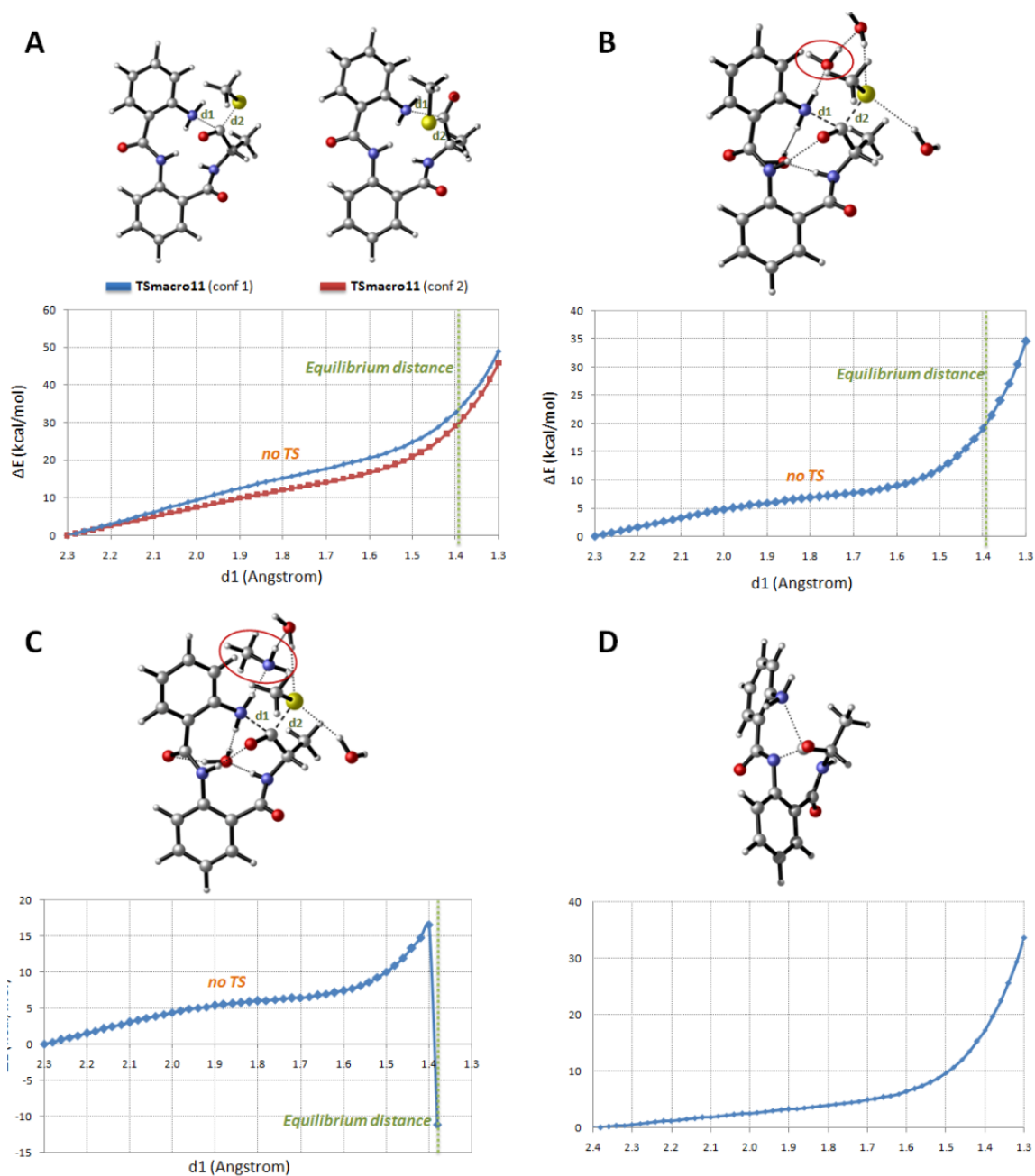
4.4   Supporting Information for Section 2.6

Supporting Information and Methods

Crystallization. LovD proteins were purified as previously described [86] except for selenomethionyl G0, which was purified using minimal medium [243]. The proteins were dialysed overnight into 50 mM Tris pH 8.0, 150 mM NaCl and 5 mM DTT (using Spectra/Por molecular porous membrane tubing MWCO 6~8,000). LovD proteins were concentrated to a desirable concentration (using Amicon Ultra 15 MWCO 30,000).

LovD G0 (7.6 mg/ml). Crystals of G0 were grown at RT by a hanging drop vapor diffusion method using a 2:1 protein to reservoir solution ratio for a total drop size of 3 μl. Diffraction quality crystals were obtained in 3~4 days when using 20% PEG 3350, 0.1M HEPES pH 7.0, 0.25M $(NH_4)_2SO_4$, and 10mM DTT, as a reservoir solution. In preparation for data collection crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen. MJA was not observed to be present in the complex.

LovD G0 selenomethionyl (5 mg/ml). Crystals of LovD selenomethionyl G0 were grown at RT by a hanging drop vapor diffusion method using a 1:2 protein to reservoir solution ratio for a total drop size of 3 μl. Diffraction quality crystals were obtained in 5~6 days when using 22% PEG 3350, 0.1M Bis-Tris pH 6.5, 0.25M $(NH_4)_2SO_4$, and 10mM DTT, as a reservoir solution. In preparation for data collection crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen.

LovD G5 (69 mg/ml). Crystals of G5 were grown at RT by a hanging drop vapor diffusion method using a 2:1 protein to reservoir solution ratio for a total drop size of 3 μl. Diffraction quality crystals were obtained in 3~4 days when using 40% PEG 400, 0.1M HEPES pH 7.5,

0.25M MgCl$_2$, and 10mM DTT as a reservoir solution. In preparation for data collection crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen.

LovD G5+ MJA (50 mg/ml) Crystals of G5 + MJA (100X molar excess of MJA) were grown at RT by a hanging drop vapor diffusion method using a 1:1 protein to reservoir solution ratio for a total drop size of 4 µl. Diffraction quality crystals were obtained in 5~6 days when using 16% PEG 3350, 0.1M magnesium formate, 0.1M Bis-Tris, pH 6.5 and 10mM DTT, as a reservoir solution. In preparation for data collection crystals were briefly soaked in 4M lithium formate and flash frozen.

LovD G5' + MJA (66 mg/ml). Crystals of G5' + MJA (100X molar excess of MJA) were grown at RT by a hanging drop vapor diffusion method using a 1:2 protein to reservoir solution ratio for a total drop size of 3 µl. Diffraction quality crystals were obtained in 4~5 days when using 12.5% PEG 1000, 0.1M calcium acetate, 0.1M imidazole, pH 8.0 and 10mM DTT, as a reservoir solution. In preparation for data collection crystals were briefly soaked in 30%/70% mixture of glycerol/reservoir solution and flash frozen.

LovD G5' + SVA (60 mg/ml). Crystals of G5' were grown at RT by a hanging drop vapor diffusion method using a 1:1 protein to reservoir solution ratio for a total drop size of 4 µl. Diffraction quality crystals were obtained in 3~4 days when using 20% PEG MME 550, 0.1M HEPES pH 8.0, 0.05M MgCl$_2$, and 10mM DTT as a reservoir solution. Stock solutions of 50% Glycerol and 30% PEG MME 550 were mixed in a 70%/30% ratio, without MgCl$_2$. SVA (200mM) was added at equal parts to the mixture above, where crystals were soaked for 10 minutes and flash frozen prior to data collection.

LovD G5' + LVA (60 mg/ml). Crystals of G5' were grown at RT by a hanging drop vapor diffusion method using a 1:1 protein to reservoir solution ratio for a total drop size of 4 μl. Diffraction quality crystals were obtained in 3~4 days when using 15% PEG MME 550, 0.1M HEPES pH 8.0, 0.05M $MgCl_2$, and 10mM DTT, as a reservoir solution. Stock solutions of 50% Glycerol and 30% PEG MME 550 were mixed in a 70%/30% ratio, without $MgCl_2$. LVA (200mM) was added at equal parts to the mixture above, where crystals were soaked for 10 minutes and flash frozen prior to data collection.

Data Collection. X-ray diffraction data were collected at the Advanced Photon Source (APS) beamline 24-ID-C using an ADSC Quantum 315 3X3 CCD array. Crystals were cooled to 100 K in a cryogenic nitrogen stream. Data reduction and scaling were performed using DENZO/SCALEPACK [244]. The LovD G0 data was indexed using XDS [231] as the diffraction pattern showed evidence of non-merohedral or epitaxial twinning, and XDS is superior for indexing such patterns. 5% of all reflections were selected at random from the reciprocal lattice using the CCP4 program FREERFLAG36. The same set of free R flags was maintained throughout the refinement. Care was taken to maintain the free R flags in switching between refinement programs. The data collected on G5'-MJA at the synchrotron was incomplete in the low resolution shell due to overloads on the detector. Data from an in-house source (Rigaku FR-E rotating anode generator equipped with HTC imaging plate) were used to supplement reflections that were overloaded.

Structure determination and refinement. The structure of LovD G0 was determined in space group P1 with molecular replacement using the program BALBES [245] and coordinates of EstB from *B. gladioli* as a search model (PDB code 1CI8) [215]. The two sequences share 32% identity

over 227 residues. Four LovD molecules were found in the asymmetric unit. The first refinement steps were performed with CNS [246], using simulated annealing and conjugate gradient algorithms and the aid of a hydrogen bond potential function [247]. Tight four-fold non-crystallographic symmetry restraints were used throughout. Later rounds of refinement were performed with REFMAC5 [233] to benefit from TLS parameterization of domain disorder [235]. After each refinement step, the model was visually inspected in Coot [248], using both 2Fo-Fc and Fo-Fc difference maps. All hydrogen atoms connected to carbon atoms and backbone nitrogen atoms were included at their geometrically calculated positions and refined using a riding model.

The structure of the selenomethionyl LovD G0 was solved in space group C2 by molecular replacement using the program PHASER [232] and coordinates of LovD G0 as a search model. Again, four LovD molecules were found in the asymmetric unit. The correctness of the molecular replacement solution was verified by the appearance of discrete peaks in the anomalous difference Fourier map over selenomethionine side chains. Both CNS and REFMAC5 were used for refinement. Tight four-fold symmetry restraints were used at first, and then reduced to medium strength at the end of the refinement.

The structure of the LovD G5 mutant was solved in space group $P2_12_12_1$ by molecular replacement using the program PHASER and coordinates of the selenomethionyl LovD as a search model. A single LovD G5 mutant molecule was found in the asymmetric unit. The structure was refined using REFMAC5 and COOT.

The structures of the LovD G5-MJA, G5'-MJA, S5'-SVA, and S5'-LVA were isomorphous with the uncomplexed LovD G5 mutant. Phases were obtained by the difference Fourier method, and refined as described above.

All models were validated with the following structure validation tools: PROCHECK [237], ERRAT [238], and VERIFY3D [239]. The coordinates of the final model and the merged structure factors have been deposited to the Protein Data Bank.

Figure S49. Whole cell activity of LovD G0 and mutants. At 25°C, the G0 produces SVA at a rate of ~1.7 mM per hour. At 32°C, G0, G1, G2.1, G2.2, and G3 did not show any activity. G7 showed a slightly higher activity compared to the WT at 25°C. At 37°C, only G7 retained residual activity of ~0.4 mM SVA per hour.
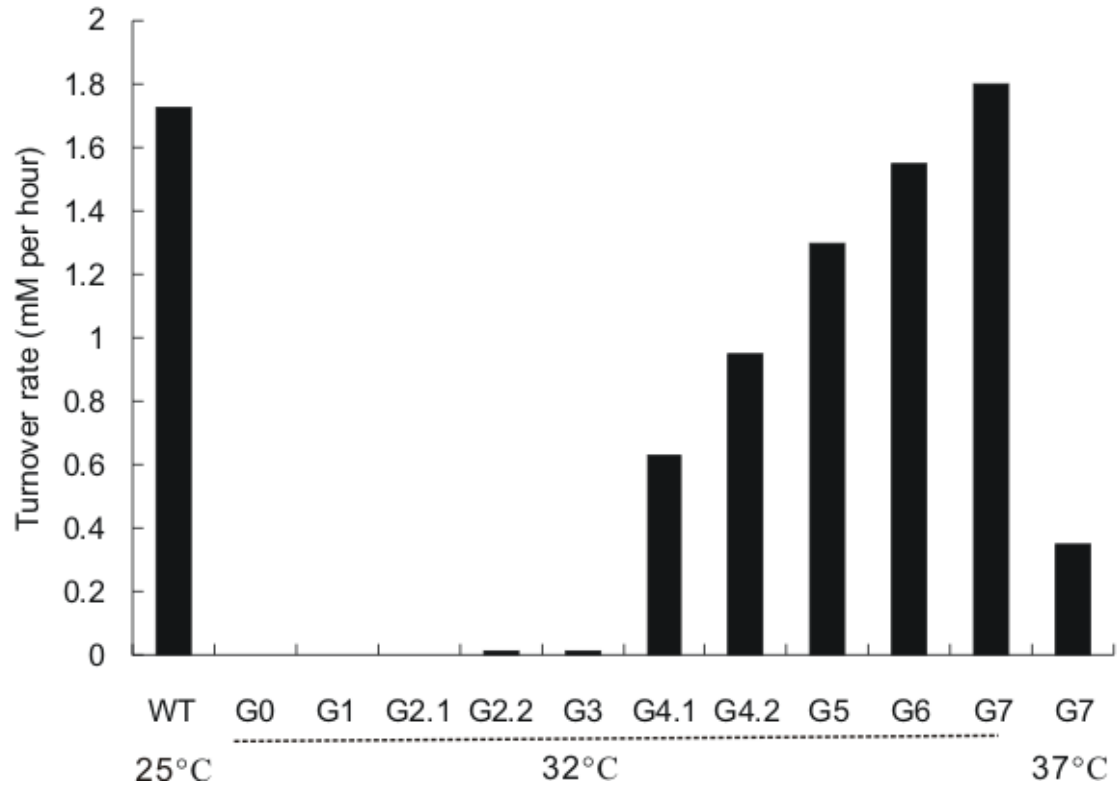
Figure S50. Kinetic characterization of LovD mutants. Turnover rate of LovD mutants toward SVA biosynthesis using DMB-SMMP as substrate (◆ blue line, k_{cat} on left y-axis), LVA biosynthesis using MB-SMMP (■, pink line, k_{cat} on left y-axis) or LovF as substrate (▲, red line, apparent turnover rate on right y-axis).
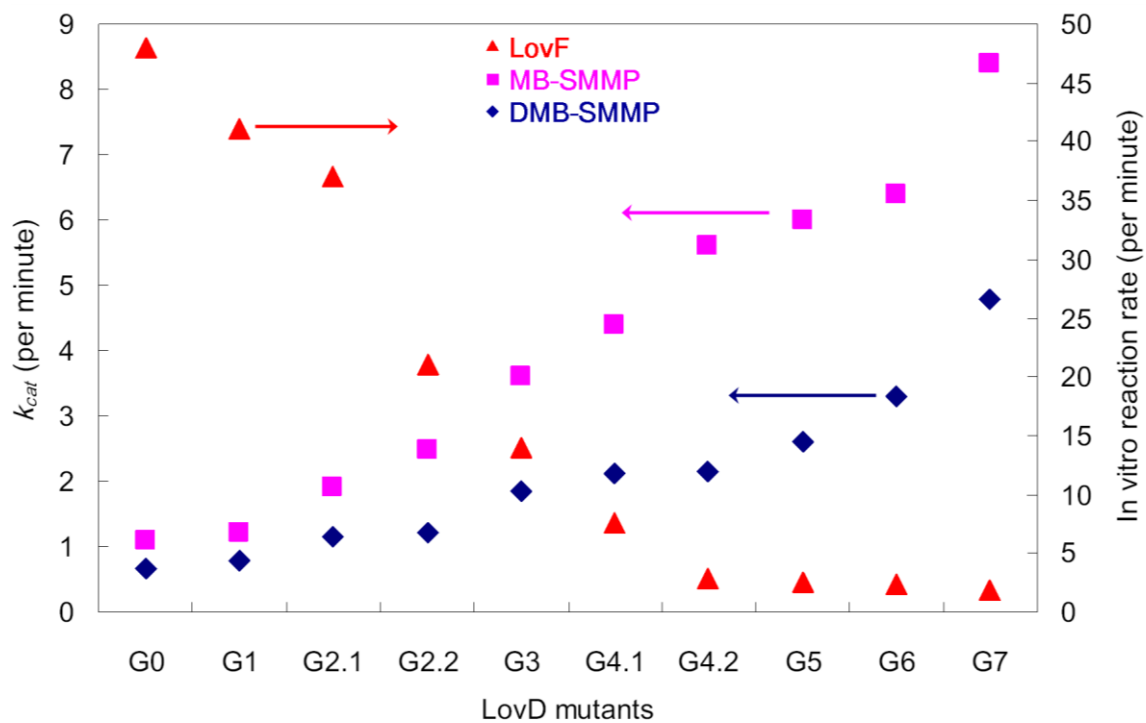
Figure S51. LovD G5' bound to LVA and MJA. (A) Structure of LovD with secondary structure element from N-terminal (blue) to C-terminal (red) and topology diagram of LovD secondary structure (arrows represent β-sheet and rectangles represent α-helix). LVA was shown in grey stick. (B) The substrate binding pockets of LovD and EstB. In LovD, encircling the active site cleft is a broad, ring-shaped ridge, whereas in EstB, the corresponded opening is blocked by an α-helix. (C) Hydrophobic and aromatic amino acids involved in the MJA binding pocket.
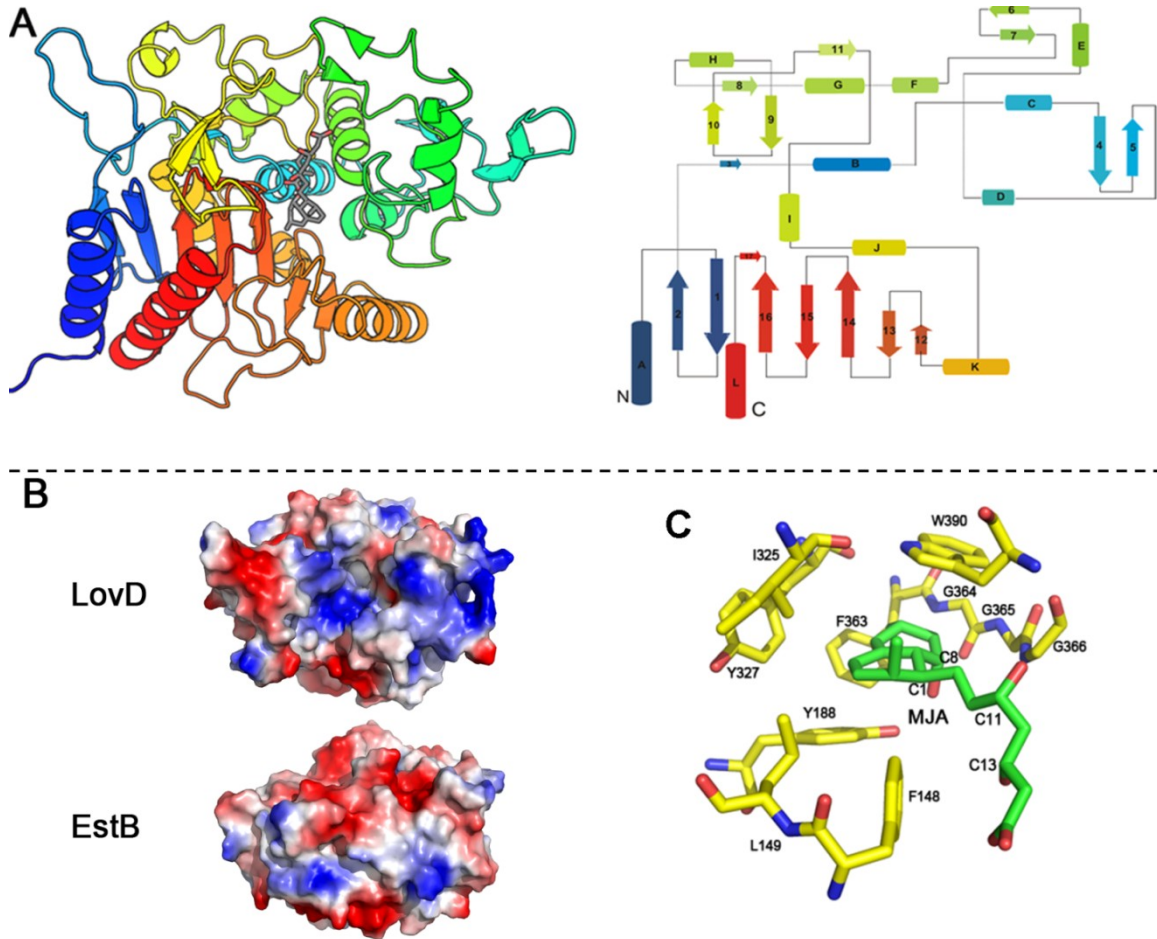
Figure 52. Movement of key residues associated with mutations and upon binding with the ligands (A) The large domains (residues 14-92 and 204-405) of G0-Semet, G5, and G5-MJA are superimposed to indicate the movement of some key residues interacting with MJA. Domain rotation from G0-Semet to G5 closes the gap between the guanido group of Arg173 and the C15 carboxylate of MJA, which is favorable towards binding of MJA. Phe148 and Tyr188 are also closer to MJA in G5. (B) Differences between LVA and SVA binding in the LovD active site. Addition of the extra α-methyl group in SVA produces an opening of the hinge between domains via contact with Phe148.
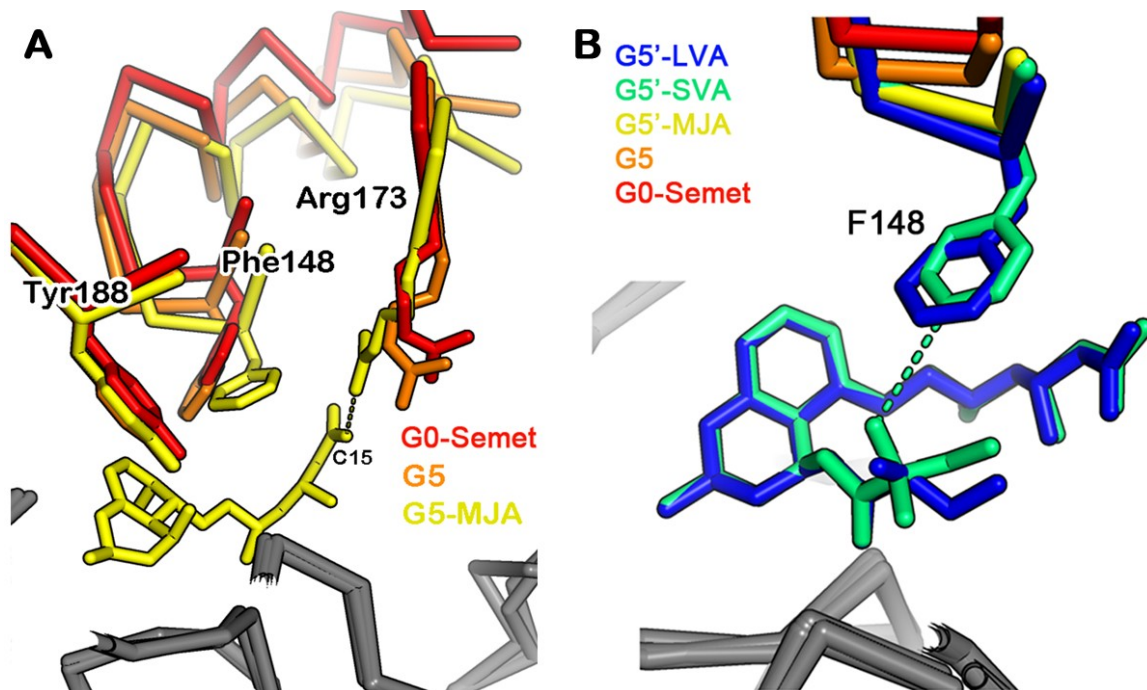
Figure S53. Proposed mechanisms of beneficial mutations in G5. (A) The K26E mutation might improve stability of the enzyme by breaking up a patch of positively charged residues (R22, K23, K26, and R28) on the surface of helix A. The patch is highlighted with a circle in both G0-Semet and G5. (B) The G275S mutation appears to improve stability of the enzyme by adding a hydrogen bond with S278 on helix I.
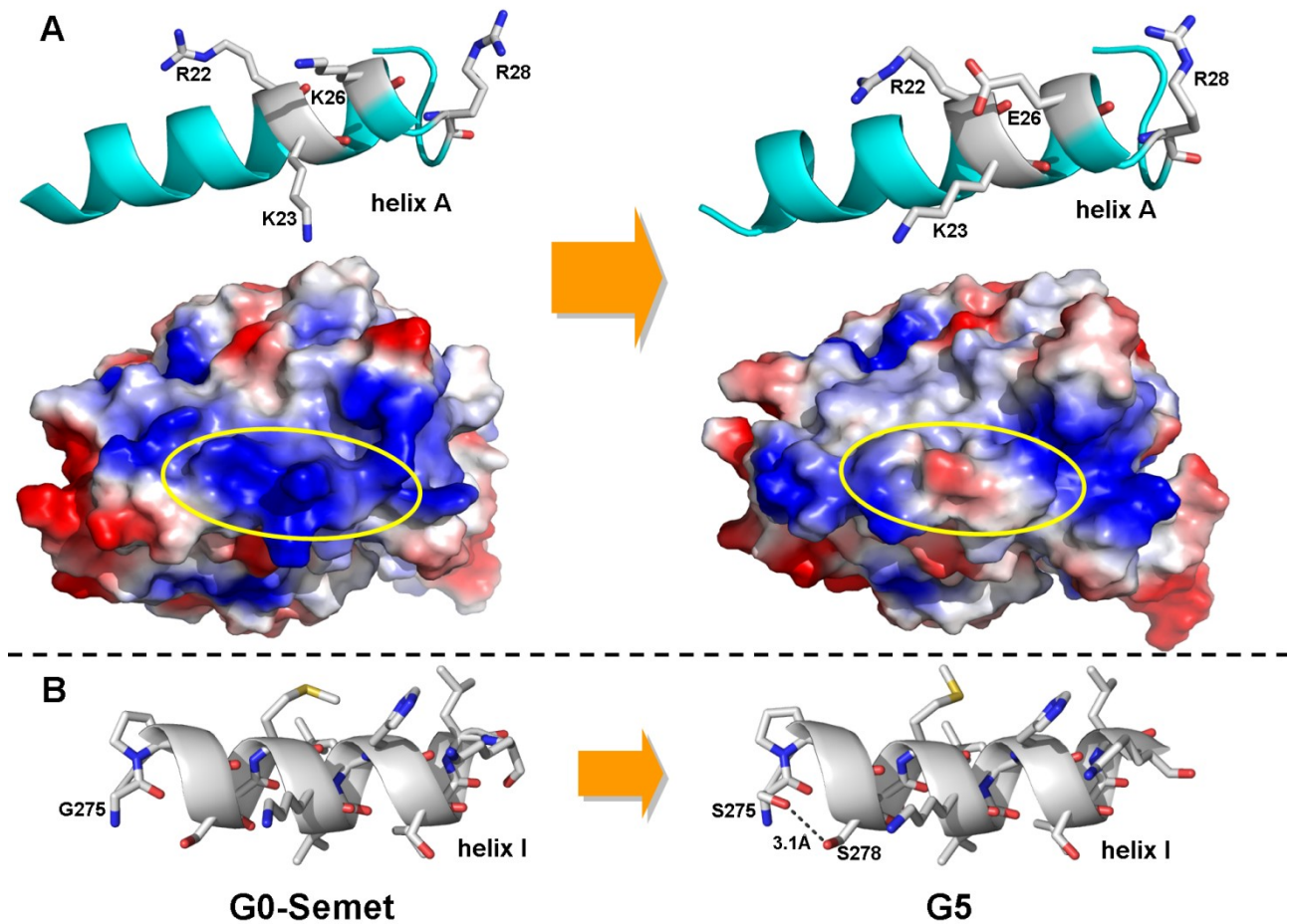
Figure S54. Catalytic residues of LovD and proposed mechanism of LVA biosynthesis. (A) Proposed residues involved in LovD reaction. (B) Superimposition of the active site residues between LovD (gray) and EstB (magenta). (C) LVA biosynthesis using MJA and MB-LovF_ACP as substrate. Ser76 acts as the catalytic nucleophile to catalyze the ester exchange reaction. The hydroxyl groups Ser76 and MJA (C8) are deprotonated by Tyr188, which is stabilized by Lys79.
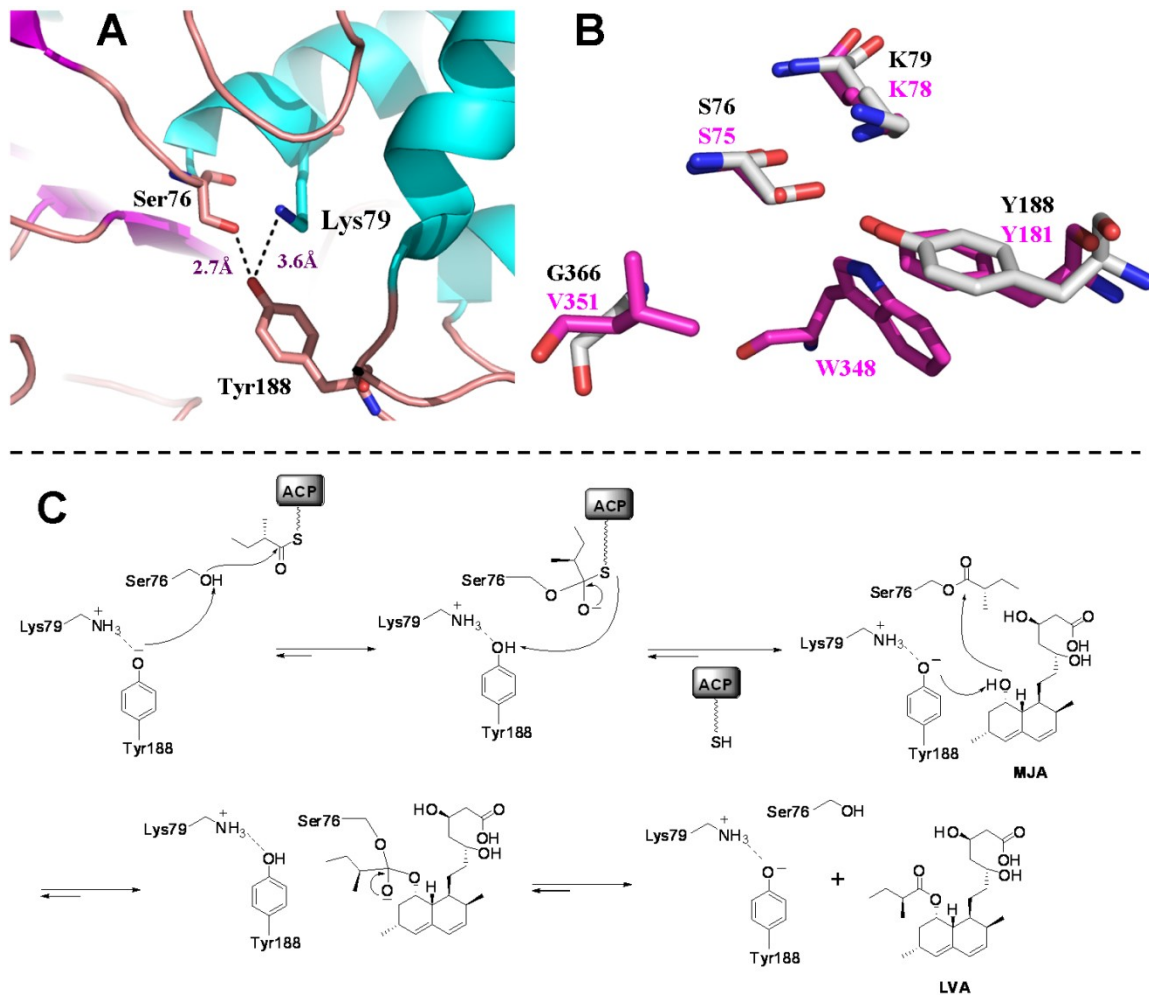
Table S11. Data collection and refinement statistics

	simh9014	simh6208
Data collection		
Space group	P1	P2$_1$
Cell dimensions		
a, b, c (Å)	56.3, 83.9, 99.2	43.8, 76.4, 113.8
α, β, γ (°)	102.6, 89.0, 107.6	90.0, 93.9, 90.0
Resolution (Å)	3.20 (3.27-3.20)	1.80 (1.85-1.80)
R_{merge}	0.175 (0.679)	0.045 (0.469)
$I/\sigma I$	4.6 (1.0)	21.0 (3.2)
CC$_{1/2}$	96.0 (51.8)	99.9 (83.1)
Completeness (%)	92.1 (89.2)	99.5 (99.5)
Redundancy	21.4 (1.8)	3.7 (3.6)
Refinement		
Resolution (Å)	3.2	1.8
No. reflections	25938	68966
R_{work}/ R_{free}	0.208/0.252	0.167/0.191
No. atoms		
Protein	12344	6395
Ligand/ion	0	391
Water	12	383
B-factors (Å^2)		
Protein	50.9	27.8
Ligand/ion	n/a	39.1
Water	3.4	31.8
R.m.s deviations		
Bond lengths (Å)	0.010	0.010
Bond angles (°)	1.3	1.0

*Highest resolution shell is shown in parenthesis.

Figure S55. Gel filtration chromatography for LovD proteins. Protein was purified from NTA agarose affinity column and was loaded onto a Superdex 200 (GE Healthcare) column and eluted with a flow rate of 0.5 ml min^{-1} in 50 mM Tris pH 8.0 and 2 mM DTT.
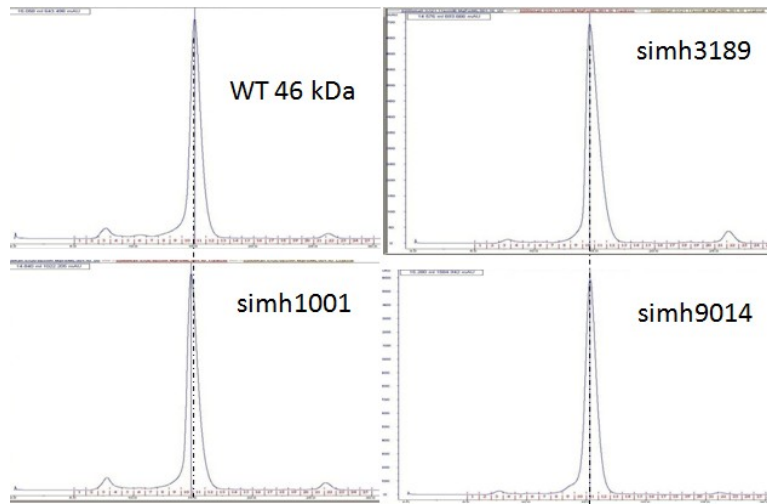
Figure S56. A stereo ribbon diagram showing the overall structures of LovD variant simh9014 and 6208. The magentas sticks showed the mutations introduced into the structures. Simh9014 was in cyans and 6208 was in green color.
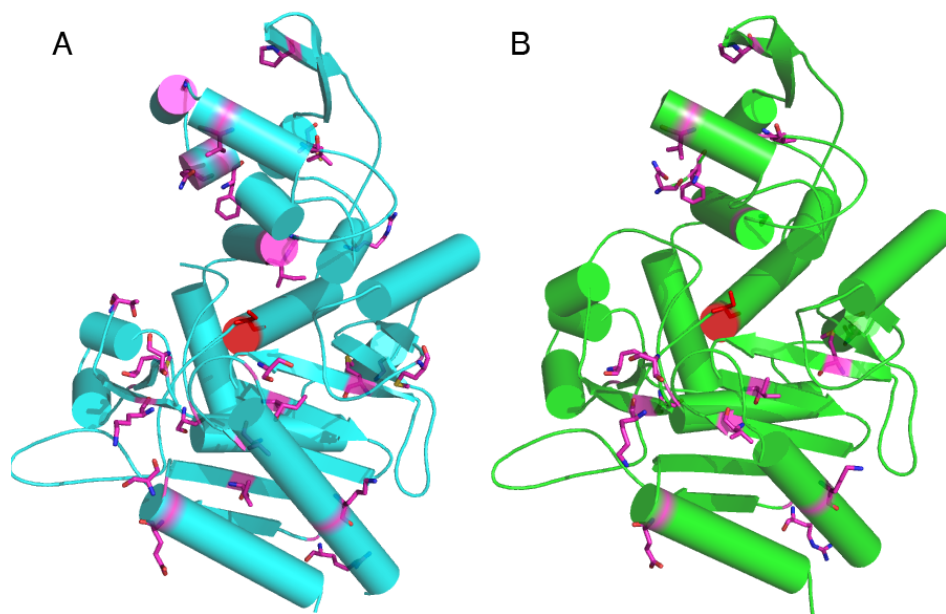
5   References

1.   Brase, S., et al., *Chemistry and Biology of Mycotoxins and Related Fungal Metabolites.* Chemical Reviews, 2009. **109**(9): p. 3903-3990.
2.   Keller, N.P., G. Turner, and J.W. Bennett, *Fungal secondary metabolism - from biochemistry to genomics.* Nat. Rev. Microbiol., 2005. **3**(12): p. 937-47.
3.   Bok, J.W., et al., *Genomic mining for Aspergillus natural products.* Chemistry & Biology, 2006. **13**(1): p. 31-37.
4.   Hoffmeister, D. and N.P. Keller, *Natural products of filamentous fungi: enzymes, genes, and their regulation.* Nat. Prod. Rep., 2007. **24**(2): p. 393-416.
5.   Keller, N.P. and T.M. Hohn, *Metabolic pathway gene clusters in filamentous fungi.* Fungal Genet. Biol., 1997. **21**(1): p. 17-29.
6.   Walton, J.D., *Horizontal gene transfer and the evolution of secondary metabolite gene clusters in fungi: An hypothesis.* Fungal Genet. Biol., 2000. **30**(3): p. 167-171.
7.   Sanchez, J.F., et al., *Advances in Aspergillus secondary metabolite research in the post-genomic era.* Nat. Prod. Rep., 2012. **29**(3): p. 351-371.
8.   van den Berg, M.A., et al., *Genome sequencing and analysis of the filamentous fungus Penicillium chrysogenum.* Nat. Biotechnol., 2008. **26**(18820685): p. 1161-1168.
9.   Hoffmeister, D. and N.P. Keller, *Natural products of filamentous fungi: enzymes, genes, and their regulation.* Natural Product Reports, 2007. **24**(2): p. 393-416.
10.   Sanchez, J.F., et al., *Advances in Aspergillus secondary metabolite research in the post-genomic era.* Natural Product Reports, 2012. **29**(3): p. 351-371.
11.   Clardy, J. and C. Walsh, *Lessons from natural molecules.* Nature, 2004. **432**(7019): p. 829-837.
12.   Byford, M.F., et al., *The mechanism of ACV synthetase.* Chemical Reviews, 1997. **97**(7): p. 2631-2649.
13.   Bierer, B.E., et al., *Cyclosporine-A and fk506 - molecular mechanisms of immunosuppression and probes for transplantation biology.* Curr. Opin. Immunol., 1993. **5**(5): p. 763-773.
14.   Alberts, A.W., et al., *Mevinolin - a highly potent competitive inhibitor of hydroxymethylglutaryl-coenzyme-a reductase and a cholesterol-lowering agent.* Proc. Nat. Acad. Sci. USA, 1980. **77**(7): p. 3957-3961.
15.   Eaton, D.L. and E.P. Gallagher, *Mechanism of aflatoxin carcinogenesis.* Annu. Rev. Pharmacol. Toxicol., 1994. **34**: p. 135-172.
16.   Gardiner, D.M., P. Waring, and B.J. Howlett, *The epipolythiodioxopiperazine (ETP) class of fungal toxins: distribution, mode of action, functions and biosynthesis.* Microbiology-Sgm, 2005. **151**: p. 1021-1032.
17.   Hanson, J.R., *The Chemistry of Fungi.* 2008, Cambridge, UK: RSC Publishing.
18.   Weber, G. and E. Leitner, *Disruption of the cyclosporin synthetase gene of Tolypocladium niveum.* Curr. Genet., 1994. **26**(7874740): p. 461-467.

19. Cacho, R.A., et al., *Identification and Characterization of the Echinocandin B Biosynthetic Gene Cluster from Emericella rugulosa NRRL 11440.* J. Am. Chem. Soc., 2012. **134**(40): p. 16781-16790.
20. Chen, S.C.A., M.A. Slavin, and T.C. Sorrell, *Echinocandin antifungal drugs in fungal infections: a comparison.* Drugs, 2011. **71**(21175238): p. 11-41.
21. Schwarzer, D., R. Finking, and M.A. Marahiel, *Nonribosomal peptides: from genes to products.* Nat. Prod. Rep., 2003. **20**(3): p. 275-287.
22. Sattely, E.S., M.A. Fischbach, and C.T. Walsh, *Total biosynthesis: in vitro reconstitution of polyketide and nonribosomal peptide pathways.* Nat. Prod. Rep., 2008. **25**(4): p. 757-793.
23. Wintermeyer, W. and M.V. Rodnina, *Biopolymers. Vol. 7, Polyamides and complex proteinaceous materials I; Ribosomal protein synthesis.* 2002, Weinheim; Cambridge: Wiley-VCH.
24. Dittmann, J., et al., *Mechanism of cyclosporin A biosynthesis. Evidence for synthesis via a single linear undecapeptide precursor.* J. Biol. Chem., 1994. **269**(8300618): p. 2841-2846.
25. Lawen, A. and R. Zocher, *Cyclosporin synthetase. The most complex peptide synthesizing multienzyme polypeptide so far described.* J. Biol. Chem., 1990. **265**(19): p. 11355-60.
26. Yin, W.-B., et al., *Acetylaszonalenin biosynthesis in Neosartorya fischeri. Identification of the biosynthetic gene cluster by genomic mining and functional proof of the genes by biochemical investigation.* J. Biol. Chem., 2009. **284**(19001367): p. 100-109.
27. Maiya, S., et al., *The fumitremorgin gene cluster of Aspergillus fumigatus: identification of a gene encoding brevianamide F synthetase.* Chembiochem, 2006. **7**(16755625): p. 1062-1069.
28. Grundmann, A. and S.-M. Li, *Overproduction, purification and characterization of FtmPT1, a brevianamide F prenyltransferase from Aspergillus fumigatus.* Microbiology, 2005. **151**(16000710): p. 2199-2207.
29. Hayashi, H., et al., *Structure and insecticidal activity of new indole alkaloids, okaramines A and B, from Penicillium simplicissimum.* Agric. Food Chem., 1989. **53**: p. 461-469.
30. Hijarrubia, M.J., et al., *Characterization of the lys2 gene of Acremonium chrysogenum encoding a functional alpha-aminoadipate activating and reducing enzyme.* Mol. Gen. Genet., 2001. **264**(11254122): p. 755-762.
31. Walsh, C.T., S.W. Haynes, and B.D. Ames, *Aminobenzoates as building blocks for natural product assembly lines.* Nat. Prod. Rep., 2012. **29**(22064917): p. 37-59.
32. Dunn, M.F., et al., *Tryptophan synthase: the workings of a channeling nanomachine.* Trends Biochem. Sci., 2008. **33**(18486479): p. 254-264.
33. McMurry, J. and T. Begley, *The organic chemistry of biological pathways.* 1995: Roberts & Company.
34. Fedorova, N.D., et al., *Genomic islands in the pathogenic filamentous fungus Aspergillus fumigatus.* PLoS Genet., 2008. **4**(18404212).

35. Wollinsky, B., et al., *Breaking the regioselectivity of indole prenyltransferases: identification of regular C3-prenylated hexahydropyrrolo[2,3-b]indoles as side products of the regular C2-prenyltransferase FtmPT1*. Org. Biomol. Chem., 2012. **10**(23090579): p. 9262-9270.

36. Yu, X., et al., *Biochemical characterization of indole prenyltransferases: filling the last gap of prenylation positions by a 5-dimethylallyltryptophan synthase from Aspergillus clavatus*. J. Biol. Chem., 2012. **287**(22123822): p. 1371-1380.

37. Zou, H.-X., et al., *Simultaneous C7- and N1-prenylation of cyclo-L-Trp-L-Trp catalyzed by a prenyltransferase from Aspergillus oryzae*. Org. Biomol. Chem., 2010. **8**(20473424): p. 3037-3044.

38. Ames, B.D. and C.T. Walsh, *Anthranilate-activating modules from fungal nonribosomal peptide assembly lines*. Biochemistry, 2010. **49**(15): p. 3351-3365.

39. Sieber, S.A. and M.A. Marahiel, *Molecular mechanisms underlying nonribosomal peptide synthesis: Approaches to new antibiotics*. Chem. Rev., 2005. **105**(2): p. 715-738.

40. Fischbach, M.A. and C.T. Walsh, *Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms*. Chem. Rev., 2006. **106**(8): p. 3468-3496.

41. Kohli, R.M., C.T. Walsh, and M.D. Burkart, *Biomimetic synthesis and optimization of cyclic peptide antibiotics*. Nature, 2002. **418**(6898): p. 658-661.

42. Kopp, F. and M.A. Marahiel, *Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis*. Nat. Prod. Rep., 2007. **24**(4): p. 735-749.

43. Eys, S., et al., *Three thioesterases are involved in the biosynthesis of phosphinothricin tripeptide in Streptomyces viridochromogenes Tu494*. Antimicrob. Agents 2008. **52**(5): p. 1686-1696.

44. Du, L.C. and L.L. Lou, *PKS and NRPS release mechanisms*. Nat. Prod. Rep., 2010. **27**(2): p. 255-278.

45. Bruner, S.D., et al., *Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE*. Structure, 2002. **10**(3): p. 301-310.

46. Samel, S.A., et al., *The thioesterase domain of the fengycin biosynthesis cluster: A structural base for the macrocyclization of a non-ribosomal lipopeptide*. J. Mol. Biol., 2006. **359**(4): p. 876-889.

47. Robbel, L., K.M. Hoyer, and M.A. Marahiel, *TioS T-TE - a prototypical thioesterase responsible for cyclodimerization of the quinoline- and quinoxaline-type class of chromodepsipeptides*. Febs J., 2009. **276**(6): p. 1641-1653.

48. Hoyer, K.M., C. Mahlert, and M.A. Marahiel, *The iterative gramicidin S thioesterase catalyzes peptide ligation and cyclization*. Chem. Biol., 2007. **14**(1): p. 13-22.

49. Kopp, F. and M.A. Marahiel, *Macrocyclization strategies in polyketide and nonribosomal peptide biosynthesis*. Natural Product Reports, 2007. **24**(4): p. 735-749.

50. Eisfeld, K., *Non-ribosomal peptide synthetases of fungi*. Physiology and Genetics, 2009: p. Springer Berlin: Heidelberg, Berlin.

51.	Daniel, J.F.D. and E. Rodrigues, *Peptaibols of Trichoderma.* Nat. Prod. Rep., 2007. **24**(5): p. 1128-1141.

52.	Liu, X.Y. and C.T. Walsh, *Cyclopiazonic Acid Biosynthesis in Aspergillus sp.: Characterization of a Reductase-like R* Domain in Cyclopiazonate Synthetase That Forms and Releases cyclo-Acetoacetyl-L-tryptophan.* Biochemistry, 2009. **48**(36): p. 8746-8757.

53.	Sims, J.W. and E.W. Schmidt, *Thioesterase-like role for fungal PKS-NRPS hybrid reductive domains.* J. Am. Chem. Soc., 2008. **130**(33): p. 11149-11155.

54.	Wilkinson, B. and J. Micklefield, *Chapter 14. Biosynthesis of Nonribosomal Peptide Precursors*, in *Methods enzymol; Complex enzymes in microbial natural product biosynthesis, Part A: overview articles and peptides*, D.A. Hopwood, Editor. 2009, Academic Press. p. 353-378.

55.	Olano, C., C. Mendez, and J.A. Salas, *Post-PKS tailoring steps in natural product-producing actinomycetes from the perspective of combinatorial biosynthesis.* Nat. Prod. Rep., 2010. **27**(4): p. 571-616.

56.	Pohlmann, V. and M.A. Marahiel, *Delta-amino group hydroxylation of L-ornithine during coelichelin biosynthesis.* Org. Biomol. Chem., 2008. **6**(10): p. 1843-8.

57.	Hollenhorst, M.A., et al., *The nonribosomal peptide synthetase enzyme DdaD tethers N(β)-fumaramoyl-l-2,3-diaminopropionate for Fe(II)/α-ketoglutarate-dependent epoxidation by DdaC during dapdiamide antibiotic biosynthesis.* J. Am. Chem. Soc., 2011. **133**(5): p. 1609-1609.

58.	Robbel, L., et al., *Consecutive enzymatic modification of ornithine generates the hydroxamate moieties of the siderophore erythrochelin.* Biochemistry, 2011. **50**(27): p. 6073-6080.

59.	Giessen, T.W., F.I. Kraas, and M.A. Marahiel, *A four-enzyme pathway for 3,5-dihydroxy-4-methylanthranilic acid formation and incorporation into the antitumor antibiotic sibiromycin.* Biochemistry, 2011. **50**(25): p. 5680-5692.

60.	Tang, M.C., C.Y. Fu, and G.L. Tang, *Characterization of SfmD as a heme peroxidase that catalyzes the regioselective hydroxylation of 3-methyltyrosine to 3-hydroxy-5-methyltyrosine in saframycin a biosynthesis.* J. Biol. Chem., 2012. **287**(7): p. 5112-21.

61.	Lin, S., S.G. Van Lanen, and B. Shen, *Characterization of the two-component, FAD-dependent monooxygenase SgcC that requires carrier protein-tethered substrates for the biosynthesis of the enediyne antitumor antibiotic C-1027.* J. Am. Chem. Soc., 2008. **130**(20): p. 6616-6623.

62.	Zhang, W.J., B.D. Ames, and C.T. Walsh, *Identification of phenylalanine 3-hydroxylase for meta-tyrosine biosynthesis.* Biochemistry, 2011. **50**(24): p. 5401-5403.

63.	Gruschow, S., E.J. Rackham, and R.J.M. Goss, *Diversity in natural product families is governed by more than enzyme promiscuity alone: Establishing control of the pacidamycin portfolio.* Chem. Sci., 2011. **2**(11): p. 2182-2186.

64. Singh, G.M., et al., *beta-Hydroxylation of the aspartyl residue in the phytotoxin syringomycin E: Characterization of two candidate hydroxylases AspH and SyrP in Pseudomonas syringae.* Biochemistry, 2008. **47**(43): p. 11310-11320.

65. Strieker, M., et al., *Stereospecific synthesis of threo- and erythro-beta-hydroxyglutamic acid during kutzneride biosynthesis.* J. Am. Chem. Soc., 2009. **131**(37): p. 13523-13530.

66. Strieker, M., et al., *Mechanistic and structural basis of stereospecific C beta-hydroxylation in calcium-dependent antibiotic, a daptomycin-type lipopeptide.* ACS Chem. Biol., 2007. **2**(3): p. 187-196.

67. Strieker, M., et al., *Non-heme hydroxylase engineering for simple enzymatic synthesis of L-threo-hydroxyaspartic acid.* Chembiochem, 2008. **9**(3): p. 374-376.

68. Li, S., et al., *Biochemical characterization of NotB as an FAD-dependent oxidase in the biosynthesis of notoamide indole alkaloids.* J. Am. Chem. Soc., 2012. **134**(2): p. 788-91.

69. Koketsu, K., et al., *Reconstruction of the saframycin core scaffold defines dual Pictet-Spengler mechanisms.* Nat. Chem. Biol., 2010. **6**(6): p. 408-10.

70. Wilkinson, B. and B.O. Bachmann, *Biocatalysis in pharmaceutical preparation and alteration.* Curr. Opin. Chem. Biol., 2006. **10**(2): p. 169-176.

71. Castle, L.A., et al., *Discovery and directed evolution of a glyphosate tolerance gene.* Science, 2004. **304**(5674): p. 1151-1154.

72. Zhao, H.M., K. Chockalingam, and Z.L. Chen, *Directed evolution of enzymes and pathways for industrial biocatalysis.* Curr. Opin. in Biotechnol., 2002. **13**(2): p. 104-110.

73. Baltz, R.H., *Molecular engineering approaches to peptide, polyketide and other antibiotics.* Nat. Biotechnol., 2006. **24**(12): p. 1533-1540.

74. Olano, C., C. Mendez, and J.A. Salas, *Antitumor compounds from actinomycetes: from gene clusters to new derivatives by combinatorial biosynthesis.* Nat. Prod. Rep., 2009. **26**(5): p. 628-660.

75. Weissman, K.J. and P.F. Leadlay, *Combinatorial biosynthesis of reduced polyketides.* Nat. Rev. Microbiol., 2005. **3**(12): p. 925-936.

76. Loncaric, C., E. Merriweather, and K.D. Walker, *Profiling a taxol pathway 10 beta-acetyltransferase: Assessment of the specificity and the production of baccatin III by in vivo acetylation in E. coli.* Chem. Biol., 2006. **13**: p. 309-317.

77. Zhang, C.S., et al., *Exploiting the reversibility of natural product glycosyltransferase-catalyzed reactions.* Science, 2006. **313**: p. 1291-1294.

78. Rix, U., et al., *Modification of post-PKS tailoring steps through combinatorial biosynthesis.* Nat. Prod. Rep., 2002. **19**(5): p. 542-80.

79. Neumann, C.S., D.G. Fujimori, and C.T. Walsh, *Halogenation strategies in natural product biosynthesis.* Chem. Biol., 2008. **15**: p. 99-109.

80. Zhou, H., X. Xie, and Y. Tang, *Engineering natural products using combinatorial biosynthesis and biocatalysis.* Curr. Opin. Biotechnol., 2008. **19**: p. 590-596.

81. Freeman, M.W., *Statins, cholesterol, and the prevention of coronary heart disease.* Faseb J., 2006. **20**(2): p. 200-201.

82. Hendrickson, L., et al., *Lavastatin biosynthesis in Aspergillus terreus: characterization of blocked mutants, enzyme activities and a multifunctional polyketide synthase gene.* Chem Biol, 1999. **6**(7): p. 429-439.

83. Kennedy, J., et al., *Modulation of polyketide synthase activity by accessory proteins during lovastatin biosynthesis.* Science, 1999. **284**(5418): p. 1368-72.

84. Manzoni, M. and N. Rollini, *Biosynthesis and biotechnological production of statins by filamentous fungi and application of these cholesterol-lowering drugs.* Appl. Environ. Microbiol., 2002. **58**(5): p. 555-564.

85. Xie, X., Meehan, MJ., Xu, W., Dorrecstein, PC., Tang, Y., *Acyltransferase mediated polyketide release from a fungal megasynthase.* J. Am. Chem. Soc., 2009. **131**: p. 8388-8389.

86. Xie, X., et al., *Biosynthesis of lovastatin analogs with a broadly specific acyltransferase.* Chem. Biol., 2006. **13**: p. 1161-9.

87. Xie, X. and Y. Tang, *Efficient synthesis of simvastatin by use of whole-cell biocatalysis.* Appl. Environ. Microbiol., 2007. **73**: p. 2054-2060.

88. Xie, X., W.W. Wong, and Y. Tang, *Improving simvastatin bioconversion in Escherichia coli by deletion of bioH.* Metab. Eng., 2007. **9**: p. 379-86.

89. Sanishvili, R., et al., *Integrating structure, bioinformatics, and enzymology to discover function: BioH, a new carboxylesterase from Escherichia coli.* J. Biol. Chem., 2003. **278**(28): p. 26039-45.

90. Cole, R.J., *Fungal tremorgens.* J. Food Prot., 1981. **44**(9): p. 715-722.

91. Evans, T. and R. Gupta, *Tremorgenic mycotoxins*, in *Veterinary toxicology: basic and clinical principles*, R. Gupta, Editor. 2007, Academic Press: New York. p. 1004-1010.

92. D' Yakonov, A. and M. Telezhenetskaya, *Quinazoline alkaloids in nature.* Chem. Nat. Compd., 1997. **33**: p. 221-267.

93. Steyn, P. and V. R., *Tremorgenic mycotoxins.* Fortschr. Chem. Org. Naturst., 1985. **48**: p. 1-80.

94. Selala, M.I., F. Daelemans, and P.J.C. Schepens, *Fungal tremorgens - the mechanism of action of single nitrogen containing toxins - a hypothesis.* Drug Chem. Toxicol., 1989. **12**(3-4): p. 237-257.

95. Yao, Y., et al., *The tremorigen aflatrem is a positive allosteric modulator of the gamma-aminobutyric acida receptor channel expressed in xenopus-oocytes.* Mol. Pharmacol., 1989. **35**(3): p. 319-323.

96. Rabindran, S.K., et al., *Fumitremorgin C reverses multidrug resistance in cells transfected with the breast cancer resistance protein.* Anticancer Res., 2000. **60**(1): p. 47-50.

97. Dowd, P.F., R.J. Cole, and R.F. Vesonder, *Toxicity of selected tremorgenic mycotoxins and related-compounds to spodoptera-frugiperda and heliothis-zea.* J. Antibiot., 1988. **41**(12): p. 1868-1872.

98. Hayashi, H., *Fungal metabolites with bioactivity to insects.* Recent Res. Dev. Agric. Biol. Chem., 1998. **2**: p. 511-525.

99. Hayashi, H., et al., *Penitrem A, as a convulsive factor against silkworm, from Penicillium simplicissimum AK-40.* Chem. Express., 1993. **8**(3): p. 177-80.

100. Ding, Y.S., et al., *Genome-based characterization of two prenylation steps in the assembly of the Stephacidin and Notoamide anticancer agents in a marine-derived Aspergillus sp.* J. Am. Chem. Soc., 2010. **132**(36): p. 12733-12740.

101. Kato, N., et al., *Identification of cytochrome P450s required for fumitremorgin biosynthesis in Aspergillus fumigatus.* Chembiochem, 2009. **10**(5): p. 920-928.

102. Saikia, S., et al., *The genetic basis for indole-diterpene chemical diversity in filamentous fungi.* Mycol. Res., 2008. **112**: p. 184-199.

103. Steffan, N., et al., *Indole prenyltransferases from fungi: A new enzyme group with high potential for the production of prenylated indole derivatives.* Curr. Med. Chem., 2009. **16**(2): p. 218-231.

104. Ariza, M.R., et al., *Penicillium digitatum metabolites on synthetic media and citrus fruits.* J. Agric. Food. Chem., 2002. **50**(22): p. 6361-6365.

105. Clardy, J., et al., *tryptoquivaline and tryptoquivalone, 2 tremorgenic metabolites of aspergillus-clavatus.* J. Am. Chem. Soc., 1975. **97**(3): p. 663-665.

106. Frisvad, J.C. and R.A. Samson, *Polyphasic taxonomy of Penicillium subgenus Penicillium - A guide to identification of food and air-borne terverticillate Penicillia and their mycotoxins.* Studies in Mycol., 2004(49): p. 1-173.

107. Takahashi, C., et al., *Fumiquinazolines a-g, novel metabolites of a fungus separated from a pseudolabrus marine fish.* J. Chem. Soc., Perkin Trans. 1, 1995(18): p. 2345-2353.

108. Wong, S.M., et al., *Fiscalins: new substance P inhibitors produced by the fungus Neosartorya fischeri. Taxonomy, fermentation, structures, and biological properties.* J. Antibiot. (Tokyo), 1993. **46**(7684734): p. 545-553.

109. Finking, R. and M.A. Marahiel, *Biosynthesis of nonribosomal peptides.* Annu. Rev. Microbiol., 2004. **58**: p. 453-488.

110. Stack, D., C. Neville, and S. Doyle, *Nonribosomal peptide synthesis in Aspergillus fumigatus and other fungi.* Microbiology-Sgm, 2007. **153**: p. 1297-1306.

111. Ames, B.D., X.Y. Liu, and C.T. Walsh, *Enzymatic processing of Fumiquinazoline F: A tandem oxidative-acylation strategy for the generation of multicyclic scaffolds in fungal indole alkaloid biosynthesis.* Biochemistry, 2010. **49**(39): p. 8564-8576.

112. Jiao, R.H., et al., *Chaetominine, a cytotoxic alkaloid produced by endophytic Chaetomium sp IFB-E015.* Org. Lett., 2006. **8**(25): p. 5709-5712.

113. Larsen, T.O., B.O. Petersen, and J.O. Duus, *Lumpidin, a novel biomarker of some ochratoxin A producing penicillia.* J. Agric. Food Chem., 2001. **49**(10): p. 5081-5084.

114. Springer, J.P., *Absolute-configuration of nortryptoquivaline.* Tetrahedron Lett., 1979(4): p. 339-342.

115. Yamazaki, M. and E. Okuyama, *Isolation of some new tryptoquivaline-related metabolites from Aspergillus fumigatus.* Chem. Pharm. Bull., 1979. **27**: p. 1611-1617.

116. Chooi, Y.H., R. Cacho, and Y. Tang, *Identification of the viridicatumtoxin and griseofulvin gene clusters from Penicillium aethiopicum.* Chem. Biol., 2010. **17**(5): p. 483-494.

117. Wang, S., et al., *Functional characterization of the biosynthesis of radicicol, an Hsp90 inhibitor resorcylic acid lactone from Chaetomium chiversii.* Chem. Biol., 2008. **15**(12): p. 1328-38.

118. Wiest, A., et al., *Identification of peptaibols from Trichoderma virens and cloning of a peptaibol synthetase.* J. Biol. Chem., 2002. **277**(23): p. 20862-20868.

119. Rausch, C., et al., *Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs).* Nucleic Acids Res., 2005. **33**(18): p. 5799-5808.

120. Pfeifer, B.A., et al., *Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli.* Science, 2001. **291**(5509): p. 1790-2.

121. Kutchan, T.M. and H. Dittrich, *Characterization and mechanism of the berberine bridge enzyme, a covalently flavinylated oxidase of benzophenanthridine alkaloid biosynthesis in plants.* J. Biol. Chem., 1995. **270**(41): p. 24475-81.

122. Ames, B.D., et al., *Complexity generation in fungal peptidyl alkaloid biosynthesis: oxidation of fumiquinazoline A to the heptacyclic hemiaminal fumiquinazoline C by the flavoenzyme af12070 from Aspergillus fumigatus.* Biochemistry, 2011. **50**(40): p. 8756-8769.

123. Kawamoto, S., et al., *Cloning and expression of Der f 6, a serine protease allergen from the house dust mite, Dermatophagoides farinae.* Biochim. Biophys. Acta, 1999. **1454**(2): p. 201-7.

124. Fujimoto, H., et al., *Isolation of new tremorgenic metabolites from an Ascomycete, Corynascus setosus.* Chem. Pharm. Bull., 1996. **44**(10): p. 1843-1848.

125. Mhaske, S.B. and N.P. Argade, *The chemistry of recently isolated naturally occurring quinazolinone alkaloids.* Tetrahedron, 2006. **62**(42): p. 9787-9826.

126. Zhang, S., et al., *A fragmented aflatoxin-like gene cluster in the forest pathogen Dothistroma septosporum.* Fungal Genet. Biol., 2007. **44**(12): p. 1342-54.

127. Bruckner, H., et al., *Aib and Iva in the biosphere: neither rare nor necessarily extraterrestrial.* Chem. Biodivers., 2009. **6**(1): p. 38-56.

128. Sansom, M.S.P., *The biophysics of peptide models of ion channels.* Prog. Biophys. Mol. Biol. , 1991. **55**(3): p. 139-235.

129. Kubicek, C.P., et al., *Facts and challenges in the understanding of the biosynthesis of peptaibols by Trichoderma.* Chem. Biodivers., 2007. **4**(6): p. 1068-1082.

130. Pojer, F., S.M. Li, and L. Heide, *Molecular cloning and sequence analysis of the clorobiocin biosynthetic gene cluster: new insights into the biosynthesis of aminocoumarin antibiotics.* Microbiology, 2002. **148**(Pt 12): p. 3901-11.

131. Pojer, F., et al., *CloR, a bifunctional non-heme iron oxygenase involved in clorobiocin biosynthesis.* J. Biol. Chem., 2003. **278**(33): p. 30661-30668.

132. Raap, J., et al., *Fungal biosynthesis of non-ribosomal peptide antibiotics and alpha,alpha-dialkylated amino acid constituents.* J. Pept. Sci., 2005. **11**(6): p. 331-338.

133. Reiber, K., et al., *A nonribosomal peptide synthetase involved in the biosynthesis of ampullosporins in Sepedonium ampullosporum.* J. Pept. Sci., 2003. **9**(11-12): p. 701-713.

134. Mohr, H. and H. Kleinkauf, *Alamethicin biosynthesis - acetylation of amino terminus and attachment of phenylalaninol.* Biochim. Biophys. Acta, 1978. **526**(2): p. 375-386.

135. Szewczyk, E., et al., *Fusion PCR and gene targeting in Aspergillus nidulans.* Nat. Protoc., 2006. **1**(6): p. 3111-3120.

136. Pall, M. and J. Brunelli, *A series of six compact fungal transformation vectors containing polylinkers with multiple unique restriction sites.* Fungal Genet. Newsl., 1993. **40**: p. 59.

137. Mullaney, E.J., et al., *Primary structure of the trpc gene from aspergillus-nidulans.* Mol. Gen. Genet., 1985. **199**(1): p. 37-45.

138. Punt, P.J., et al., *Functional elements in the promoter region of the aspergillus-nidulans gpda gene encoding glyceraldehyde-3-phosphate dehydrogenase.* Gene, 1990. **93**(1): p. 101-109.

139. Chooi, Y.H., et al., *Cloning and sequence characterization of a non-reducing polyketide synthase gene from the lichen Xanthoparmelia semiviridis.* Mycol. Res., 2008. **112**: p. 147-161.

140. Jin, J.-M., et al., *Functional characterization and manipulation of the apicidin biosynthetic pathway in Fusarium semitectum.* Mol. Microbiol., 2010. **76**(2): p. 456-466.

141. Slightom, J.L., et al., *Cloning and molecular characterization of the gene encoding the Aureobasidin A biosynthesis complex in Aureobasidium pullulans BP-1938.* Gene, 2009. **431**(1-2): p. 67-79.

142. Winterberg, B., et al., *Elucidation of the complete ferrichrome A biosynthetic pathway in Ustilago maydis.* Mol. Microbiol., 2010. **75**(5): p. 1260-1271.

143. Gao, X., et al., *Fungal indole alkaloid biosynthesis: Genetic and biochemical investigation of the tryptoquialanine pathway in Penicillium aethiopicum.* J. Am. Chem. Soc., 2011. **133**(8): p. 2729-2741.

144. Bergendahl, V., U. Linne, and M.A. Marahiel, *Mutational analysis of the C-domain in nonribosomal peptide synthesis.* Eur. J. Biochem., 2002. **269**(2): p. 620-629.

145. Keating, T.A., et al., *The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.* Nat. Struct. Mol. Biol., 2002. **9**(7): p. 522-526.

146. Samel, S.A., et al., *Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase.* Structure, 2007. **15**(7): p. 781-792.

147. von Doehren, H., *A survey of nonribosomal peptide synthetase (NRPS) genes in Aspergillus nidulans.* Fungal Genet. Biol., 2009. **46**: p. S45-S52.

148. Shao, Z., H. Zhao, and H. Zhao, *DNA assembler, an in vivo genetic method for rapid construction of biochemical pathways.* Nucleic Acids Res., 2009. **37**(2).

149. Ma, S.M., et al., *Complete reconstitution of a highly reducing iterative polyketide synthase.* Science, 2009. **326**(5952): p. 589-592.

150. Mootz, H.D., K. Schorgendorfer, and M.A. Marahiel, *Functional characterization of 4 '-phosphopantetheinyl transferase genes of bacterial and fungal origin by complementation of Saccharomyces cerevisiae lys5.* FEMS Microbiol. Lett. , 2002. **213**(1): p. 51-57.

151. Rusnak, F., et al., *Biosynthesis of the Escherichia-coli siderophore enterobactin - sequence of the entf gene, expression and purification of entf, and analysis of covalent phosphopantetheine.* Biochemistry, 1991. **30**(11): p. 2916-2927.

152. Stachelhaus, T. and C.T. Walsh, *Mutational analysis of the epimerization domain in the initiation module PheATE of gramicidin S synthetase.* Biochemistry, 2000. **39**(19): p. 5775-5787.

153. Linne, U. and M.A. Marahiel, *Control of directionality in nonribosomal peptide synthesis: Role of the condensation domain in preventing misinitiation and timing of epimerization.* Biochemistry, 2000. **39**(34): p. 10439-10447.

154. Quadri, L.E.N., et al., *Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases.* Biochemistry, 1998. **37**(6): p. 1585-1595.

155. Balibar, C.J. and C.T. Walsh, *GliP, a multimodular nonribosomal peptide synthetase in Aspergillus fumigatus, makes the diketopiperazine scaffold of gliotoxin.* Biochemistry, 2006. **45**(50): p. 15029-15038.

156. Keating, T.A., D.A. Miller, and C.T. Walsh, *Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of yersiniabactin synthetase.* Biochemistry, 2000. **39**(16): p. 4729-4739.

157. Lee, J.H., et al., *In vitro characterization of a heterologously expressed nonribosomal peptide synthetase involved in phosphinothricin tripeptide biosynthesis.* Biochemistry, 2009. **48**(23): p. 5054-5056.

158. Wang B, K.Q., Lu Y, Bai L, Wang C., *Unveiling the biosynthetic puzzle of destruxins in Metarhizium species* Proc. Nat. Acad. Sci. USA, 2012.

159. Grindberg, R.V., et al., *Single Cell Genome Amplification Accelerates Identification of the Apratoxin Biosynthetic Pathway from a Complex Microbial Assemblage.* Plos One, 2011. **6**(4).

160. Gatto, G.J., et al., *Elucidating the substrate specificity and condensation domain activity of FkbP, the FK520 pipecolate-incorporating enzyme.* Biochemistry, 2005. **44**(16): p. 5993-6002.

161. Schwecke, T., et al., *The biosynthetic gene-cluster for the polyketide immunosuppressant rapamycin.* Proc. Nat. Acad. Sci. USA, 1995. **92**(17): p. 7839-7843.

162. Sieber, S.A., et al., *Peptidyl thiophenols as substrates for nonribosomal peptide cyclases.* Angew. Chem. Int. Ed., 2004. **43**(4): p. 493-498.

163. Sieber, S.A., C.T. Walsh, and M.A. Marahiel, *Loading peptidyl-coenzyme A onto peptidyl carrier proteins: A novel approach in characterizing macrocyclization by thioesterase domains.* J. Am. Chem. Soc., 2003. **125**(36): p. 10862-10866.

164.	Roche, E.D. and C.T. Walsh, *Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis.* Biochemistry, 2003. **42**(5): p. 1334 - 1344.

165.	Fischbach, M.A., et al., *Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes.* Proc. Nat. Acad. Sci. USA, 2007. **104**(29): p. 11951-11956.

166.	Stachelhaus, T., A. Schneider, and M.A. Marahiel, *Rational design of peptide antibiotics by targeted replacement of bacterial and fungal domains.* Science, 1995. **269**(5220): p. 69-72.

167.	Mootz, H.D., D. Schwarzer, and M.A. Marahiel, *Construction of hybrid peptide synthetases by module and domain fusions.* Proc. Nat. Acad. Sci. USA, 2000. **97**(11): p. 5848-5853.

168.	Sattely, E.S. and C.T. Walsh, *A Latent Oxazoline Electrophile for N−O−C Bond Formation in Pseudomonine Biosynthesis.* Journal of the American Chemical Society, 2008. **130**(37): p. 12282-12284.

169.	Keating, T.A., C.G. Marshall, and C.T. Walsh, *Vibriobactin Biosynthesis in Vibrio cholerae: VibH Is an Amide Synthase Homologous to Nonribosomal Peptide Synthetase Condensation Domains†.* Biochemistry, 2000. **39**(50): p. 15513-15521.

170.	Horikawa, S., et al., *Molecular-cloning and nucleotide-sequence of cdna-encoding the rat-kidney s-adenosylmethionine synthetase.* J. Biol. Chem., 1990. **265**(23): p. 13683-13686.

171.	Soding, J., A. Biegert, and A.N. Lupas, *The HHpred interactive server for protein homology detection and structure prediction.* Nucleic Acids Res., 2005. **33**(suppl_2): p. W244-248.

172.	Chang, R., et al., *A potent nonpeptide cholecystokinin antagonist selective for peripheral tissues isolated from Aspergillus alliaceus.* Science, 1985. **230**(4722): p. 177-179.

173.	Karwowski, J.P., et al., *5-N-acetylardeemin, a novel heterocyclic compound which reverses multiple drug resistance in tumor cells. I. Taxonomy and fermentation of the producing organism and biological activity.* J. Antibiot. (Tokyo), 1993. **46**(3): p. 374-9.

174.	Liesch, J.M., et al., *Novel cholecystokinin antagonists from Aspergillus alliaceus. II. Structure determination of asperlicins B, C, D, and E.* J. Antibiot. (Tokyo), 1988. **41**(7): p. 878-81.

175.	Haynes, S.W., et al., *Unraveling terminal C-domain-mediated condensation in fungal biosynthesis of imidazoindolone metabolites.* Biochemistry, 2011. **50**(25): p. 5668-5679.

176.	Haynes, S.W., et al., *Complexity generation in fungal peptidyl alkaloid biosynthesis: a two-enzyme pathway to the hexacyclic mdr export pump inhibitor ardeemin.* ACS Chem. Biol., 2013. **doi: 10.1021/cb3006787**.

177.	Haynes, S.W., et al., *Assembly of asperlicin peptidyl alkaloids from anthranilate and tryptophan: a two-enzyme pathway generates heptacyclic scaffold complexity in asperlicin E.* J. Am. Chem. Soc., 2012. **134**(42): p. 17444-7.

178. Walsh, C.T., et al., *Short pathways to complexity generation: fungal peptidyl alkaloid multicyclic scaffolds from anthranilate building blocks.* ACS Chem. Biol., 2013: p. in press.

179. Gao, X., et al., *Cyclization of fungal nonribosomal peptides by a terminal condensation-like domain.* Nat. Chem. Biol., 2012. **8**: p. 823-830.

180. Nayak, T., et al., *A versatile and efficient gene-targeting system for Aspergillus nidulans.* Genetics, 2006. **172**(3): p. 1557-66.

181. Bruice, T.C., *A view at the millennium: the efficiency of enzymatic catalysis.* Acc. Chem. Res., 2002. **35**(Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.): p. 139-148.

182. Bruice, T.C. and F.C. Lightstone, *Ground state and transition state contributions to the rate of intramolecular and enzymic reactions.* Acc. Chem. Res., 1999. **32**(Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.): p. 127-136.

183. Hur, S. and T.C. Bruice, *The near attack conformation approach to the study of the chorismate to prephenate reaction.* Proc. Nat. Acad. Sci. USA, 2003. **100**(Copyright (C) 2011 American Chemical Society (ACS). All Rights Reserved.): p. 12015-12020.

184. Becke, A.D., *Density-functional thermochemistry. III. The role of exact exchange.* J. Chem. Phys., 1993. **98**(7): p. 5648-5652.

185. Lee, C., W. Yang, and R.G. Parr, *Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density.* Phys. Rev. B Condens. Matter., 1988. **37**(2): p. 785-789.

186. Gaussian 09, R.A., et al., *Gaussian, Inc., Wallingford CT.* 2009.

187. Ribeiro, R.F., et al., *Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation.* J. Phys. Chem. B, 2011. **115**(49): p. 14556-14562.

188. Bauschlicher Jr, C.W., *A comparison of the accuracy of different functionals.* Chem. Phys. Lett. , 1995. **246**(1–2): p. 40-44.

189. Merrick, J.P., D. Moran, and L. Radom, *An evaluation of harmonic vibrational frequency scale factors.* J. Phys. Chem. A, 2007. **111**(45): p. 11683-11700.

190. Gonzalez, C. and H.B. Schlegel, *An improved algorithm for reaction path following.* J. Chem. Phys., 1989. **90**(4): p. 2154-2161.

191. Gonzalez, C. and H.B. Schlegel, *Reaction path following in mass-weighted internal coordinates.* J. Chem. Phys., 1990. **94**(14): p. 5523-5527.

192. Marenich, A.V., C.J. Cramer, and D.G. Truhlar, *Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions.* J. Phys. Chem. B, 2009. **113**(18): p. 6378-6396.

193. Fremlin, L.J., et al., *Cottoquinazoline A and cotteslosins A and B, metabolites from an Australian marine-derived strain of Aspergillus versicolor.* J. Nat. Prod., 2009. **72**(19245260): p. 666-670.

194. Liesch, J.M., et al., *Novel cholecystokinin antagonists from Aspergillus alliaceus. II. Structure determination of asperlicins B, C, D, and E.* Journal of Antibiotics (Tokyo), 1988. **41**(7): p. 878-81.

195. Hochlowski, J.E., et al., *5-N-acetylardeemin, a novel heterocyclic compound which reverses multiple drug resistance in tumor cells. II. Isolation and elucidation of the structure of 5-N-acetylardeemin and two congeners.* J. Antibiot. (Tokyo), 1993. **46**(3): p. 380-6.

196. Karwowski, J.P., et al., *5-N-acetylardeemin, a novel heterocyclic compound which reverses multiple drug resistance in tumor cells. I. Taxonomy and fermentation of the producing organism and biological activity.* Journal of Antibiotics (Tokyo), 1993. **46**(3): p. 374-9.

197. Hochlowski, J.E., et al., *5-N-acetylardeemin, a novel heterocyclic compound which reverses multiple drug resistance in tumor cells. II. Isolation and elucidation of the structure of 5-N-acetylardeemin and two congeners.* Journal of Antibiotics (Tokyo), 1993. **46**(3): p. 380-6.

198. Juliano, R.L. and V. Ling, *A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants.* Biochimica et Biophysica Acta, 1976. **455**(990323): p. 152-162.

199. Kartner, N., J.R. Riordan, and V. Ling, *Cell surface P-glycoprotein associated with multidrug resistance in mammalian cell lines.* Science, 1983. **221**(6137059): p. 1285-1288.

200. Chou, T.C., et al., *Reversal of anticancer multidrug resistance by the ardeemins.* Proc. Nat. Acad. Sci. USA, 1998. **95**(9653193): p. 8369-8374.

201. Depew, K.M., et al., *Total Synthesis of 5-N-Acetylardeemin and Amauromine: Practical Routes to Potential MDR Reversal Agents.* J. Am. Chem. Soc., 1999. **121**(51): p. 11953-11963.

202. Avalos, J., R.F. Geever, and M.E. Case, *Bialaphos resistance as a dominant selectable marker in Neurospora crassa.* Curr. Genet., 1989. **16**(2532965): p. 369-372.

203. Broberg, A., A. Menkis, and R. Vasiliauskas, *Kutznerides 1-4, depsipeptides from the actinomycete Kutzneria sp. 744 inhabiting mycorrhizal roots of Picea abies seedlings.* J. Nat. Prod., 2006. **69**(1): p. 97-102.

204. Fujimori, D.G., et al., *Cloning and characterization of the biosynthetic gene cluster for kutznerides.* Proc. Nat. Acad. Sci. USA, 2007. **104**(42): p. 16498-503.

205. Akrigg, A., S.R. Ayad, and G.R. Barker, *The nature of a competence-inducing factor in Bacillus subtilis.* Biochem. Biophys. Res. Commun., 1967. **28**(4965071): p. 1062-1067.

206. Magnuson, R., J. Solomon, and A.D. Grossman, *Biochemical and genetic characterization of a competence pheromone from B. subtilis.* Cell, 1994. **77**(8168130): p. 207-216.

207. Okada, M., et al., *Structure of the Bacillus subtilis quorum-sensing peptide pheromone ComX.* Nat. Chem. Biol., 2005. **1**(16407988): p. 23-24.

208. Okada, M., et al., *Chemical structure of posttranslational modification with a farnesyl group on tryptophan.* Bioscience, Biotechnology, and Biochemistry, 2008. **72**(18323630): p. 914-918.

209. Walsh, C.T., S.W. Haynes, and B.D. Ames, *Aminobenzoates as building blocks for natural product assembly lines.* Natural Product Reports, 2012. **29**(1): p. 37-59.

210. Xu, W., et al., *Analysis of Intact and dissected fungal polyketide synthase-nonribosomal peptide synthetase in Vitro and in Saccharomyces cerevisiae.* J. Am. Chem. Soc., 2010. **132**(39): p. 13604-13607.

211. Istvan, E.S. and J. Deisenhofer, *Structural mechanism for statin inhibition of HMG-CoA reductase.* Science, 2001. **292**: p. 1160-1164.

212. Hoffman, W.F., et al., *3-Hydroxy-3-methylglutaryl-coenzyme A reductase inhibitors .4. side-chain ester derivatives of mevinolin.* J. Med. Chem., 1986. **29**: p. 849-852.

213. Berg, V.A., M. Hans, and H. Steekstra, *Method for the production of simvastatin.* WO 2007147801 (A1) 2009.

214. Morgan, B., et al., *Methods for making simvastatin and intermediates.* WO 2005040107 (A2), 2006.

215. Wagner, U.G., et al., *EstB from Burkholderia gladioli: a novel esterase with a beta-lactamase fold reveals steric factors to discriminate between esterolytic and beta-lactam cleaving activity.* Protein Sci., 2002. **11**: p. 467-78.

216. Petersen, E.I., et al., *A novel esterase from Burkholderia gladioli which shows high deacetylation activity on cephalosporins is related to beta-lactamases and DD-peptidases.* J. Biotechnol., 2001. **89**(1): p. 11-25.

217. Arnold, F.H., *Combinatorial and computational challenges for biocatalyst design.* Nature, 2001. **409**: p. 253-257.

218. Arnold, F.H. and A.A. Volkov, *Directed evolution of biocatalysts.* Curr Opin Chem Biol, 1999. **3**: p. 54-9.

219. Kumar, M.S., et al., *A rapid technique for screening of lovastatin-producing strains of Aspergillus terreus by agar plug and Neurospora crassa bioassay.* J. Microbiol. Methods, 2000. **40**: p. 99-104.

220. Xie, X., et al., *Rational improvement of simvastatin synthase solubility in Escherichia coli leads to higher whole-cell biocatalytic activity.* Biotechnol. Bioeng., 2009. **102**: p. 20-8.

221. Fromant, M., S. Blanquet, and P. Plateau, *Direct random mutagenesis of gene-sized DNA fragments using polymerase chain-reaction.* Anal. Biochem., 1995. **224**: p. 347-353.

222. Heikinheimo, P., et al., *Of barn owls and bankers: a lush variety of alpha/beta hydrolases.* Structure, 1999. **7**: p. R141-R146.

223. Nardini, M. and B.W. Dijkstra, *Alpha/beta hydrolase fold enzymes: the family keeps growing.* Curr. Opin. Struct. Biol., 1999. **9**: p. 732-7.

224. Lai, J.R., A. Koglin, and C.T. Walsh, *Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis.* Biochemistry, 2006. **45**: p. 14869-79.

225. Hsu, C.C., et al., *Directed evolution of D-sialic acid aldolase to L-3-deoxy-manno-2-octulosonic acid (L-KDO) aldolase.* Proc. Nat. Acad. Sci. USA, 2005. **102**: p. 9122-9126.

226. Oue, S., et al., *Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues.* J. Biol. Chem., 1999. **274**: p. 2344-2349.

227. Zhao, H.M. and F.H. Arnold, *Directed evolution converts subtilisin E into a functional equivalent of thermitase.* Protein Eng., 1999. **12**: p. 47-53.

228. Schweiker, K.L., et al., *Computational design of the Fyn SH3 domain with increased stability through optimization of surface charge charge interactions.* Protein Sci., 2007. **16**(12): p. 2694-702.

229. Oefner, C., et al., *Refined crystal-structure of beta-lactamase from citrobacter-freundii Indicates a mechanism for beta-lactam hydrolysis.* Nature, 1990. **343**: p. 284-288.

230. Gao, X., et al., *Directed evolution and structural characterization of a simvastatin synthase.* Chem. Biol., 2009. **16**(10): p. 1064-1074.

231. Kabsch, W., *XDS.* Acta Crystallogr. Sect. D-Biol. Crystallogr. , 2010. **66**: p. 125-132.

232. McCoy, A.J., et al., *Phaser crystallographic software.* J. Appl. Crystallogr., 2007. **40**: p. 658-674.

233. Murshudov, G.N., A.A. Vagin, and E.J. Dodson, *Refinement of macromolecular structures by the maximum-likelihood method.* Acta Crystallogr. Sect. D-Biol. Crystallogr., 1997. **53**: p. 240-255.

234. Blanc, E., et al., *Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT.* Acta Crystallogr. Sect. D-Biol. Crystallogr., 2004. **60**: p. 2210-2221.

235. Winn, M.D., G.N. Murshudov, and M.Z. Papiz, *Macromolecular TLS refinement in REFMAC at moderate resolutions.* Methods Enzymol., 2003. **374**: p. 300-321.

236. Emsley, P., et al., *Features and development of Coot.* Acta Crystallogr. D Biol. Crystallogr., 2010. **66**: p. 486-501.

237. Laskowski, R.A., et al., *Procheck - a program to check the stereochemical quality of protein structures.* J. Appl. Crystallogr., 1993. **26**: p. 283-291.

238. Colovos, C. and T.O. Yeates, *Verification of protein structures: patterns of nonbonded atomic interactions.* Protein Sci., 1993. **2**: p. 1511-9.

239. Luthy, R., J.U. Bowie, and D. Eisenberg, *Assessment of protein models with three-dimensional profiles.* Nature, 1992. **356**: p. 83-5.

240. The PyMOL Molecular Graphics System, V., Schrödinger, LLC (http://www.pymol.org).

241. Tamura, K., et al., *MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0.* Mol. Biol. Evol., 2007. **24**(8): p. 1596-1599.

242. Larkin, M.A., et al., *Clustal W and clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.

243. Doublie, S., *Production of selenomethionyl proteins in prokaryotic and eukaryotic expression systems.* Methods Mol. Biol., 2007. **363**: p. 91-108.

244. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Method Enzymol., 1997. **276**: p. 307-326.

245. Long, F., et al., *BALBES: a molecular-replacement pipeline.* Acta Crystallogr., Sect D: Biol Crystallogr, 2008. **64**: p. 125-132.

246. Brunger, A.T., et al., *Crystallography & NMR system: A new software suite for macromolecular structure determination.* Acta Crystallogr. D Biol. Crystallogr., 1998. **54**: p. 905-21.

247. Fabiola, F., et al., *An improved hydrogen bond potential: impact on medium resolution protein structures.* Protein Sci., 2002. **11**: p. 1415-23.

248. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr., Sect D: Biol Crystallogr, 2004. **60**: p. 2126-2132.