Discovery and biosynthetic investigation of a new antibacterial dehydrated non-ribosomal tripeptide

Shan Wang*[1, #], Qing Fang[1, #], Zhou Lu[1,2], Yingli Gao[1,3], Laurent Trembleau[1], Rainer Ebel[1], Jeanette H. Andersen[4], Carol Philips[5], Samantha Law [5], and Hai Deng*[1]

1. Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Meston Walk, Aberdeen AB24 3UE, Scotland, UK
2. Guangdong Provincial Key Laboratory of Microbial Culture Collection and Application, State Key Laboratory of Applied Microbiology Southern China, Guangdong Institute of Microbiology, Guangdong Academy of Sciences
3. College of Marine Life and Fisheries, Jiangsu Ocean University, Lianyungang, Jiangsu Province, China
4. Marbio, UiT - The Arctic University of Norway, Tromsø N-9037, Norway
5. NCIMB Ltd, Ferguson Building, Craibstone Estate, Bucksburn, Aberdeen, AB21 9YA, Scotland, UK

[#] These authors contributed equally.

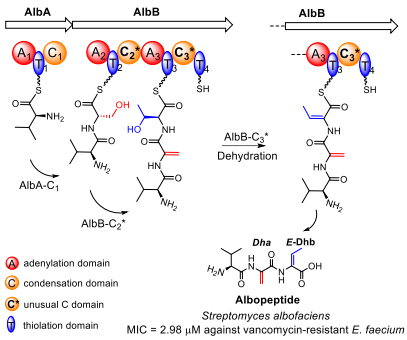
* Corresponding authors: Dr Hai Deng, h.deng@abdn.ac.uk

Dr Shan Wang, shan.wang@abdn.ac.uk

Abstract

Dehydroalanine (Dha) and dehydrobutyrine (Dhb) display considerable flexibility in a variety of chemical and biological reactions. Natural products containing Dha and/or Dhb residues are often found to possess diverse biological activities. While the (*Z*)-geometry is predominant in nature, only a handful of metabolites containing (*E*)-Dhb have been found thus far. Here we report discovery of a new antimicrobial peptide, albopeptide, through NMR analysis and chemical synthesis, which contains two contiguous unsaturated residues, Dha-(*E*)-Dhb. It displays narrow-spectrum activity against vancomycin-resistant *Enterococcus faecium. In vitro* biochemical assays show that albopeptide originates from a noncanonical NRPS pathway, featured with dehydration processes, catalysed by unusual condensation domains. Finally, we provide evidence of the occurrence of a previously untapped group of short unsaturated peptides in the bacterial kingdom, suggesting an important biological function in bacteria.

Table of Content



Timing of dehydration: Albopeptide, a new antibacterial short peptidyl metabolite, was found to contain a unique contiguous Dha-(*E*)-Dhb moiety. *In vitro* reconstitution and biochemical analysis together with chemical capture of online biosynthetic intermediates demonstrated that two condensation domains in the corresponding NRPSs catalyse the dehydration reactions on the Ser and Thr residues, respectively, to generate Dha and (*E*)-Dhb residues.

1 Introduction

Dehydroalanine (Dha) and dehydrobutyrine (Dhb) are noncanonical amino acids found in in a wide array of natural products (NPs) from various organisms, many of which display antibacterial, antifungal, antitumor and phytotoxic activity[1]. Both Dha and Dhb have profound effects on the properties of the molecule. They display significant synthetic flexibility, readily undergoing reactions such as Michael additions, transition-metal-catalysed cross-couplings, and cycloadditions. They are also used as orthogonal chemical handles for later stage modification of biomolecules[2].

Dha is the simplest dehydroamino acid (dhAA) without geometrical isomers while Dhb is the simplest dhAA displaying *E*/*Z* isomerism. The *Z* isomer of Dhb is sterically favourable [3] and is present in a majority of NPs containing the Dhb residue. Dha and (*Z*)-Dhb motifs are commonly found in ribosomally synthesized and post-translationally modified peptides (RiPPs). Dha and Dhb motifs in RiPPs are generated from Ser/Cys and Thr residues, respectively, via in *trans* activation, followed by elimination with *anti* stereoselectivity (in the case of L-Thr) by dedicated biosynthetic enzymes[4].

Although less common, Dha and Dhb motifs do appear in non-ribosomal peptides (NRPs). For example, they are present in the cyclic pentadepsipeptide, phomalide **1**[5], the cycloheptapeptide, [D-Asp^3, (*Z*)-Dhb^7] microcystin **2**[6], the β-lactam antibiotic norcardicin **3**[7], the pyrrolizidine alkaloid azabicyclene **4**[8] and the nonproteinogenic amino acid methoxyvinylglycine **5**[9] (Figure 1A). In contrast to RiPPs, both (*Z*)- and (*E*) -Dhb geometries have been found in NRPs. A typical example is microcystins (MCs), the most widespread class of cyanotoxins discovered in cyanobacteria. The majority of MCs consists of *N*-methyl-Dha residue at position 7[6]. Around 30 MC congeners were found to contain the Dhb moiety, among which only two have been characterized by NMR structural elucidation to be (*Z*)-Dhb residues, including MC congener **2** discovered from an environmental *Oscillatoria agardhii* isolate [10]. Ten of MC congeners consist of (*E*)-Dhb moiety [6], for example, the isomer of **2** containing *E*-Dhb motif was isolated from *O. agardhii* CCAP1459/14[11]. It is worth to note that the geometry of Dhb residue in some of these NPs is crucial for their biological activity[13,14]. An interesting example can be found in phomalide **1** (Figure 1A) and its synthetic isomer, isophomalide[13]. Phomalide **1** is a selective phytotoxin, causing leaf spot and stem cankers (blackleg)[5], while its isomer, isophomalide, is not a causative agent[13].

Recent biochemical studies indicated that one of the condensation domains (NocB-M5C) in the pathway of nocardicin **3** catalyses the dehydration of a seryl-containing tetrapeptidyl thioester to generate a transient Dha-containing intermediate for subsequent cyclization to afford the β-lactam ring system[6,11]. This C domain consists of the H(790)HHxxxDG catalytic motif, in contrast to the conventional C domains which only contain HHxxDG. The third His

residue (H790) was proposed to catalyse the β-elimination to generate cryptic Dha-containing intermediate[14]. *In vitro* biosynthetic reconstitution of two NRPS-derived metabolites, azabicyclene **4** and methoxyvinylglycine **5**, also suggested that two noncanonical C* domains are likely to be responsible for the formation of cryptic Dha and *cis*-α,β dehydroglumate intermediates, respectively, in the corresponding NRPS pathways[8,9]. However, the formation of Dhb residues in NRPs remains to be determined. This is the particular case for (*E*)-Dhb residues. Unlike (*Z*)-Dhb, the bio-origin of (*E*)-Dhb is less obvious. It has raised a question of whether (*E*)-Dhb residues in NRPs are derived from L-*allo*-Thr [12] via facile *anti*-elimination or they directly descend from *syn*-elimination in which two vicinal substituents (e.g. α-H and β-hydroxide) on the L-Thr residue leave simultaneously, an analogous process of chemical precedent under elevated temperatures and harsh conditions[3].

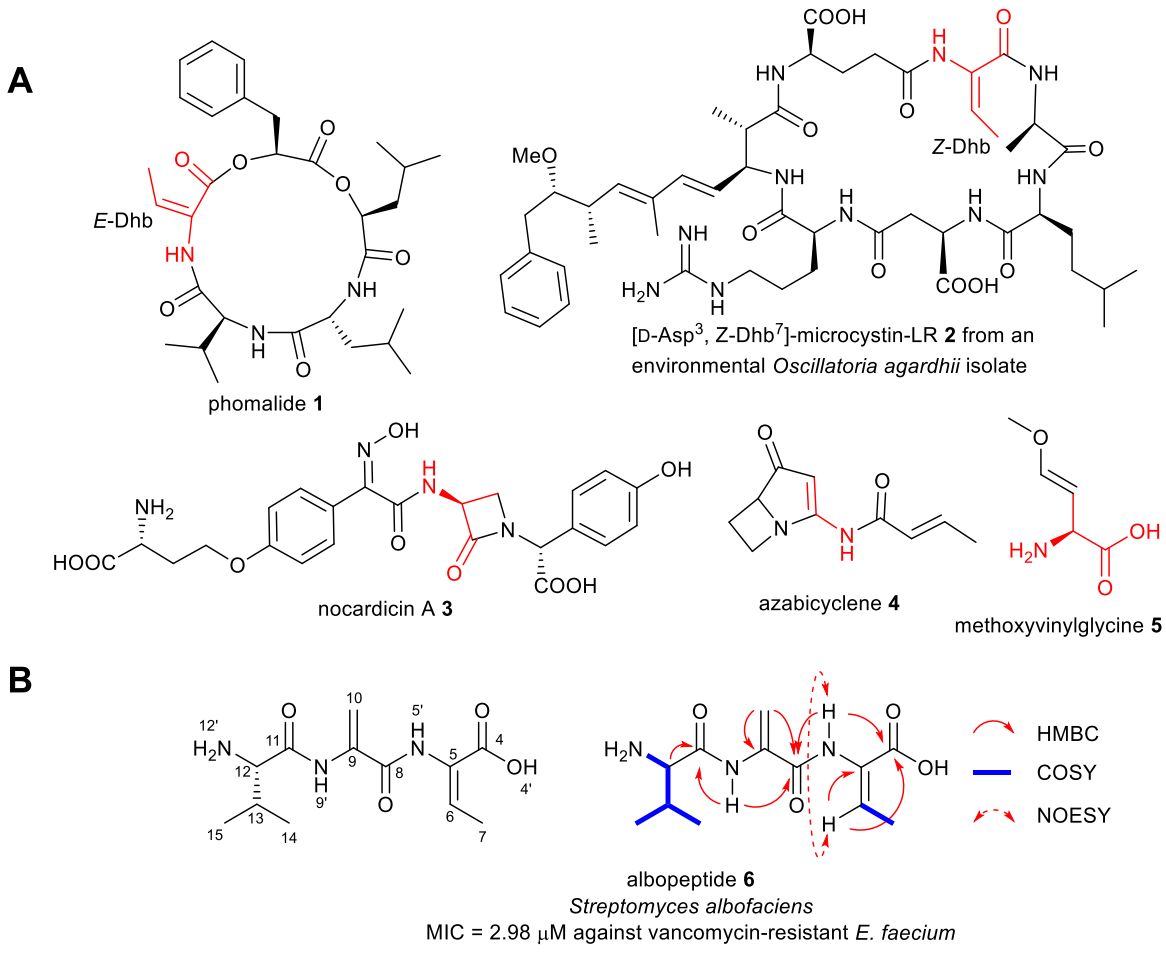


Figure 1. A. structural examples of NRPs containing residues derived from dhAAs (highlighted in red); **B**. Structure of the new short peptide, albopeptide **6**, isolated from *Streptomyces albofaciens,* with selective COSY, NOESY and HMBC correlations shown.

Herein, we report the identification of a new short linear peptide, albopeptide **6**, from the soil bacterium *Streptomyces albofaciens* (NCIMB 10975) (Figure 1B). The complete structure of

6 was determined to be L-valyl-Dha-(*E*)-Dhb through comprehensive NMR, high-resolution MS analyses and chemical synthesis together with NMR spectral comparisons of natural **6** with synthetic **6** and its *Z*-isomer **6a**. To the best of our knowledge, the contiguous Dha-(*E*)Dhb combination is unique in NPs. Bioinformatics and *in vitro* reconstitution showed that only two multidomain non-ribosomal peptide synthetases AlbA and AlbB are required for the production of albopeptide **6**. Our detailed biochemical analysis not only demonstrated the timing of the dehydration events during the **6** biosynthesis but also confirmed that two unusual C* domains, AlbB-C_2 and -C_3, catalyse the dehydration to generate Dha and (*E*)-Dhb residues, respectively. Comparative genomic analysis revealed the occurrence of similar biosynthetic gene clusters (BGCs) in Gram-negative and Gram-positive bacteria that potentially produce a previously unnoticed group of short dehydrated NRPs. Albopeptide **6** displays a potent narrow-spectrum antibacterial activity against a vancomycin-resistant *Enterococcus faecium* hospital isolate with a MIC of 2.98±0.07 µM.

2 Results

2.1 Discovery of a new short peptide 6

Over the course of efforts toward the search of new natural antimicrobial metabolites, we investigated metabolic profiles of *S. albofaciens* using liquid chromatography coupled to high resolution electrospray ionization mass spectroscopy (HR-ESIMS) analysis. The MS data were examined by molecular network analysis[15]. The crude extract generated from the Modified Bennett's medium was identified to contain structurally diverse classes of metabolites (Figure S1-4, Table S1). In addition, clusters of compounds having low (<300 Da) mass-to-charge ratios (*m*/*z*) were identified in the network. The absence of these observed masses in NP databases[16],[17] suggested the presence of new metabolites.

Subsequent large-scale fermentation (12L), followed by chemical workup, afforded the isolation of compound **6** (0.6 mg). The molecular formula of **6** (Figure 1) was established as $C_{12}H_{20}O_4N_3^+$ by HR-ESIMS (*obsd.* $[M+H]^+$ = 270.1449; *calcd.* $[M+H]^+$ = 270.1448; Δ = 0.37 ppm), with five degrees of unsaturation. The planar structure of **6** was deduced to be Val-Dha-Dhb by HR-ESIMS, 1D and 2D NMR analyses (Figure S2-7, Table S2).

Advanced Marfey's analysis confirmed that **6** contains L-valine (Figure S8). The determination of the geometry of Dhb moiety in **6** was based on the interpretation of ^{1}H NMR and ^{1}H-^{1}H NOESY spectra in DMSO-*d*6 (Figure S6). The olefinic proton in Dhb unit of **6** appears at 6.03 ppm, and displays a, albeit weak, correlation with 5'-H at the amide moiety, suggesting that the geometry of the Dhb unit in **6** is *E*. The absence of any (*Z*)-Dhb containing **6** isomer during the chemical workup and structural elucidation indicated that the isolated **6** is a pure *E* configurational isomer.

2.2 Synthesis of the isomers of 6

Finally support for the proposed structure of **6** was provided by the total synthesis of isomers **6** and **6a**, containing the *E* and *Z* configurations of the Dhb motifs, respectively (Figure 2). L-Thr was converted into its allyl ester and coupled with Boc-protected Ser to yield the dipeptide **7**. This peptide was further coupled with Boc-protected valine to give the Boc-L-Val-L-Ser-L-Thr allyl ester **8**. Subsequent elimination of the hydroxy groups of the Ser and Thr residues was performed by treatment with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and copper (II) chloride to generate the desired Dha and Dhb residues[18]. Another portion of **8** was treated with mesyl chloride and triethylamine to give **6a**, containing the (*Z*)-Dhb unit (100% ee)[3]. Finally, removal of the allyl and Boc groups using conventional reaction conditions afforded a mixture of **6** and **6a** containing the *E/Z* configured-Dhb motif (approx. 50:50) in the first case, and pure isomer **6a** containing the *Z*-configured Dhb unit in the second. Attempts to synthesize the pure *E*-isomer from the ester **8** was unsuccessful. It is well known that the thermodynamic instability of an (*E*)-Dhb residue in comparison to its (Z) counterpart makes its stereoselective construction highly challenging. [18,19] In general, the 1, 3-allylic strain of the (*E*)-Dhb residue is higher than that of the (Z) derivative due to steric interactions in (*E*)-Dhb. This led to generation of a mixture containing the thermodynamically stable *Z*-**6a** under the reaction conditions[3]. An earlier study also revealed difficulties in the introduction of a pure *E*-

Dhb residue in the synthesis of vioprolides[18]. Nonetheless, the *E/Z* isomeric mixture of **6** and **6a** provided the required information on the olefin geometry of isolated peptide **6**.

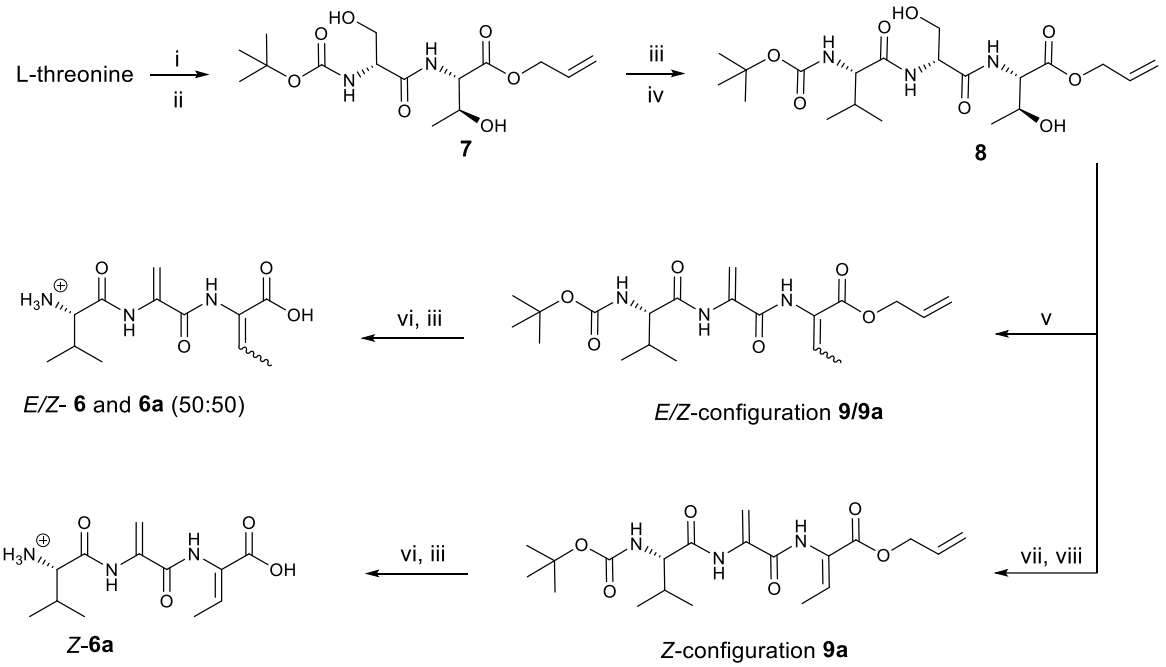


Figure 2. Synthesis route of *E*/*Z* albopeptide **6**, **6a**. Reagents and conditions: i, allyl alcohol, *p*-toluenesulfonic acid, 16 h, 110 °C; ii, Boc-L-Ser, PyBoP, HOBt, DIPEA, 22 h, RT, 56.1%; iii, TFA (quant.); iv, Boc-L-Val, PyBoP, HOBt, DIPEA, 22 h, RT, 39.4%; v, EDC 2eq. $CuCl_2$ 0.2eq, anhydrous toluene 80 °C , 2 h, argon, 20% (50.0%); vi, $Pd(PPh_3)_4$, $PhSiH_3$, DCM, 2 h, RT, 75%, vii, MsCl, Et_3N, 0 °C RT, 1h; viii, DBU, DCE, 85 °C, 17 h, 40.2 % (ee% 100%).

2.3 NMR spectral comparison among natural 6 and synthetic isomers

With the synthetic isomers **6** and **6a** available, we carried out detailed analyses and comparisons of the ^{1}H-NMR spectra in CD_3OD (Table S2, Figure 3) of the natural peptide **6** with the synthetic peptides **6** and **6a**. The natural compound **6** had identical proton chemical shifts to the synthetic one containing the *E*-Dhb motif (δ_H 6.46 and 2.09 (Hβ and CH_3 in Dhb, respectively) which are clearly different to the ones containing the *Z*-Dhb motif (δ_H 6.98 and 1.8, respectively) (Figure 3). This confirmed that the Dhb residue in natural **6** has indeed the *E* configuration.

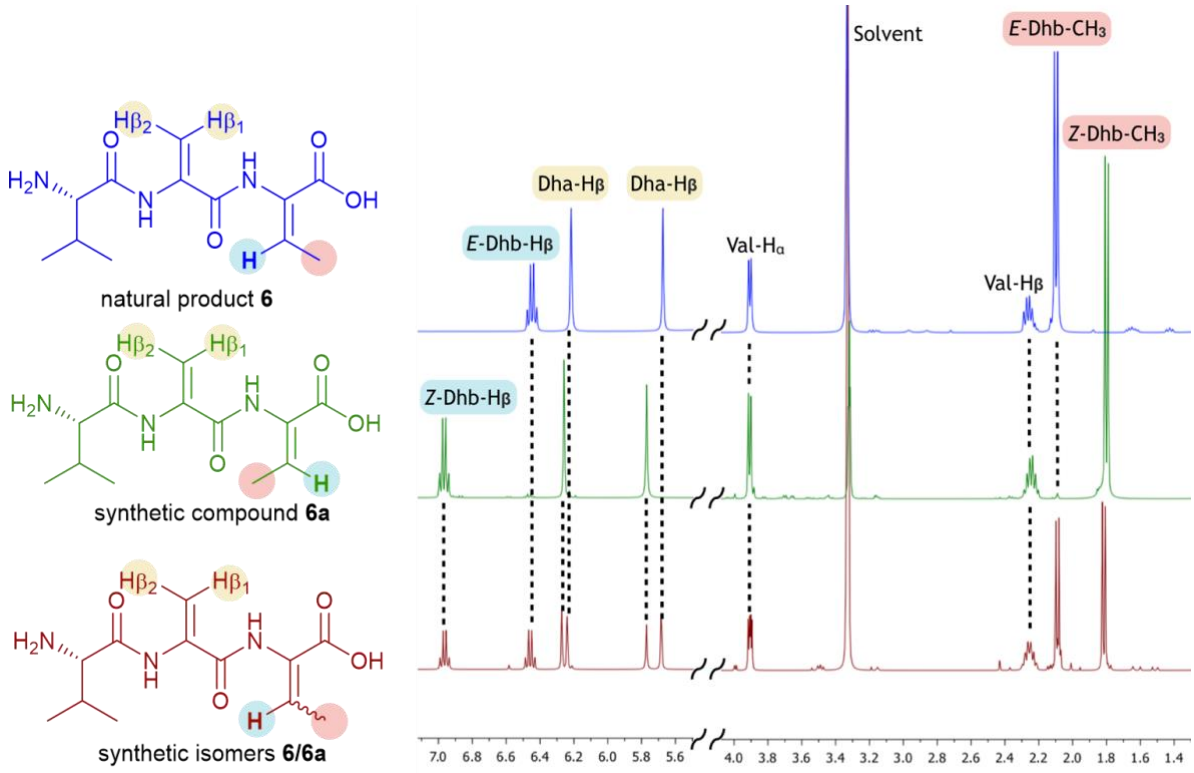


Figure 3. ^{1}H-NMR spectral comparison of natural compound **6,** synthetic Z-isomer **6a**, and synthetic mixtures of **6** and **6a**, confirming that the new short peptidyl metabolite, **6**, contains (*E*)-Dhb residue.

Another interesting observation in this NMR comparison is the chemical shifts of Dha residues in these two isomers. The geometry of the double bond of Dhb in these contiguous Dha-Dhb residues strongly influences the chemical shifts of two olefinic protons in the adjacent Dha residue in **6**, which appear upfield, compared to the respective signals in **6a** that contains a (*Z*)-Dhb moiety (Figure 3).

Taken together, our experimental data demonstrated that compound **6** is a new short peptide metabolite, which we named albopeptide, by association with the strain name.

2.4 Biological Tests

Due to the availability of sub-milligram quantities of natural albopeptide, the synthetic materials **6/6a** (50:50) and **6a**, were used to investigate their biological properties against a panel of standard bacterial pathogens and clinical isolates of bacterial pathogens (Materials and Methods). The peptide mixture **6/6a** displayed no significant activities against most of bacterial pathogens tested except for the clinical isolate vancomycin-resistant *Enterococcus faecium* K60-39[20] with a potent MIC value of 2.98±0.07 µM (Figure S24). Interestingly, *Z*-peptide **6a** showed no activity against *E. faecium* K60-39, indicating that albopeptide **6** is solely responsible for the observed antibiotic activity. Further analysis of the mode of action of **6** would be of considerable interest.

2.5 Identification of the biosynthetic gene cluster of 6

Molecular network analysis (Figure S1, Table S1) indicated that albopeptide **6** is only associated with one ion, suggesting that **6** may not be originated from ribosomally synthesized and post-translationally modified peptides (RiPPs), which normally have large molecular weights (>500 Da)[21]. This prompted us to investigate whether **6** may originate from a NRPS assembly. *In silico* analysis of the annotated draft genome (NCBI accession: NZ_PDCM01000002.1) allowed identification of the putative biosynthetic gene cluster (*alb* BGC) (Figure 4A). The centre of the *alb* BGC encodes two multidomain NRPSs (AlbA and AlbB), both of which possess arrangements of A_1-T_1-C_1 and A_2-T_2-C_2-A_3-T_3-C_3-T_4, respectively (Figure 4A). The A domains of AlbA and AlbB were predicted to activate Val, Ser and Thr [22], respectively, which is consistent with the predicted precursors. No gene/domain encoding thioesterase was found within the NRPS gene or in the close proximity of the *alb* BGC, suggesting an unusual release mechanism of thiolate-template peptidyl intermediate generated by NRPS.

Although only two peptide bonds are present in **6**, three condensation domains were observed, suggesting that some of these C domains may be involved in unconventional processes. Phylogenetic analysis of C domains[23] revealed that, while AlbA-C_1 is a normal L-amino acid condensation domain, AlbB-C_2 and C_3 belong to a unique group of C* domains which are also found in the biosynthetic pathways of NPs, such as methoxyvinylglycine[9], legonmycin[24], azabicyclene[8], puwainaphycin[25], FK228[26], bleomycin[27], hassallidin[28], nodularin[29] and microcystin[30] (Figure 4B). All of these NPs contain either dhAA motifs or motifs derived from dhAA residues.

2.6 In vitro pathway reconstitution and biochemical analysis.

To confirm the identity of these NRPSs, we set out to investigate the roles of both AlbA and AlbB in the biosynthesis of **6**. Overexpression of both *alb*A and *alb*B genes in BL21-CodonPlus

(DE3)-RP allowed for the purification of the corresponding recombinant proteins to near homogeneity, respectively (Figure S25).

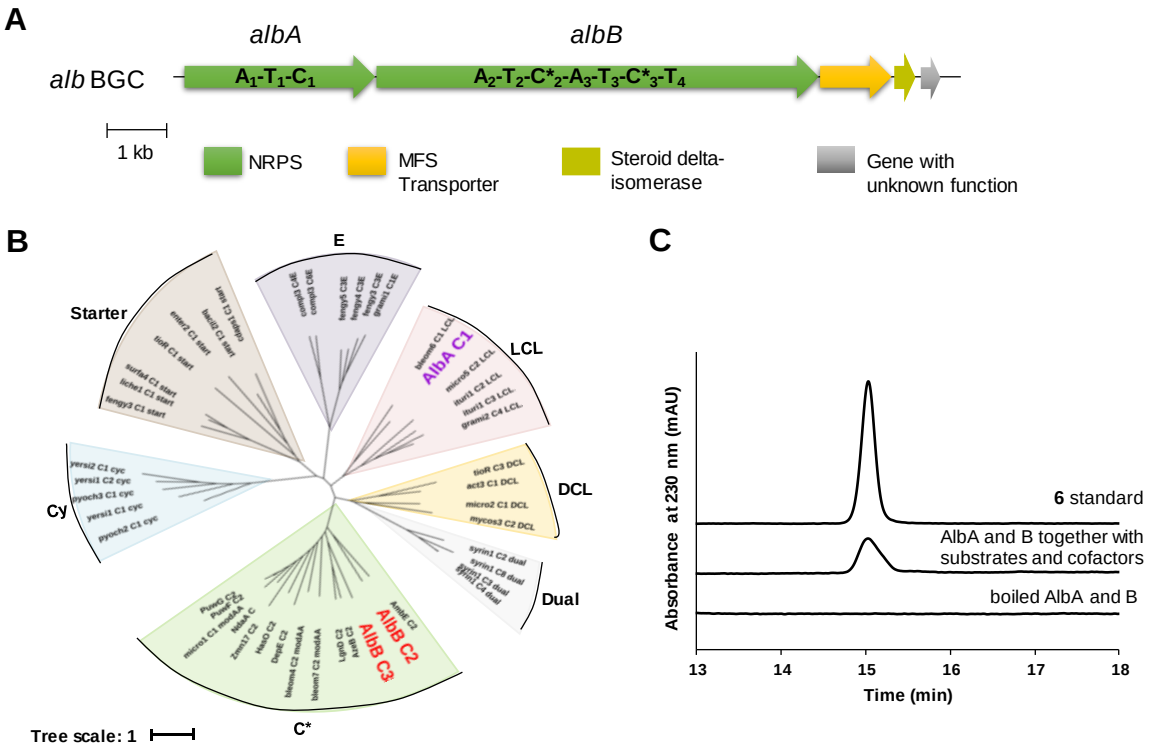


Figure 4. **A**. The arrangement of the *alb* gene cluster responsible for the biosynthesis of **6**. **B**. Phylogenetic analysis of C* domains in AlbB with C* domains identified in other biosynthetic pathways, suggesting that C_2 and C_3 domains belong to a unique group of C domains in bacteria. LCL: condensation between two L amino acids; DCL: condensation between a D- and a L-amino acid; Starter: acylation of the first amino acid with a β-hydroxy-carboxylic acid; Cy: cyclization; E: epimerization; Dual: epimerization and condensation. **C**. HPLC analysis for *in vitro* reconstitution catalysed by AlbA and AlbB, demonstrating that these two multidomain NRPSs are responsible for the production of **6**.

We first sought to reconstitute the biosynthesis of **6** *in vitro*. To this end, we performed a one-pot bioassay. AlbA and AlbB were first converted into their *holo* form by the phosphopantetheine (Ppant) transferase Sfp in the presence of Coenzyme A (CoA) and $MgCl_2$, followed by incubation with substrates Val, Ser and Thr and ATP. This assay yielded a product, which has the same retention time as the standard **6** as observed in HPLC analysis (Figure 4C). The corresponding ion is also identical to **6** based on LC-HR-MS and tandem MS analyses (Figure S26-27). Omission of each substrate, or each individual enzyme abolished the formation of **6** (Figure S26), confirming that each enzyme and substrate is necessary for the biosynthesis of **6**. It is worth to note that using L-*allo*-Thr to replace L-Thr in the above biotransformation resulted in no production of **6**, indicating that (*E*)-Dhb moiety in **6** is not derived from L-*allo*-Thr.

We next set out to investigate the dehydration events of the Ser and Thr residues during the biotransformation. In this case, we attempted a chemical derivatization strategy to capture the biosynthetic intermediate attached to the full length NRPSs, which involves the use of cysteamine as a thioester reactive probe[9]. Incubation of AlbA or/and AlbB (*holo* forms) with the amino acids together with ATP in the presence of cysteamine allowed chemical cleavage of the NRPS-bound intermediates to form cystamine adducts that can be detected by LC-HRMS (Figure 5 and Figure S28). Using this strategy, we detected cystamine-bound Val **15** as the major product accumulated when AlbA was assayed with L-Val and ATP (Figure 5B and Figure S28A), confirming that Val is indeed activated by A_1. In an assay of *holo*-AlbA and -AlbB with L-Val, L-Ser, and ATP, we were able to detect ions corresponding to both cystamine adducted Val-Ser **16** intermediate (Figure 5B and Figure S28B) and cystamine adducted Val-aminoethylcysteine (AEC) **17** (Figure 5B and Figure S28C). **17** is generated from the non-enzymatic product of Val-Dha in the presence of cysteamine under mild alkaline conditions[31]. Addition of L-Thr under the aforementioned assay allowed the identification of ions corresponding to cystamine adducted Val-AEC-Thr **18** and Val-AEC-Dhb **19** as well as cysteamine modified albopeptide **20** (Figure 5B and Figure S28D-F).

To finally confirm the functions of individual domains of AlbB, we purified a series of truncated AlbB proteins, AlbB-A_2-T_2, AlbB-A_2-T_2-C_2, AlbB-A_2-T_2-C_2-A_3-T_3 and AlbB-A_2-T_2-C_2-A_3-T_3-C_3 (Figure S29). In an assay with *holo*-AlbA and AlbB-A_2-T_2 in presence of cysteamine, only **16** intermediate was observed, confirming that A_2 domain indeed activates Ser residue and the C_1 domain in AlbA is responsible for the first condensation between L-Val and L-Ser (Figure S30). Incubation of *holo*-AlbB-A_2-T_2-C_2 with AlbA allowed the identification of both **16** and **17** (Figure S30), consolidating that C_2 catalyses the dehydration reaction of Ser residue. In the case of the assay of *holo*-AlbA with AlbB-A_2-T_2-C_2-A_3-T_3, we observed the presence of **18** in our LC-MS analysis (Figure S30), indicating that A_3 domain activates L-Thr and C_2 domain is responsible for the condensation of activated Thr residue with growing Val-Dha thioester to generate **13**. Finally, incubation of *holo*-AlbA and AlbB-A_2-T_2-C_2-A_3-T_3-C_3 provided the evidence of the presence of **19**-**20** (Figure S30), confirming that C_3 domain catalyses the dehydration of Thr residue and the hydrolytic offload of the mature albopeptide. The last T_4 domain plays no essential role in the **6** biosynthesis.

3 Discussion

The presence of Dha and Dhb residues can have pronounced effects on the peptides, such as enhanced peptide-receptor affinity, increased stabilities to degradative enzymes and sometimes good enzyme inhibitors by serving as electrophiles in nucleophilic addition reactions[32]. The contiguous Dha-(*E*)-Dhb residue in albopeptide **6** is structurally unique. Based on our *in vitro* pathway reconstitution and biochemical assays together with *in silico* analysis and previous knowledge[25, 27, 35], we propose the biosynthetic pathway for albopeptide **6** as depicted in Figure 5A. L-Val, L-Ser and L-Thr are activated and loaded onto T domains by AlbA and AlbB, respectively. The thiolated Ser is condensed with Val by C_1 domain,

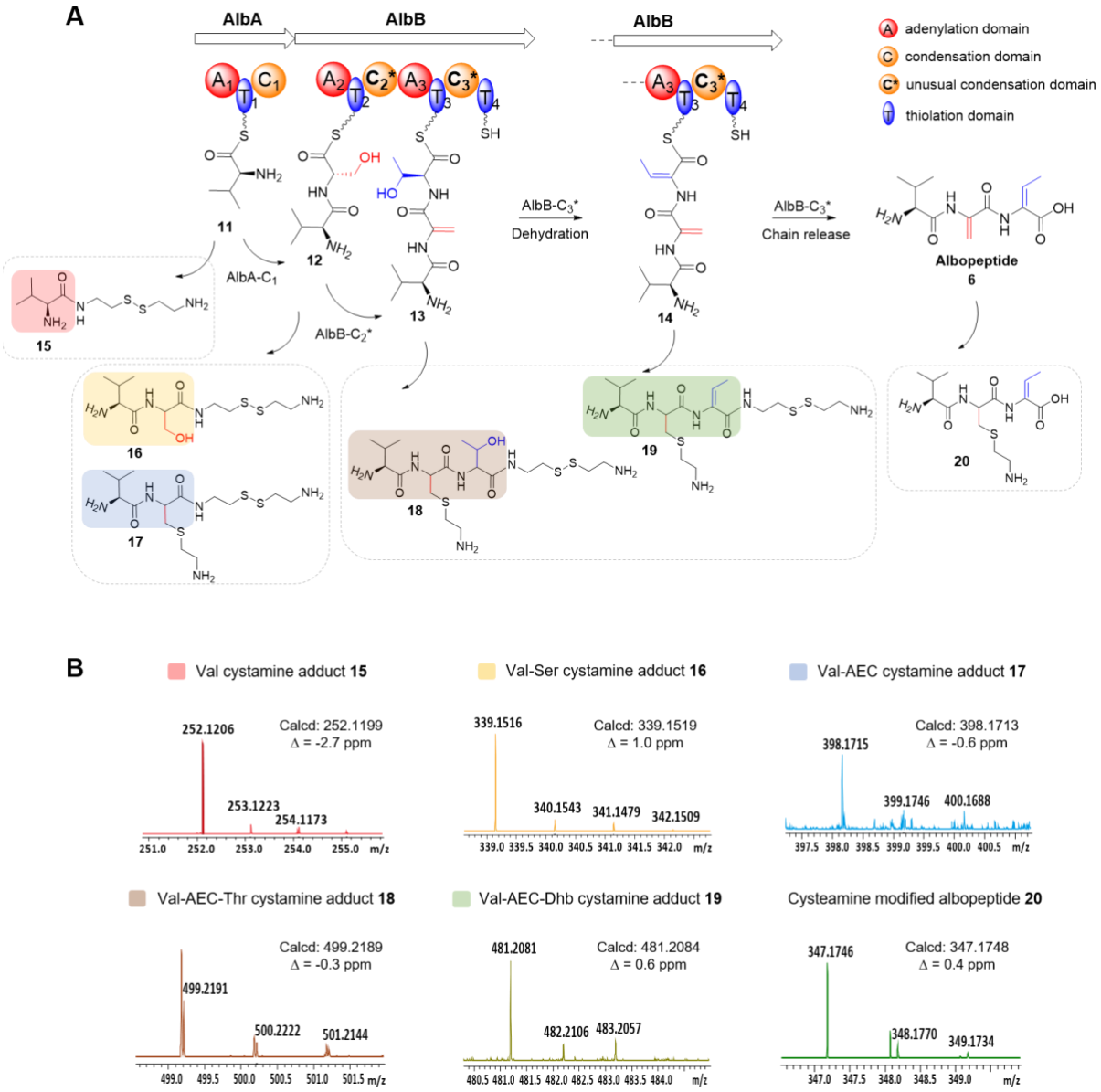


Figure 5. **A**. Overview of NRP assembly of **6** and *in vitro* capture of putative biosynthetic intermediates via cysteamine. **B**. Mass spectra of intermediates captured by cysteamine in the reconstitution reactions.

followed by the dehydration, catalysed by C_2, to form the dipeptide Val-Dha thioester **12**.

Subsequent condensation with the thiolated Thr yields the Val-Dha-Thr thioester **13**, followed by the second dehydration event on Thr residue, by C_3, to produce the last intermediate, Val-Dha-(*E*)-Dhb thioester **14**. Finally, C_3 catalyses the hydrolysis of the T_3 bound intermediate to afford **6** (Figure 5A). Among four chain release mechanisms have been discovered thus far in the NRPS assembly lines[33–35], terminal C-like domains (C_T) have been found to be responsible for the formation of macrolactone[36,37], macrolactam[38], amides[39,40] and carboxylic acids[41]. Phylogenetic analysis of selected C_T domains and C_3 showed that C_3 is clearly diverged from most of C_T domains[30-34] but is closely related to Crok-C_T in the biosynthesis of crocacin catalysing the hydrolysis of the fully elongated intermediate[41] (Figure S31). Interestingly, we also observed the presence of the ion corresponding to cysteamine adduct albopeptide, Val-AEC-Dhb **20** (Figure 5A), demonstrating its reactivity of **6** towards the Michael donors.

Taken together, our biochemical analyses provided direct evidence of the accurate timing of dehydration events on Ser and Thr residues in this AlbB NRPS, where the formation of Dha occurs following the first condensation and the formation of Dhb occurs following the second condensation prior hydrolysis to yield **6**.

The identification of the NRPS pathway of **6** enabled the use of comparative genomic approaches to identify other BGCs for the production of short dehydrated **6**-like metabolites. Blast search in NCBI using AlbB as the sequence query allowed identification of hundreds of AlbB-like open reading frames (ORFs). Phylogenetic analysis of these ORFs suggested that one clade is closely related to AlbB (Figure S32). Using this guidance to analyse the genes around these *orf*s in NCBI, we identified a few BGCs from Gram-positive and Gram-negative bacteria (Figure 6; Table S4), which are potential producers of **6**-like metabolites. Interestingly, these BGCs can be grouped into three subfamilies (Figure 6). While the first group (top box, Figure 6) is likely to direct the production of metabolites containing the Val-Dha-Dhb scaffold, the second one (middle box, Figure 6) is likely to produce metabolites containing a hydrophilic moiety, i.e. Glu or Asp residue, instead of Val residue in **6** in the N-terminal. Such examples can be found in *Vibro parahaemolyticus*, the leading causal agent of human acute gastroenteritis[42] and *Dickerya fangzhongdai*, plant pathogens that mostly cause soft-rot diseases[43]. The products of the third subgroup (bottom box, Figure 6) are unclear due to the limited information of A domains. The common feature of these multidomain NRPSs is that they all contain C* domains which are likely to possess dual functions, condensation and dehydration. Taken together, these BGCs are likely to direct the production of a previously neglected group of short dehydrated peptidyl NPs.

Many NRPs containing dhAA residues or moieties likely derived from dhAA have been discovered thus far, only few of which have been carried out to examine the formation of dhAA

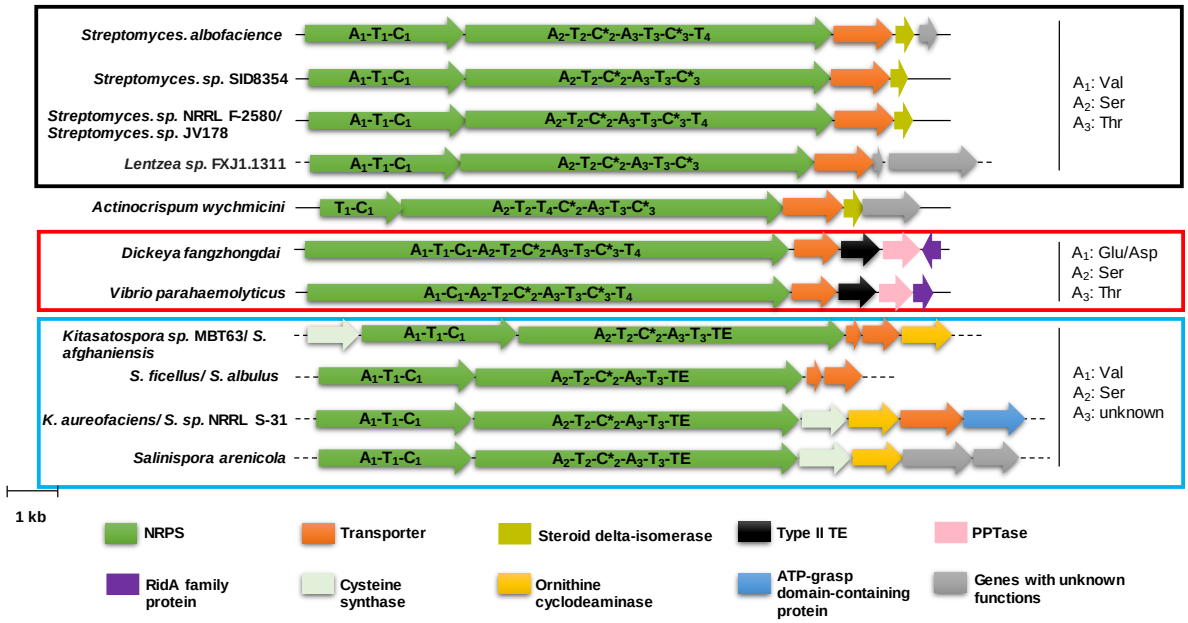


Figure 6. Conserved genomic context of representatives of a new group of *alb*-like BGCs identified in bacterial strains, suggesting that this previously unnoticed group of short dehydrated NRPS peptides is widely spread across the bacterial kingdom.

residues in these metabolites. Recently NocB-M5C in the nocardicin **3** biosynthesis has been shown to be responsible for the generation of the cryptic transient Dha species during the course of the β-lactam formation [7,14]. Our phylogenetic analysis suggested that NocB-M5C belongs to the clade of normal $^{D}C_{L}$ domains[23] that catalyse the condensation reaction with D-amino acids (Figure S33).Unlike NocB-M5C, both C_2 and C_3 form a clade with other C* domains (Figures S33). This group of C* domains possesses a typical HHxxxDG catalytic motif emblematic of conventional C domains and lacks the third His residue in case of nocardicin[14], suggesting that these C* domains are evolutionarily diverged from NocB-M5C. Structural modelling of C_2 and C_3 in Phyre2 Server[44] suggested that the overall folding of these two C* domains share good similarity (100% confidence, 23% identity) to one of the C domain (Txo2-C_2) in the biosynthesis of the potent antibiotic, teixobactin, which catalyses the condensation reaction between the growing peptidyl thioester and L-Ser[45]. However, it was noticed that one loop (A382-T389 in the case of C_2) in both predicted structures of C_2 and C_3 appear to be in the closer proximity toward the catalytic centres of both enzymes, compared to Txo2-C_2 (Figure S34 and S35). Protein MEME analysis[46] suggested that one Glu residue (E387 in the case of C_2) is highly conserved among these C* domains, which is absent in other types of C domains, implying that this residue and its surrounding may involve the dehydration reaction (Figure S36). A combination of biophysics and site-directed mutagenesis will shed

light on the dehydration mechanism of AlbB-C_2 and C_3 domains, which requires further investigation in our laboratory.

NPs have been considered privileged structures in drug discovery. The molecular scaffolds (or pharmacophores) of NPs provide biologically relevant chemical space which are inspirations for modern antimicrobial drug discovery and development[47]. **6** displays potent activity against a VRE isolate. VRE is one of ESKAPE pathogens that acquires resistance against virtually all clinically used antibiotics[48],[49] and is assigned as a high priority pathogen by WHO on its global priority list of antibiotic-resistant bacteria[50]. Therefore, there is a growing interest in discovery of novel antibacterial agents against VRE. In this respect, **6** may offer alternative option for the future lead development against VRE. The mode of action of **6** against VRE will be the focus of our future studies.

Conclusion

We have discovered a new short antimicrobial peptide, albopeptide **6**, in the culture broth of *S. albofacience*. HR-MS and comprehensive NMR analyses, together with chemical synthesis of isomers, confirmed that the structure of compound **6** as L-Val-Dha-(*E*)-Dhb. Albopeptide displays potent narrow-spectrum antibacterial activity against a VRE hospital isolate with an MIC_{50} value of 2.98 μM, potentially offering a new lead to combat VRE. *In vitro* pathway reconstitution and biochemical assays combined with genomic analysis determine that **6** originates from an unusual NRPS assembly line. This work provides evidence on the existence of a new group of short dehydrated NRPS peptides and exemplifies the importance of noncanonical NRPS biosynthetic pathways in bacterial genomes in the search for new leads.

Acknowledgement: QF and HD are grateful to the University of Aberdeen Elphinstone Scholarship and Scottish Funding Council/ScotCHEM (PEER/PERCE) for financial support. HD, ZL and SW thank the financial supports of Biotechnology and Biological Sciences Research Council UK (BBSRC, BB/P00380X/1) and the Royal Society-NSFC Newton Mobility Grant Award (IEC\NSFC\170617 to HD). HD, SAM and CP thank Business Interaction Vouchers (BIV009) from BBSRC funded Natural Products discovery and bioengineering Network (NPRONET). Y.G. thanks NSFC oversea scholarship, Natural Science Foundation of Jiangsu Province (BK20170450), and the Open Research fund of Jiangsu Key Laboratory of Marine Biotechnology (HS2017003).

Author Contributions: S.W., Q.F., Z.L, Y.G., J.H.A., R.E., L.T., H.D. formal analysis and investigation. H.D., S.W., S.L., C.P., funding acquisition and methodology. S.W., Q.F., H.D. writing of original draft. H.D., L.T., R.E., C.P. review and editing. S.W., H.D. supervision and project administration.

Conflicts of Interest: The authors declare no conflict of interest.

Keywords: natural product discovery • peptide synthesis • nonribosomal peptide synthetases • dehydroamino acids • *in vitro* pathway reconstitution

References:

[1] D. Siodłak, *Amino Acids* **2015**, *47*, 1–17.

[2] J. W. Bogart, A. A. Bowers, *Org. Biomol. Chem.* **2019**, *17*, 3653–3669.

[3] T. Kuranaga, Y. Sesoko, M. Inoue, *Nat. Prod. Rep.* **2014**, *31*, 514–532.

[4] G. A. Hudson, D. A. Mitchell, *Curr. Opin. Microbiol.* **2018**, *45*, 61–69.

[5] M. S. C. Pedras, J. L. Taylor, T. T. Nakashima, *J. Org. Chem.* **1993**, *58*, 4778–4780.

[6] N. Bouaicha, C. O. Miles, D. G. Beach, Z. Labidi, A. Djabri, N. Y. Benayache, T. Nguyen-Quang, *Toxins (Basel).* **2019**, *11*, 714.

[7] D. H. Long, C. A. Townsend, *Biochemistry* **2018**, *57*, 3353–3358.

[8] J. B. Patteson, A. R. Lescallette, B. Li, *Org. Lett.* **2019**, *21*, 4955–4959.

[9] J. B. Patteson, Z. D. Dunn, B. Li, *Angew. Chemie - Int. Ed.* **2018**, *57*, 6780–6785.

[10] T. Sano, K. A. Beattie, G. A. Codd, K. Kaya, *J. Nat. Prod.* **1998**, *61*, 851–853.

[11] T. Sano, K. Kaya, *Tetrahedron* **1998**, *54*, 463–470.

[12] H. A. Grab, V. C. Kirsch, S. A. Sieber, T. Bach, *Angew. Chemie - Int. Ed.* **2020**, 1–6.

[13] D. E. Ward, A. Vázquez, M. S. C. Pedras, *J. Org. Chem.* **1999**, *64*, 1657–1666.

[14] N. M. Gaudelli, D. H. Long, C. A. Townsend, *Nature* **2015**, *520*, 383–387.

[15] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Crüsemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderón, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C. C. Liaw, Y. L. Yang, H. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. P. Boya, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodríguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P. M. Allard, P. Phapale, L. F. Nothias, T. Alexandrov, M. Litaudon, J. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D. T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Müller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. Palsson, K. Pogliano, R. G. Linington, M. Gutiérrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, *Nat. Biotechnol.* **2016**, *34*, 828–837.

[16] H. Laatsch, *AntiBase, a Database for Rapid Dereplication and Structure Determination of Microbial Natural Products.*, John Wiley & Sons, Inc.,Hoboken, **2010**.

[17] J. A. Van Santen, G. Jacob, A. L. Singh, V. Aniebok, M. J. Balunas, D. Bunsko, F. C. Neto, L. Castaño-Espriu, C. Chang, T. N. Clark, J. L. Cleary Little, D. A. Delgadillo, P. C. Dorrestein, K. R. Duncan, J. M. Egan, M. M. Galey, F. P. J. Haeckl, A. Hua, A. H. Hughes, D. Iskakova, A. Khadilkar, J. H. Lee, S. Lee, N. Legrow, D. Y. Liu, J. M. Macho, C. S. McCaughey, M. H. Medema, R. P. Neupane, T. J. O'Donnell, J. S. Paula, L. M. Sanchez, A. F. Shaikh, S. Soldatou, B. R. Terlouw, T. A. Tran, M. Valentine, J. J. J. Van Der Hooft, D. A. Vo, M. Wang, D. Wilson, K. E. Zink, R. G. Linington, *ACS Cent. Sci.* **2019**, *5*, 1824–1833.

[18] E. Butler, L. Florentino, D. Cornut, G. Gomez-Campillos, H. Liu, A. C. Regan, E. J. Thomas, *Org. Biomol. Chem.* **2018**, *16*, 6935–6960.

[19] H. Sai, T. Ogiku, H. Ohmizu, *Tetrahedron* **2007**, *63*, 10345–10353.

[20] T. Wagner, B. Joshi, J. Janice, F. Askarian, N. Škalko-Basnet, O. C. Hagestad, A. Mekhlif, S. N. Wai, K. Hegstad, M. Johannessen, *J. Proteomics* **2018**, *187*, 28–38.

[21] M. A. Ortega, W. A. Van Der Donk, *Cell Chem. Biol.* **2016**, *23*, 31–44.

[22] K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema, T. Weber, *Nucleic Acids Res.* **2019**, *47*, W81–W87.

[23] N. Ziemert, S. Podell, K. Penn, J. H. Badger, E. Allen, P. R. Jensen, *PLoS One* **2012**, *7*, 1–9.

[24] S. Huang, J. Tabudravu, S. S. Elsayed, J. Travert, D. Peace, M. H. Tong, K. Kyeremeh, S. M. Kelly, L. Trembleau, R. Ebel, M. Jaspars, Y. Yu, H. Deng, *Angew. Chemie - Int. Ed.* **2015**, *54*, 12697–12701.

[25] J. Mareš, J. H. Jek, P. Urajová, J. Kopecký, P. Hrouzek, *PLoS One* **2014**, *9*, 1–11.

[26] Y. Q. Cheng, M. Yang, A. M. Matter, *Appl. Environ. Microbiol.* **2007**, *73*, 3460–3469.

[27] B. Shen, L. Du, C. Sanchez, D. J. Edwards, M. Chen, J. M. Murrell, *J. Nat. Prod.* **2002**, *65*, 422–431.

[28] J. Vestola, T. K. Shishido, J. Jokela, D. P. Fewer, O. Aitio, P. Permi, M. Wahlsten, H. Wang, L. Rouhiainen, K. Sivonen, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 1909–1917.

[29] M. C. Moffitt, B. A. Neilan, *Appl. Environ. Microbiol.* **2004**, *70*, 6353–6362.

[30] D. Tillett, E. Dittmann, M. Erhard, H. Von Döhren, T. Börner, B. A. Neilan, *Chem. Biol.* **2000**, *7*, 753–764.

[31] H. Wang, W. A. Van Der Donk, *ACS Chem. Biol.* **2012**, *7*, 1529–1535.

[32] J. M. Humphrey, A. R. Chamberlin, *Chem. Rev.* **1997**, *97*, 2243–2266.

[33] L. Du, L. Lou, *Nat. Prod. Rep.* **2010**, *27*, 255–278.

[34] T. Kuranaga, K. Matsuda, A. Sano, M. Kobayashi, A. Ninomiya, K. Takada, S. Matsunaga, T. Wakimoto, *Angew. Chemie - Int. Ed.* **2018**, *57*, 9447–9451.

[35] Y. Zhou, X. Lin, C. Xu, Y. Shen, S.-P. Wang, H. Liao, L. Li, H. Deng, H.-W. Lin, *Cell Chem. Biol.* **2019**, *26*, 737–744.

[36] T. Schwecke, J. F. Aparicio, I. Molnár, A. König, L. E. Khaw, S. F. Haydock, M. Oliynyk, P. Caffrey, J. Cortés, J. B. Lester, G. A. Böhm, J. Staunton, P. F. Leadlay,

Proc. Natl. Acad. Sci. U. S. A. **1995**, *92*, 7839–7843.

[37] G. J. Gatto, S. M. McLoughlin, N. L. Kelleher, C. T. Walsh, *Biochemistry* **2005**, *44*, 5993–6002.

[38] X. Gao, S. W. Haynes, B. D. Ames, P. Wang, L. P. Vien, C. T. Walsh, Y. Tang, *Nat. Chem. Biol.* **2012**, *8*, 823.

[39] C. G. Marshall, N. J. Hillson, C. T. Walsh, *Biochemistry* **2002**, *41*, 244–250.

[40] K. Ishida, G. Christiansen, W. Y. Yoshida, R. Kurmayer, M. Welker, N. Valls, J. Bonjoch, C. Hertweck, T. Börner, T. Hemscheidt, E. Dittmann, *Chem. Biol.* **2007**, *14*, 565–576.

[41] S. Müller, S. Rachid, T. Hoffmann, F. Surup, C. Volz, N. Zaburannyi, R. Müller, *Chem. Biol.* **2014**, *21*, 855–865.

[42] V. Letchumanan, K. G. Chan, L. H. Lee, *Front. Microbiol.* **2014**, *5*, 1–13.

[43] Alič, F. Van Gijsegem, J. Pédron, M. Ravnikar, T. Dreo, *Plant Pathol.* **2018**, *67*, 1612–1620.

[44] L. A. Kelley, S. Mezulis, C. M. Yates, M. N. Wass, M. J. E. Sternberg, *Nat. Protoc.* **2015**, *10*, 845–858.

[45] K. Tan, M. Zhou, R. P. Jedrzejczak, R. Wu, R. A. Higuera, D. Borek, G. Babnigg, A. Joachimiak, *Curr. Res. Struct. Biol.* **2020**, *2*, 14–24.

[46] T. L. Bailey, M. Boden, F. A. Buske, M. Frith, C. E. Grant, L. Clementi, J. Ren, W. W. Li, W. S. Noble, *Nucleic Acids Res.* **2009**, *37*, W202–W208.

[47] A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, *Nat. Rev. Drug Discov.* **2015**, *14*, 111–129.

[48] C. A. Arias, B. E. Murray, *Nat. Rev. Microbiol.* **2012**, *10*, 266–278.

[49] F. Lebreton, W. van Schaik, A. M. McGuire, P. Godfrey, A. Griggs, V. Mazumdar, J. Corander, L. Cheng, S. Saif, S. Young, Q. Zeng, J. Wortman, B. Birren, R. J. L. Willems, A. M. Earl, M. S. Gilmore, *MBio* **2013**, *4*, DOI 10.1128/mBio.00534-13.

[50] E. Tacconelli, E. Carrara, A. Savoldi, S. Harbarth, M. Mendelson, D. L. Monnet, C. Pulcini, G. Kahlmeter, J. Kluytmans, Y. Carmeli, M. Ouellette, K. Outterson, J. Patel, M. Cavaleri, E. M. Cox, C. R. Houchens, M. L. Grayson, P. Hansen, N. Singh, U. Theuretzbacher, N. Magrini, *Lancet. Infect. Dis.* **2018**, *18*, 318–327.