

Cite this: *Chem. Sci.*, 2021, 12, 7138 All publication charges for this article have been paid for by the Royal Society of Chemistry

Understanding thioamitide biosynthesis using pathway engineering and untargeted metabolomics†

Tom H. Eyles, Natalia M. Vior,  Rodney Lacret  and Andrew W. Truman *

Thiostreptamide S4 is a thioamitide, a family of promising antitumour ribosomally synthesised and post-translationally modified peptides (RiPPs). The thioamitides are one of the most structurally complex RiPP families, yet very few thioamitide biosynthetic steps have been elucidated, even though the biosynthetic gene clusters (BGCs) of multiple thioamitides have been identified. We hypothesised that engineering the thiostreptamide S4 BGC in a heterologous host could provide insights into its biosynthesis when coupled with untargeted metabolomics and targeted mutations of the precursor peptide. Modified BGCs were constructed, and in-depth metabolomics enabled a detailed understanding of the biosynthetic pathway to thiostreptamide S4, including the identification of a protein critical for amino acid dehydration that has homology to HopA1, an effector protein used by a plant pathogen to aid infection. We use this biosynthetic understanding to bioinformatically identify diverse RiPP-like BGCs, paving the way for future RiPP discovery and engineering.

Received 14th December 2020

Accepted 19th April 2021

DOI: 10.1039/d0sc06835g

rsc.li/chemical-science

Introduction

Thioviridamide is an apoptosis-inducing compound that was isolated from *Streptomyces olivoviridis* during a screen for cytotoxic compounds¹ and represents the founding member of the thioamitides, a structurally complex family of ribosomally synthesised and post-translationally modified peptides² (RiPPs). RiPPs derive from a ribosomally synthesised precursor peptide that is modified by a series of tailoring enzymes encoded in a biosynthetic gene cluster (BGC). The discovery of the thioviridamide BGC initiated the genomics-led discovery of other thioamitides, including thioholgamide³ (also known as neo-thioviridamide⁴), thioalbamide, thiostreptamide S87, and thiostreptamide S4 (1, Fig. 1A).⁵ It was recently determined that thioamitides inhibit mitochondrial ATP synthase,⁶ which induces mitochondrial dysfunction and triggers apoptosis.⁷

1 features multiple post-translational modifications that are common to most thioamitides but are otherwise rare in nature, including four thioamide bonds, a 2-aminovinyl-3-methylcysteine (AviMeCys) macrocycle,⁸ histidine bis-*N*-methylation, histidine β -hydroxylation, and an *N*-terminal pyruvyl group. 1 also features tyrosine *O*-methylation, which is not found in

other thioamitides.^{5,9} These features are interesting due to their structural and biosynthetic rarity, along with the possible influence they have on bioactivity. For example, histidine bis-*N*-methylation is a modification not found in other RiPPs, and the installation of multiple thioamide bonds is extremely rare.¹⁰ However, there was limited data on thioamitide biosynthesis at the onset of this study.^{11,12} We hypothesised that understanding thiostreptamide S4 biosynthesis would reveal new biosynthetic machinery involved in RiPP maturation, which could inform future pathway engineering and genome mining for new RiPPs with related biosynthetic machinery. Notably, thioamitide biosynthesis is predicted to require lanthipeptide-like Ser/Thr dehydrations, but homologues of the Lan proteins that usually catalyse this step are not encoded in thioamitide BGCs.²

Gene deletions are commonly used to understand natural product biosynthesis, as they can lead to the production of intermediates and therefore reveal the role of a gene, especially as there can be substantial challenges in the *in vitro* reconstitution of complex multi-step pathways from Actinobacteria. However, there are significant difficulties in using gene deletions to understand RiPP biosynthesis.¹³ If the deleted biosynthetic gene produces a protein that acts early in a biosynthetic pathway, then the resulting precursor peptide intermediate is often unstructured and minimally modified. These peptides can be readily digested by endogenous proteases and acetylated endogenously.¹³ Therefore, the identification of intermediates and shunt metabolites can be very challenging, especially if these issues are combined with low productivity in complex media (Fig. S1†).

Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, UK. E-mail: andrew.truman@jic.ac.uk

† Electronic supplementary information (ESI) available: Methods, ESI tables and ESI figures, ESI dataset 1 (XLSX): co-occurrence data for HopA1 proteins, ESI dataset 2 (XLSX): data for networked peptides from RiPPER, ESI dataset 3 (Cytoscape file): networked short peptides and associated data. See DOI: 10.1039/d0sc06835g

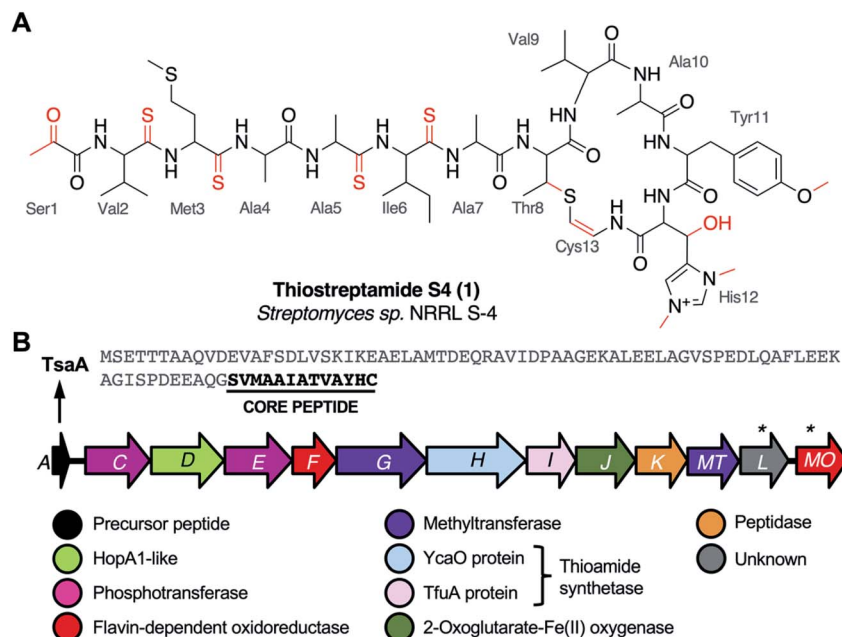


Fig. 1 (A) Thiostreptamide S4 (1) structure with posttranslational modifications highlighted in red and core peptide numbering. (B) Thiostreptamide S4 (*tsa*) BGC and precursor peptide sequence. (*) indicates genes tested in this study that are unlikely to be involved in biosynthesis.

Here, we use a combination of heterologous expression, gene deletions, untargeted metabolomics, and yeast-mediated core peptide engineering to understand the biosynthesis of thiostreptamide S4. This provides a genetic basis for almost every post-translational modification in thioamitide biosynthesis. In addition, the identification of genes associated with Ser/Thr dehydration enables the discovery of diverse RiPP-like BGCs across multiple bacterial taxa.

Results and discussion

Identification of essential biosynthetic genes

We had previously cloned the thiostreptamide S4 (*tsa*) BGC (Fig. 1B, Table S1†) from *Streptomyces* sp. NRRL S-4 using transformation associated recombination (TAR) cloning^{14,15} in yeast to produce plasmid pCAPtsa.⁵ Heterologous expression of pCAPtsa in *Streptomyces coelicolor* M1146¹⁶ produced complete thiostreptamide S4, providing evidence that every gene required to produce thiostreptamide S4 was present. However, the region captured *via* TAR cloning covered a larger region than the predicted BGC (*tsaA*–*tsaMO*; Fig. S2A†). Three genes, *tsa-3*, *tsa-2*, and *tsa-1*, were captured upstream from the predicted BGC and were predicted to encode a DNA polymerase III δ subunit, a hydrolase, and a threonine tRNA respectively. Three genes were captured downstream from the predicted BGC, *tsa+1*, *tsa+2*, and *tsa+3*, predicted to encode a transporter, sulphur-transferase, and serine/threonine protein kinase respectively.

All genes in the predicted *tsa* BGC were independently deleted in pCAPtsa using PCR targeting, replacing most of the target gene with an in-frame 81 bp scar sequence while retaining the original start and stop codons.¹⁷ The up- and downstream regions described above were also deleted. These

mutated plasmids were then expressed in *S. coelicolor* M1146. This revealed that *tsaA*–*tsaJ* and *tsaMT* were required for the biosynthesis of 1, whereas the molecule was still produced in Δ *tsaK*, Δ *tsaL* and Δ *tsaMO* (Fig. S2B; † for simplicity, each *S. coelicolor* M1146 strain harbouring a mutated version of pCAPtsa will herein be referred to by the mutation only). Production of 1 following deletion of *tsaK* was surprising given that this gene is conserved amongst thioamitide BGCs⁵ and encodes a cysteine protease that we predicted was involved in leader peptide removal. It is possible that native peptidases from the heterologous host, *S. coelicolor* M1146, can complement this deletion as there was a small drop in the production of 1 (Fig. S3†). TsaK may only be necessary when 1 is produced in the native host. Similarly, *tsaL*-like genes are conserved amongst almost all thioamitide BGCs, although there is no clear catalytic domain in TsaL (Table S1†). This analysis also demonstrated that *tsaMO* is not required for the biosynthesis of 1.

The deletion of the trio of upstream genes, *tsa-3*, *tsa-2*, and *tsa-1*, caused a significant drop in production (Fig. S3†), whereas production was unaffected by deletion of *tsa+1*, *tsa+2*, and *tsa+3*. With the exception of Δ *tsaA*, each deletion that abolished production was successfully complemented (Fig. S2B†), which ensured there were no unwanted polar effects of each gene deletion. Genetic complementation experiments were carried out by expressing the gene from the strong constitutive promoter *Perme*^{*18} in pIJ10257,¹⁹ which integrates into a ϕ BT1 site in the *S. coelicolor* M1146 genome. This enabled us to determine the correct start codon of each gene (Table S3, Fig. S4†), which revealed that there are two series of genes with overlapping start and stop codons within the BGC, *tsaC*–*G* and

tsaH-MT, with an untranslated 28 bp region between *tsaG* and *tsaH*.

Identifying the genetic basis for macrocycle modifications

The C-terminal macrocycle of **1** features a bis-*N*-methylated and β -hydroxylated histidine, as well as an *O*-methylated tyrosine. There are homologues of TsaG (methyltransferase, pfam06325) and TsaJ (2-oxoglutarate-Fe(II) oxygenase, pfam05721) encoded across almost all thioamitide BGCs, so these were predicted to install the conserved histidine *N*-methylations and β -hydroxylation, respectively. Amongst characterised thioamitide BGCs, TsaMT (methyltransferase, pfam13649) is only encoded in the thiostreptamide S4 BGC, so was predicted to catalyse *O*-methylation, which is unique to characterised thioamitides.⁹ We therefore searched liquid chromatography-mass spectrometry (LC-MS) spectra of wild type (WT), Δ *tsaG*, Δ *tsaJ* and Δ *tsaMT* cultures for masses matching the loss of one to three methyl groups, and/or one hydroxyl group. In total, five of these masses were detected: **1** (m/z 1377.55), **2** (m/z 1363.53, –1 methyl), **3** (m/z 1361.55, –1 hydroxyl), **4** (m/z 1347.54, –1 methyl and –1 hydroxyl), and **5** (m/z 1319.50, –3 methyl and –1 hydroxyl). Tandem MS (MS/MS) fragmentation data confirmed that the mass differences were on the macrocycle and accurate masses are consistent with these proposed structures (Fig. S5†). **1**, **3**, and **4** are seen in the WT, **5** is seen in Δ *tsaG*, **3** and **4** are seen in Δ *tsaJ*, and **2** and **4** are seen in Δ *tsaMT* (Fig. 2). The proposed structures are entirely consistent with detailed analysis of MS data, but have not been fully characterised by NMR.

Deletion of *tsaMT*, which encodes a class I SAM-dependant methyltransferase, resulted in the loss of **1** and **3** (Fig. 2). Instead, **2** was produced, which lacks the tyrosine methylation but is otherwise identical to **1**, therefore confirming that TsaMT is the protein responsible for this modification. Tyrosine *O*-methylation is a rare modification but is observed prior to assembly line biosynthesis of the fungal phytotoxin pyrichalasin H.²⁰ A retro-aldol MS/MS fragmentation that provides a loss of m/z 125.07 is consistent with histidine hydroxylation and bis-*N*-methylation in **2** (Fig. S5†). This shows that the tyrosine methylation is not required for the histidine hydroxylase and methyltransferase to function. Δ *tsaJ* produces **3** and **4** (Fig. 2), which both lack the histidine hydroxylation. This indicates that TsaJ, a non-heme Fe(II) and α -ketoglutarate dependent dioxygenase, is responsible for histidine hydroxylation. The production of **3** shows that histidine hydroxylation is not a prerequisite for tyrosine methylation or histidine bis-*N*-methylation.

Deletion of *tsaG*, which encodes a SAM-dependant methyltransferase, abolished production of **1** and instead led to production of **5**, a version of **1** that lacks all modifications to the macrocycle but is otherwise fully mature (Fig. 2). This means that histidine bis-*N*-methylation is a prerequisite for TsaJ-catalysed histidine hydroxylation and TsaMT-catalysed tyrosine methylation. This indicates that TsaJ and TsaMT are unable to recognise a TsaA-derived substrate without the histidine methylations, which provide a permanent positive charge. The thioamitides are the only RiPPs that feature a bis-*N*-

methylated histidine. Given that the macrocycle is correctly formed in each mutant, these results are consistent with a biosynthetic model where these modifications to the macrocycle are among the final steps in thiostreptamide S4 biosynthesis. TsaG-catalysed histidine methylation occurs first, which is then followed by TsaJ-catalysed histidine hydroxylation and TsaMT-catalysed tyrosine methylation in an undefined order. The role of TsaJ is consistent with a parallel study on the homologue in thioholgamide biosynthesis, Thoj.¹²

To see if other thiostreptamide-like metabolites were produced by these mutants, the characteristic fragmentation pattern of these metabolites was used to search the LC-MS/MS data from the WT, Δ *tsaG*, Δ *tsaJ*, and Δ *tsaMT* strains. The macrocycle is one of the main fragments of **1**–**5**, and so the masses of the different macrocycle fragments seen in **1**–**5** (m/z 687.33, 673.31, 671.33, 657.32, and 629.29, respectively) were used to search all fragmentation events in LC-MS/MS spectra. This enabled the preliminary identification of six new metabolites, **6**–**11** (Fig. S7 and S8†). **6**–**10** are proposed to be versions of **1**–**5** that are hydrolysed between Ala4 and Ala5 (Fig. 3, Fig. S7†), while **11** is predicted to be a version of **5** where the other non-thioamide bond between Ala7 and the macrocycle is hydrolysed (Fig. S8†). These therefore result from hydrolysis of the only non-thioamidated peptide bonds in the tail portion of the molecule, which supports previous evidence that thioamide bonds protect molecules from proteolysis.²¹

Identifying thiostreptamide-related metabolites using untargeted metabolomics

In contrast to the genes encoding macrocycle-modifying enzymes, it was difficult to predict likely pathway products for all other gene deletions (Δ *tsaC*, Δ *tsaD*, Δ *tsaE*, Δ *tsaF*, Δ *tsaH* and Δ *tsaI*), and the targeted analysis described above was unable to identify any macrocycle-containing molecules from these mutants. To address this challenge, MS-based untargeted metabolomics was employed to detect any pathway-associated metabolites. Mutants were compared to *S. coelicolor* M1146-pCAPtsa, Δ *tsaA*, and a medium only control. By filtering out all metabolites present in Δ *tsaA*, we were able to identify multiple metabolites across almost all mutant strains that were likely to derive from TsaA, the precursor peptide (Fig. 3, Table S4†).

There were numerous difficulties in interpreting these data. Identification by MS/MS initially proved difficult due to limited fragmentation, and fragments that were observed could not be accounted for by the simple loss of proteinogenic amino acids. Notably, molecules containing thioamide bonds can undergo fragmentation to lose SH₂; corresponding to a mass loss of 33.9877 Da that does not break the backbone of the molecule.²² This signature loss can be seen very clearly in the fragmentation of **1** (Fig. S6†) and can be used as a tool to identify metabolites that contain thioamides. This indicated that previously unidentified metabolites (m/z 552.23, 503.15, 465.22, 453.16, 392.11, 370.12, 348.14, 330.13 and 259.09) have thioamide bonds in their structure due to this signature fragmentation and are not produced by the Δ *tsaA* mutant (Table S4†). These



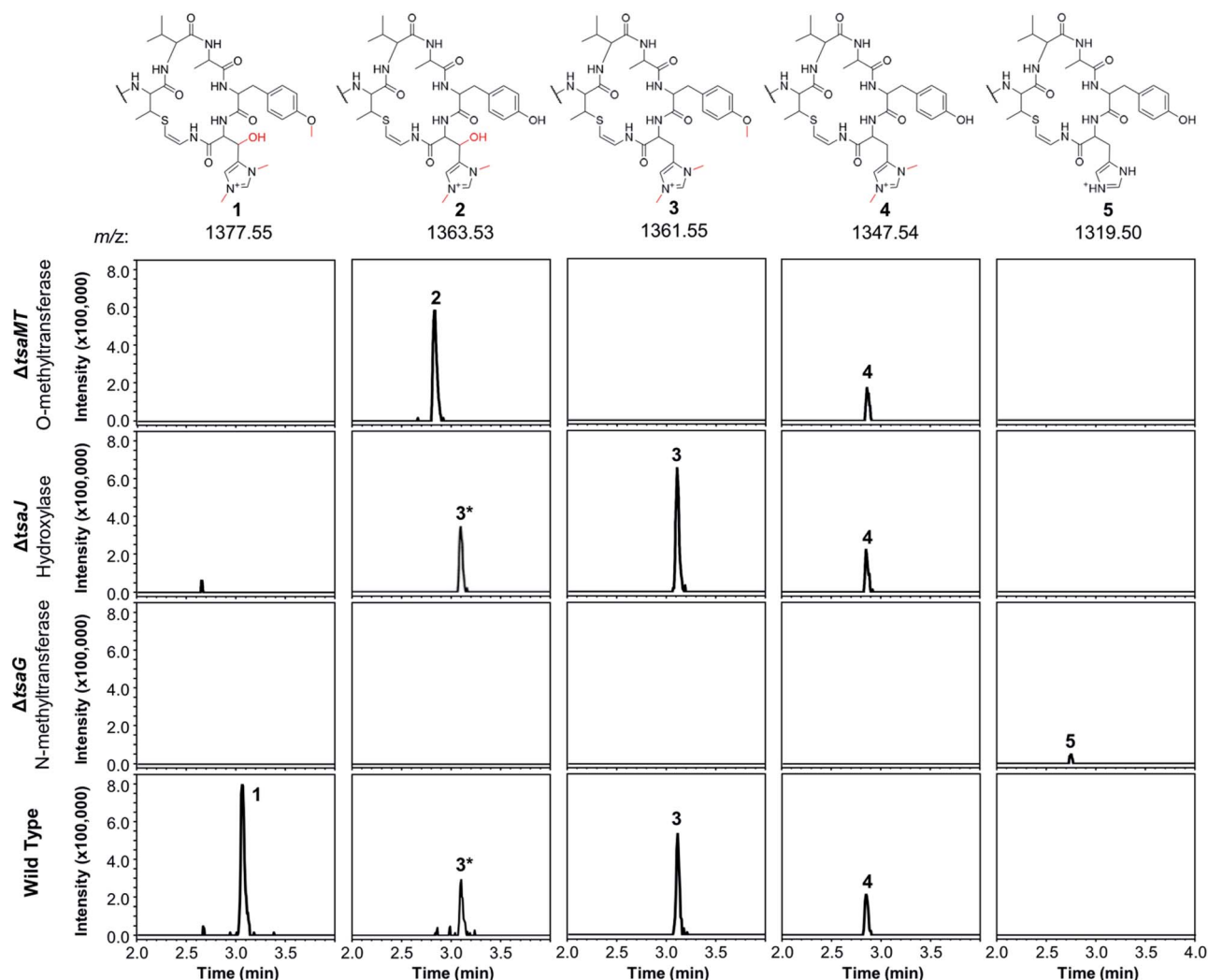


Fig. 2 Extracted ion chromatograms (EICs) normalised for intensity showing the varied methylation and hydroxylation patterns of thio-streptamide-like molecules produced in *S. coelicolor* M1146 expressing the WT, Δ *tsaG*, Δ *tsaJ* and Δ *tsaMT* BGCs. Structures were inferred using detailed MS/MS analysis (Fig. S5†), which are consistent with 2–5 featuring full modifications on the N-terminal linear peptide portion (as in 1). The structure of the macrocycle from each metabolite is shown above the relevant traces; in each case the rest of the molecule is identical to 1. The 3* label indicates the +2 isotope peak of 3.

molecules are hypothesised to be short shunt metabolites that are protected from proteolytic degradation by thioamidation.²¹

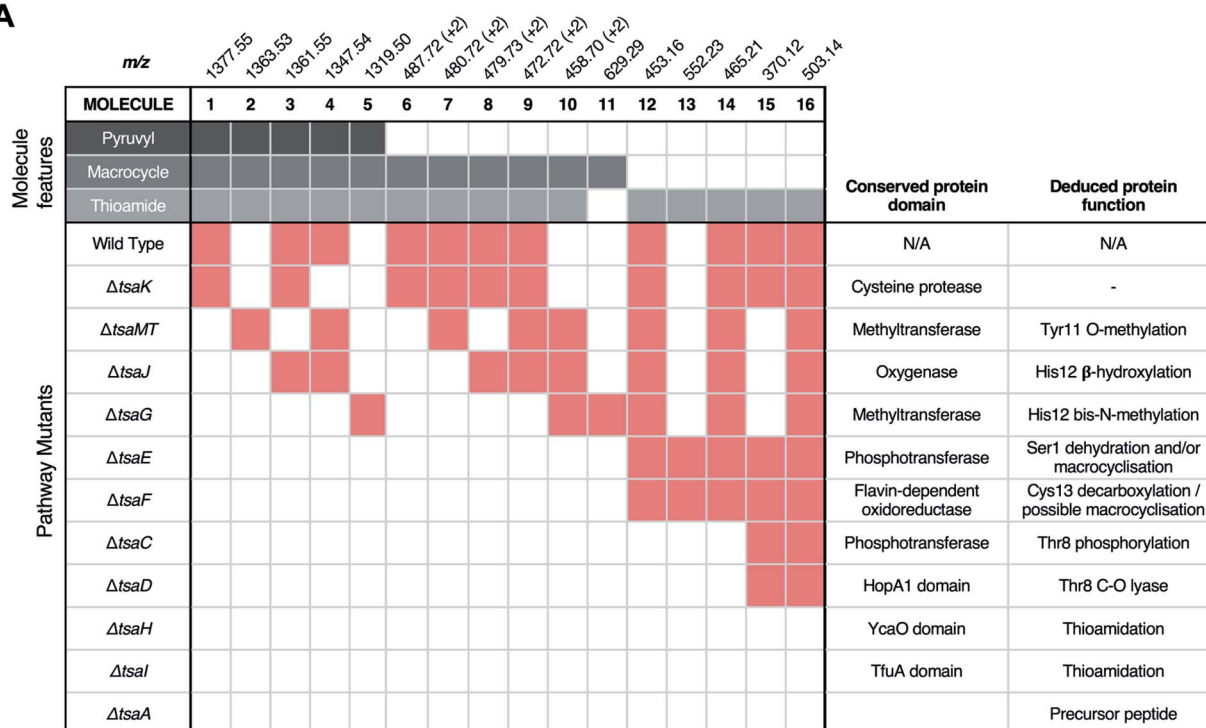
Chemical characterisation of a thioamidated shunt metabolite

To support the preliminary interpretation of thioamidated shunt metabolites, the signature SH_2 loss was used to target metabolites for detailed chemical characterisation. **12** (m/z 453.16) was targeted for purification due to its high production levels and because numerous metabolites featured comparable MS/MS fragmentation (Fig. S9†). **12** was purified from *S. coelicolor* M1146 expressing the Δ *tsaE* BGC, yielding 0.7 mg of pure compound. This was characterised by NMR (^1H , COSY, DEPTQ, HSQCed and HMBC, Fig. S11–S16, Table S5†). ^{13}C shifts of 206.0 and 203.7 ppm were indicative of two thioamides, while a ^{13}C shift of 136.4 ppm was consistent with an olefinic methine.

Two-dimensional experiments established the amino acid connectivity to show that **12** is a modified tetrapeptide, *N*-acetyl-Ala^SIle^SAlaDhb (Dhb = 2,3-dehydrobutyryne; superscript S = thioamidated amino acid) (Fig. 4), whose molecular formula of $\text{C}_{18}\text{H}_{30}\text{N}_4\text{O}_4\text{S}_2$ was consistent with a high-resolution MS peak of m/z 453.1584 ($[\text{M} + \text{Na}]^+$, calc. m/z 453.1601). **12** is therefore a portion of TsaA (Ala5 to Thr8, Fig. 1A) that has undergone expected thioamidation of Ala5–Ile6 and dehydration of Thr8 to Dhb8, but has not been macrocyclised and instead has been hydrolysed at unmodified peptide bonds. These modifications help explain the difficult to interpret MS/MS data, as does an atypical MS/MS fragmentation pattern that occurs for sodiated peptides^{23,24} (Fig. S9B†).

Following the characterisation of **12**, the structures of **13** (m/z 552.23), **14** (m/z 465.21), and **15** (m/z 370.12) could also be proposed to be related acetylated and thioamidated short peptides, based on similar MS/MS fragmentation patterns,

A



B

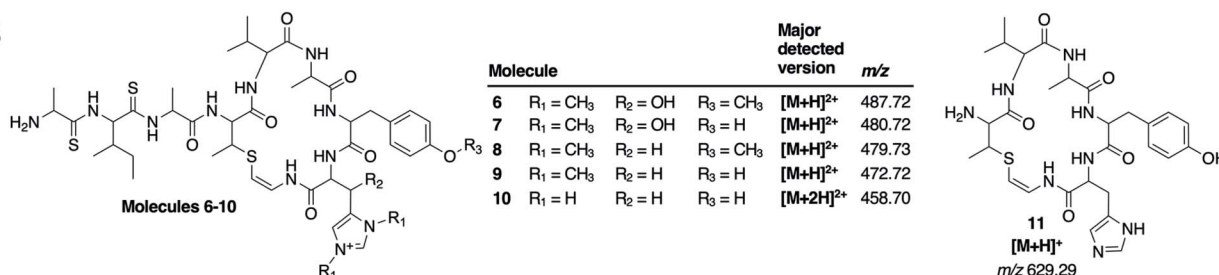


Fig. 3 Untargeted metabolomic analysis of thiostreptamide S4 biosynthesis. (A) Matrix of detected molecules versus mutants. Red shading indicates the production of a molecule in a given mutant. "+2" indicates that the doubly charged m/z is detected. (B) Proposed structures of molecules 6–11 based on MS/MS and accurate mass data (Fig. S7 and S8†). *Permanent charge on bis-methylated histidine means that a single protonation generates a doubly charged molecule. Proposed structures of 13–16 and associated MS data are shown in Fig. S9–S10 and Fig. S17.†

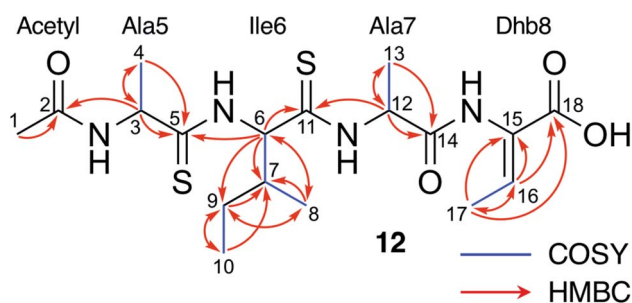


Fig. 4 NMR characterisation of 12 in CD₃OD. See Fig. S11–S16 and Table S5† for NMR assignments.

accurate mass data and predicted thioamidations (Fig. S9A†). This similarly enabled us to propose the structure of 16 (m/z 503.14), a metabolite produced by the WT and all $\Delta tsacF$ mutants (Fig. 3 and Table S4†). 16 is proposed to be *N*-acetyl-

SerVal^SMet^SAla, which we hypothesise derives from the precursor peptide (Ser1 to Ala4, Fig. 1A) that has undergone expected thioamidation of Val2-Met3 (Fig. S10†). Further support for this structure was provided by precursor peptide modifications (S1T and M3I), which led to expected mass shifts to this metabolite (Fig. S10; † see later section for a description of site-directed mutagenesis).

Thioamidation requires a YcaO protein and a TfuA protein

Prior studies have demonstrated that YcaO and TfuA domain proteins are required for thioamidation in archaea^{25,26} and bacteria,^{22,27} and a recent study showed that an archaeal TfuA protein delivers sulphide to its cognate YcaO protein.²⁸ We therefore predicted that YcaO protein TsaH and TfuA protein TsaI would iteratively introduce the four thioamides in 1. Deletion of either *tsaH* or *tsaI* led to the abolition of every detectable metabolite associated with the BGC (Fig. 3). This is in contrast to every other mutant, which were all able to make

thioamidated peptides (Fig. 3). This supports a biosynthetic model where TsaH and TsaI cooperate to catalyse thioamidation. The absence of detectable metabolites is consistent with TsaH/TsaI functioning as essential steps at a very early stage in the pathway, as other modifications that could protect TsaA from proteolysis, such as macrocyclisation, were not detected in these mutants.

Identification of a new amino acid dehydratase

Along with thioamidation, a characteristic feature of thioamides is a C-terminal Avi(Me)Cys macrocycle. In **1**, this is predicted to be generated by the Michael-type addition of oxidatively decarboxylated Cys13 with Dhb8, which is formed by dehydration of Thr8. Another conserved thioamide feature is an N-terminal pyruvyl or lactyl moiety that we previously predicted to be derived from dehydrated Ser1.⁵ In lanthipeptide biosynthesis, the dehydration of Thr to Dhb is catalysed by Lan proteins.²⁹ However, no Lan proteins are encoded in thioamide BGCs. We were unable to detect full-length core peptides featuring Thr8, but we hypothesised that the presence/absence of metabolites **12–16** would help identify genes involved in dehydration and potentially cyclisation. **12–16** were therefore mapped to the metabolomes of mutants of genes that had not yet been functionally annotated (Δ tscA, Δ tscD, Δ tscE, Δ tscF; Fig. 3 and 5).

The structure of **12** (Fig. 4), and the predicted structures of **13–16** (Fig. S17†), provide key information towards the proteins involved in dehydration (Fig. 5). **12–14** are shunt metabolites of an intermediate that lacks the macrocycle but contains the

Dhb8 residue that is required for macrocycle formation. In contrast, **15** can derive from an intermediate that contains an unmodified Thr8; the lack of modification making it susceptible to proteolysis. In all strains containing deletions of any of tscA–F, all detected metabolites lack the macrocycle (**12–16**), which implies that they are involved in steps previous to its formation. Of these mutants, Δ tscA and Δ tscD produce thioamidated **15** and **16** (Fig. 3) but do not produce shunt metabolites containing Dhb8. We hypothesise that these metabolites derive from a modified TsaA that is not yet dehydrated at Thr8 and is therefore more susceptible to proteolysis at that position. This would indicate that TsaC and TsaD cooperate to catalyse dehydration of Thr8 to Dhb8.

TsaC contains an aminoglycoside phosphotransferase (APH)-like domain (pfam01636). APHs are structurally similar to eukaryotic protein kinases³⁰ and it has been shown that some APH enzymes can also phosphorylate serine residues.³¹ We therefore propose that TsaC is responsible for phosphorylation of Thr8, allowing for a subsequent elimination reaction to dehydrate Thr8 (Fig. 4). The role of TsaD in threonine dehydration is currently unclear, although TsaD contains a HopA1 effector family domain (pfam17914). HopA1 itself is a type III effector that aids plant infection by the plant pathogen *Pseudomonas syringae*,^{32,33} although the mechanistic basis for this activity is unknown. Other effectors, such as the OspF family,³⁴ do function as lyases to inactivate protein kinases in the host cell. OspF family effectors are also known as HopAI1-like proteins, but have no sequence homology to the similarly named HopA1-like proteins. The N-terminal lyase domain of the LanL family of lanthionine synthetases has sequence homology

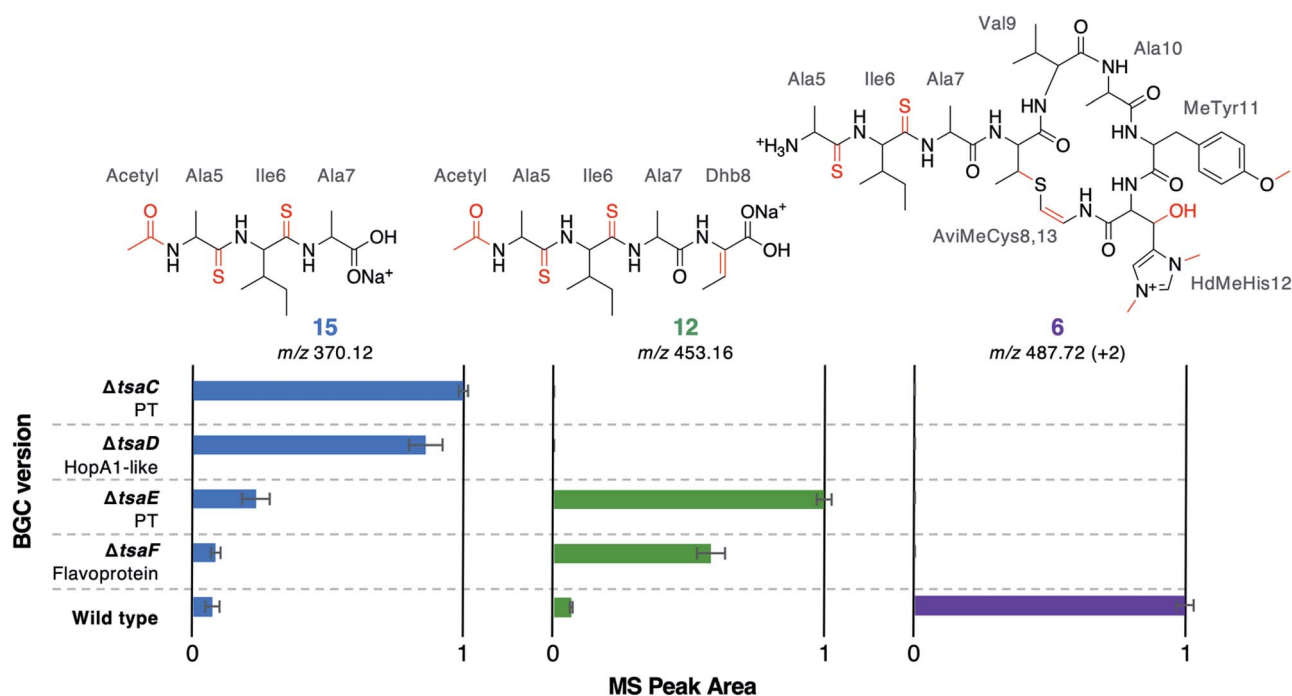


Fig. 5 MS peak areas of selected metabolites produced by mutants (Δ tscA–F) and WT thiostreptamide S4 BGC. Each bar chart is normalised to the highest mass spectral peak area for that metabolite. The error bars represent the standard error of three biological repeats. PT = phosphotransferase; "+2" indicates that the doubly charged m/z is shown.

with OspF proteins.³⁵ TsaD may therefore act as a C–O lyase to catalyse the elimination of a TsaC-installed phosphate group to dehydrate Thr8. Alternatively, TsaD may have a non-catalytic role that is essential for TsaCD-catalysed dehydration, such as precursor peptide binding. Very recently, two RiPPs containing lanthionine cross-links and AviMeCys macrocycles were reported (cacaoidin³⁶ and lexapeptide³⁷), whose BGCs encode homologues of TsaC and TsaD. Intriguingly, our results are in contrast to a very recent report on the biosynthesis of the thioisparsoamide, which indicated that lanthipeptide synthetases encoded outside of the BGC catalyse thioamidite dehydration.³⁸

Macrocyclisation is dependent on TsaE and TsaF

In contrast to $\Delta tsaC$ and $\Delta tsaD$, **12**, **13** and **14** are produced by both $\Delta tsaE$ and $\Delta tsaF$ (Fig. 3 and 5). Therefore, neither gene is required for the dehydration of Thr8 to Dhb8. However, no macrocyclised molecules (**1–11**) are produced by either $\Delta tsaE$ or $\Delta tsaF$. TsaF is predicted to be a flavoprotein (pfam02441) that has similarity to flavin-dependent cysteine decarboxylases that catalyse the formation of AviCys,³⁹ AviMeCys, and avionin macrocycles⁴⁰ in RiPP biosynthesis. In Avi(Me)Cys-containing natural products this oxidative decarboxylation forms the thioenolate moiety required for Avi(Me)Cys formation.⁴¹ The lack of macrocyclised molecules when *tsaF* is deleted supports the role of TsaF as a cysteine decarboxylase that decarboxylates Cys13. This is consistent with a recent co-expression study using the thioviridamide orthologue, TvaF, which was shown to catalyse oxidative decarboxylation of the thioviridamide precursor peptide.¹¹ In lanthipeptide biosynthesis, a cyclase is required to catalyse cyclisation, where one of the roles of the cyclase is to stabilise the thiolate involved in macrocycle formation.²⁹ The formation of the AviMeCys thioether may be spontaneous, as the enethiolate that results from cysteine decarboxylation has a significantly lower pK_a than the thiol side chain of cysteines.⁴² This could explain the lack of a cyclase homologue encoded in thioamidite BGCs or in other Avi(Me)Cys containing RiPP BGCs, such as the linaridins,⁴³ although studies on the linaridin cypemycin indicate that decarboxylation is not sufficient for cyclisation.^{44,45}

TsaE possesses weak homology to the APH-like phosphotransferase domain (pfam01636) that is also found in TsaC. However, $\Delta tsaE$ has a very similar metabolite profile to $\Delta tsaF$ (Fig. 3 and 5, Table S4†). It is somewhat surprising that the macrocycle cannot form in $\Delta tsaE$, given that Dhb8-containing molecules are produced by this mutant and the cysteine decarboxylase, TsaF, is present. The lack of macrocycle could be explained if TsaE assists with AviMeCys cyclisation. However, it is unclear what role a phosphotransferase could play in cyclisation and there are no *tsaE* homologues encoded in BGCs for other Avi(Me)Cys RiPPs, such as the linaridins. An alternative hypothesis is that TsaE functions as the phosphotransferase involved in Ser1 dehydration to 2,3-dehydroalanine (Dha), which we predicted to be necessary for the formation of the N-terminal pyruvyl group of **1**.

To demonstrate that the pyruvyl group originates from a dehydrated Ser1 instead of a pyruvyl transferase,⁴⁶ a S1T

mutant of TsaA was generated, producing the construct pCAPtsaS1T. *S. coelicolor* M1146-pCAPtsaS1T produced a molecule with m/z 1391.5598 (**17**) that was absent in the WT (Fig. S18†). This mass reflects an extra methyl group compared to **1** and MS/MS fragmentation is consistent with **17** containing an N-terminal 2-oxobutyryl moiety instead of a pyruvyl moiety (Fig. S18†). This therefore confirms that the natural N-terminal pyruvyl group originates from a dehydrated serine. Hydrolytic removal of the leader peptide generates an enamine in equilibrium with an imine that is predicted to spontaneously hydrolyse to the pyruvyl group (Fig. 6E). A Thr1 residue is seen naturally in the predicted core peptides for uncharacterised thioamidites from *Micromonospora eburnea* and *Salinispora pacifica*.⁵ The amino acid origin of the pyruvyl group is consistent with previous co-expression studies on epilancin 15X⁴⁷ and polytheonamide dehydratases.⁴⁸ The serine origin of the pyruvyl group means that it is plausible that TsaC, TsaD and/or TsaE are involved in Ser1 dehydration, given that **16** is proposed to contain an unmodified Ser1 residue and is produced by each of these mutants (Fig. 3 and S10†), but further experimental work is required to confirm the serine dehydratase.

Proposed biosynthetic pathway to **1**

The metabolomic data from the gene deletions enables a plausible biosynthetic pathway to be proposed (Fig. 6A). The absence of any detectable metabolites in $\Delta tsaH$ and $\Delta tsaI$, as well as the presence of thioamidated peptides in all other mutants, indicates that the first step is the thioamidation of the TsaA core peptide by TsaH (YcaO-like) and TsaI (TfuA-like) (Fig. 6B). The absence of metabolites containing Dhb8 (or macrocycles) in $\Delta tsaC$ and $\Delta tsaD$ is consistent with TsaC/TsaD-catalysed dehydration of Thr8 to Dhb8. TsaC is a phosphotransferase, which would be consistent with a phosphorylation and elimination mechanism typical of class II, III, and IV lanthipeptide dehydratases (Fig. 6C). Our data suggest that HopA1-like TsaD is either a lyase responsible for phosphate elimination, or a non-catalytic partner protein. At some stage following this step, Ser1 is dehydrated, which is predicted to proceed *via* the same mechanism. This may involve TsaE-catalysed phosphorylation, but it also cannot be ruled out that TsaC catalyses this step. The lack of macrocyclic molecules detected in $\Delta tsaE$ indicates that this precedes macrocycle formation, but could alternatively indicate a cryptic role for TsaE in macrocyclisation.

We propose that the AviMeCys macrocycle is then formed, which first involves Cys13 decarboxylation to generate a reactive thioenolate. This is proposed to be catalysed by flavoprotein TsaF, given the absence of cyclised molecules produced in $\Delta tsaF$ and the prior characterisation of the TsaF orthologue from the thioviridamide pathway. Macrocyclisation itself may be non-enzymatic (Fig. 6D), given the lack of an obvious cyclase protein, although the apparent absence of multiple diastereoisomers of **1** suggests stereochemical control during AviMeCys formation.

The next step is bis-N-methylation of His12. Deletion experiments show that TsaG is responsible for this step. Histidine bis-methylation is not present in any other natural product



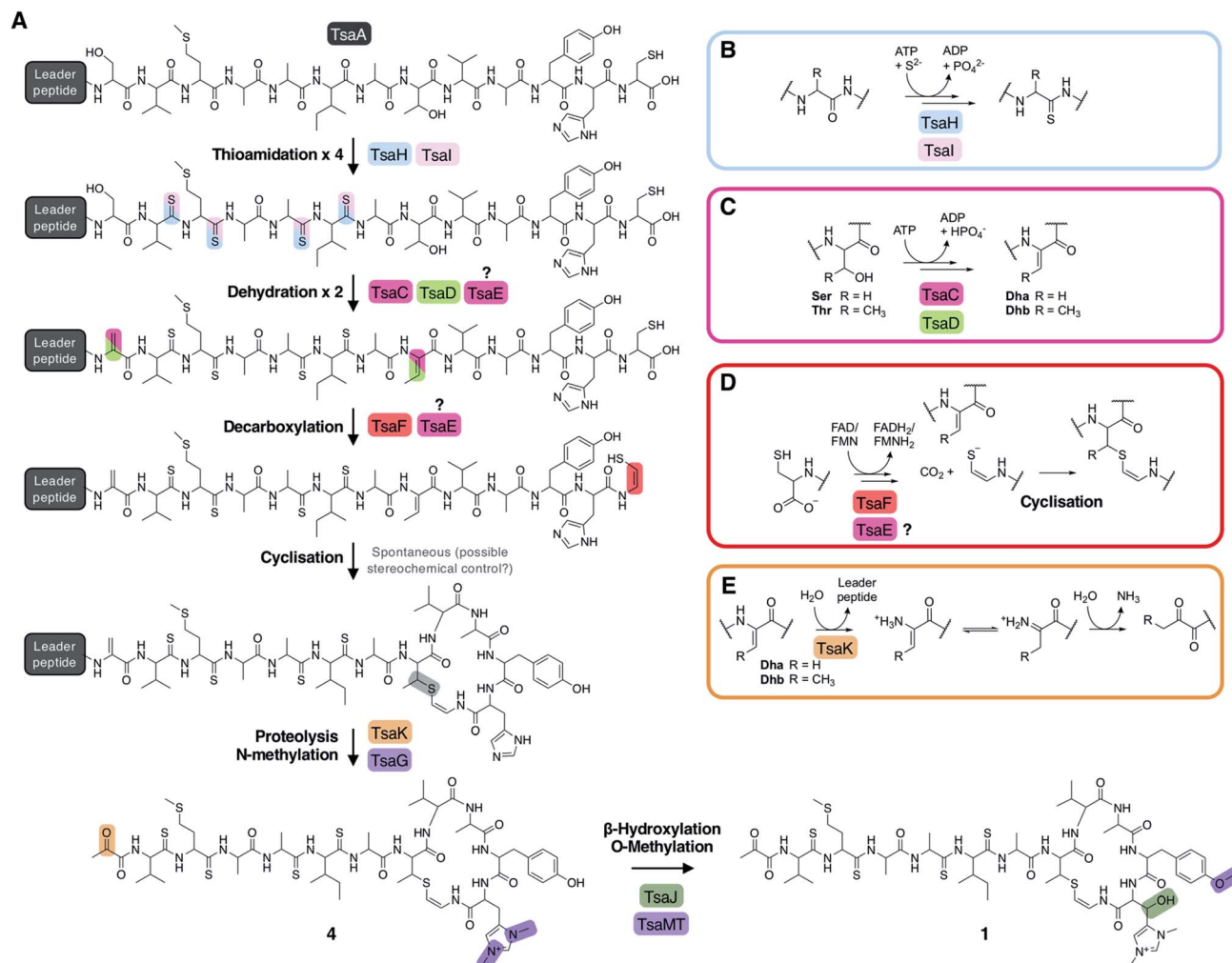


Fig. 6 (A) Proposed biosynthetic pathway to thiostreptamide S4 (**1**). (B) Proposed route to thioamidation based on prior studies of archaeal thioamidation. (C) Proposed route to dehydration of serine ($R = H$) or threonine ($R = CH_3$) via phosphorylation and elimination. (D) Proposed route to cysteine decarboxylation and cyclisation. (E) Proposed production of pyruvyl or 2-oxobutyryl moiety from an N -terminal Ser1 or Thr1 (S1T mutant) following leader peptide proteolysis.

family and provides a positive charge that may be important for biological activity.⁴⁹ Gene deletion experiments show that this methylation acts as a gatekeeper for subsequent modifications: His12 β -hydroxylation and Tyr11 O -methylation, installed by TsaJ and TsaMT respectively. Our data indicate that these proteins preferentially act on substrates containing a bis-methylated histidine. Whilst mature thiostreptamide S4-like molecules detected in this study rarely lack the histidine N -methylations, we could readily detect mature thiostreptamide S4-like molecules lacking the histidine β -hydroxylation and tyrosine O -methylation (Fig. 2). Therefore, these modifications are not a prerequisite for leader peptide cleavage and associated pyruvyl formation, and may happen following leader peptide removal.

There are no clear data to allow the assignment of an enzyme for leader peptide cleavage. This was unexpected, as bioinformatic analysis shows that TsaK is a C1A family cysteine protease and homologues are encoded in other thioamitide

BGCs. This protease family is rare in bacteria, although a C1A family protease catalyses removal of the leader peptide in polypeptide biosynthesis.⁵⁰ It is possible that the small change in production observed when *tsaK* is deleted (Fig. S3†) is because endogenous proteases catalyse hydrolysis of the leader peptide, as in the biosynthesis of many class III lanthipeptides.⁵¹ Following proteolysis, the pyruvyl group is likely to be formed spontaneously from Dha1 (Fig. 6E). This is supported by the production of **17** (featuring an N -terminal 2-oxobutyryl group) when the precursor peptide contains a S1T mutation (Fig. S18†).

Yeast assembly enables site-directed mutagenesis of precursor peptide TsaA

Precursor peptide mutagenesis can be important to probe specific biosynthetic steps or hypotheses, to test the substrate specificity of enzymes, and to generate RiPP libraries. However, it was not possible to complement the precursor peptide Δ *tsaA*



mutant with an intact copy of the *tsaA* gene present in the integrative plasmid pIJ10257, which may be due to insufficient levels of expression from the non-native *PerME** promoter. Therefore, we employed a yeast-mediated recombination strategy⁵² to introduce clean modifications to the precursor peptide in pCAPtsa. Here, the vector was digested using naturally occurring unique restriction enzyme sites near *tsaA* (*Afl*III and *Srf*I) and then reassembled in a single step in yeast using PCR fragments and a single-stranded synthetic oligonucleotide that contains a mutated core peptide sequence (Fig. S19A†).

This strategy was used to generate the S1T mutant that was discussed earlier. To test the tolerance of the biosynthetic enzymes to modifications to the macrocycle amino acids, we made four further mutants of the TsaA core peptide: T8S, Y11V, H12A and H12W (Fig. S19B†). T8S was constructed to assess whether dehydration and macrocyclisation takes place when Thr8 is swapped with a serine residue, which is found in this position in some related precursor peptides.⁵ This led to the production of **18**, which has a mass (calc. m/z 1363.5311, obs. m/z 1363.5261) and MS/MS fragmentation that is consistent with a fully modified derivative of **1** featuring the expected *AviCys* moiety (Fig. S20†). In contrast, the other modifications were not tolerated, as no macrocyclised molecules were detected with the Y11V, H12A and H12W mutants. However, an increase in the production of **12** (Fig. S21†) in each mutant indicated that early stage thioamidation and Thr8 dehydration took place, but either TsaE or TsaF would not function.

A common metabolite detected throughout growth and extraction of thiostreptamide S4 (**1**) is the methionine sulphoxide derivative (**19**; Fig. S22†). Met3 is particularly susceptible to oxidation, which is problematic if this molecule was used in a clinical setting, as the methionine sulphoxide version of a similar molecule, thioholgamide, is around ten times less active than un-oxidised thioholgamide.³ To engineer thiostreptamide S4 into a more stable molecule, a version was made with Met3 swapped for an isoleucine (M3I), which is naturally found at this position in the thioalbamide precursor peptide.⁵ This modification was tolerated and led to the production of **20** (m/z 1359.59, Fig. S23†). As with a site-directed mutagenesis study on thioviridamide,⁵³ these data indicate that precursor peptide mutagenesis represents a viable route to novel thioamitides, although the complexity of these pathways means that there are mutants that are not tolerated by all tailoring enzymes (Fig. S19B†).

Understanding thioalbamide biosynthetic modifications

The *Amycolatopsis alba* thioalbamide BGC contains genes that encode a predicted cytochrome P450 (TaaCYP; pfam00067) and a NAD(P)H-dependent reductase (TaaRed; pfam00106) that are absent from almost all other thioviridamide-like BGCs (Fig. S24†). The function of these genes could not be tested directly in *A. alba* because attempts to genetically manipulate this strain were unsuccessful. Therefore, *taaCYP* and *taaRed* were expressed in *S. coelicolor* M1146-pCAPtsa to test if their activity could be reconstituted on a similar molecule. Thioalbamide contains an *N*-terminal lactyl group, and we

previously predicted that TaaRed catalyses the reduction of the Ser1-derived pyruvyl group to a lactyl group.⁵ This would be analogous to the generation of a lactyl group in epilancin 15X biosynthesis by *ElxO*, a NAD(P)H-dependant reductase.⁵⁴ Co-expression of TaaRed with pCAPtsa generated a thio-streptamide S4 derivative (**21**) that is 2 Da greater than **1**. The accurate mass (m/z 1379.5624) and MS/MS fragmentation data for **21** is consistent with an *N*-terminal lactyl group (Fig. S25†). Our data provide preliminary evidence that TaaRed has broad substrate tolerance, given the significant differences between thioalbamide and **1**.

Thioalbamide has a hydroxylated Phe5 not seen in other characterised thioviridamide-like compounds, so it was hypothesised that TaaCYP is responsible for this hydroxylation. To test this, we used yeast-mediated assembly to generate two new versions of the thiostreptamide S4 BGC with mutated *tsaA* genes: one encoding a core peptide with a containing a phenylalanine at position 5 (A5F), and TsaCoreTaa, where the entire thiostreptamide S4 core peptide was replaced with the thioalbamide core peptide (Fig. S19B†). Unfortunately, no related metabolites could be detected when these clusters were expressed in *S. coelicolor* M1146, meaning that these modifications were not tolerated by the thiostreptamide S4 tailoring enzymes.

Genome mining reveals that the HopA1 and phosphotransferase protein pair are widely found in RiPP-like BGCs

A key step of thiostreptamide biosynthesis is dehydration, which we propose is catalysed by phosphotransferase TsaC and HopA1-like protein TsaD. To determine whether these proteins represent an overlooked signature of RiPP BGCs, we carried out a similarity-based search for TsaD-like proteins in GenBank, which identified 1,340 non-redundant HopA1 domain proteins across multiple bacterial phyla. In 96% of cases, the HopA1 protein is encoded alongside a phosphotransferase (ESI dataset 1 and Fig. S26†), supporting the theory that these are partner proteins that cooperate to catalyse one reaction. We hypothesised that if conserved short peptides were encoded near these proteins then they could represent novel RiPP BGCs. To assess this, we used RiPPER, which we previously developed to identify short peptides encoded near bait proteins²² (ESI dataset 2). Short peptides encoded near HopA1 proteins were then grouped into families using sequence similarity networking⁵⁵ (Fig. S27 and ESI dataset 3) and the associated genomic loci were assessed for co-occurring proteins (Fig. S26†).

These putative BGCs were manually assessed for characteristic features of RiPP BGCs: co-linearity of putative biosynthetic genes and a position at the beginning of biosynthetic genes for the short peptide gene. The thioamitides themselves belong to peptide Family 10. The majority of BGCs belong to Family 1A (Cyanobacteria) and Family 1B (Actinobacteria), which includes the BGCs for the antibiotics cacaoidin and lexa peptide, the first members of the recently described lanthidin RiPP family.^{36,37} Genes *cao7*, *cao9* and *caoD* in the cacaoidin BGC encode a HopA1-like protein, a phosphotransferase and a cysteine



decarboxylase homologous to TsaD, TsaC and TsaF respectively, which suggests that the AviMeCys group found in this molecule is installed following a similar mechanism as in the thioamides. The precursor peptides in this family feature C-termini with highly conserved Thr and Cys residues (Fig. 7A), consistent with the production of diverse AviMeCys containing RiPPs. In parallel with our study, a new RiPP genome mining algorithm, decRiPPter, also identifies the discovery of a similar set

of actinobacterial RiPP BGCs encoding HopA1-like proteins and phosphotransferases, which led to the discovery of pristinin A3.⁵⁶

Family 1A precursor peptides (570 peptides across 183 BGCs) are exclusive to Cyanobacteria and are encoded in partially conserved BGCs (Fig. 7A). These peptides have high conservation of their leader region, which features a conserved double glycine motif that is a common cleavage motif in

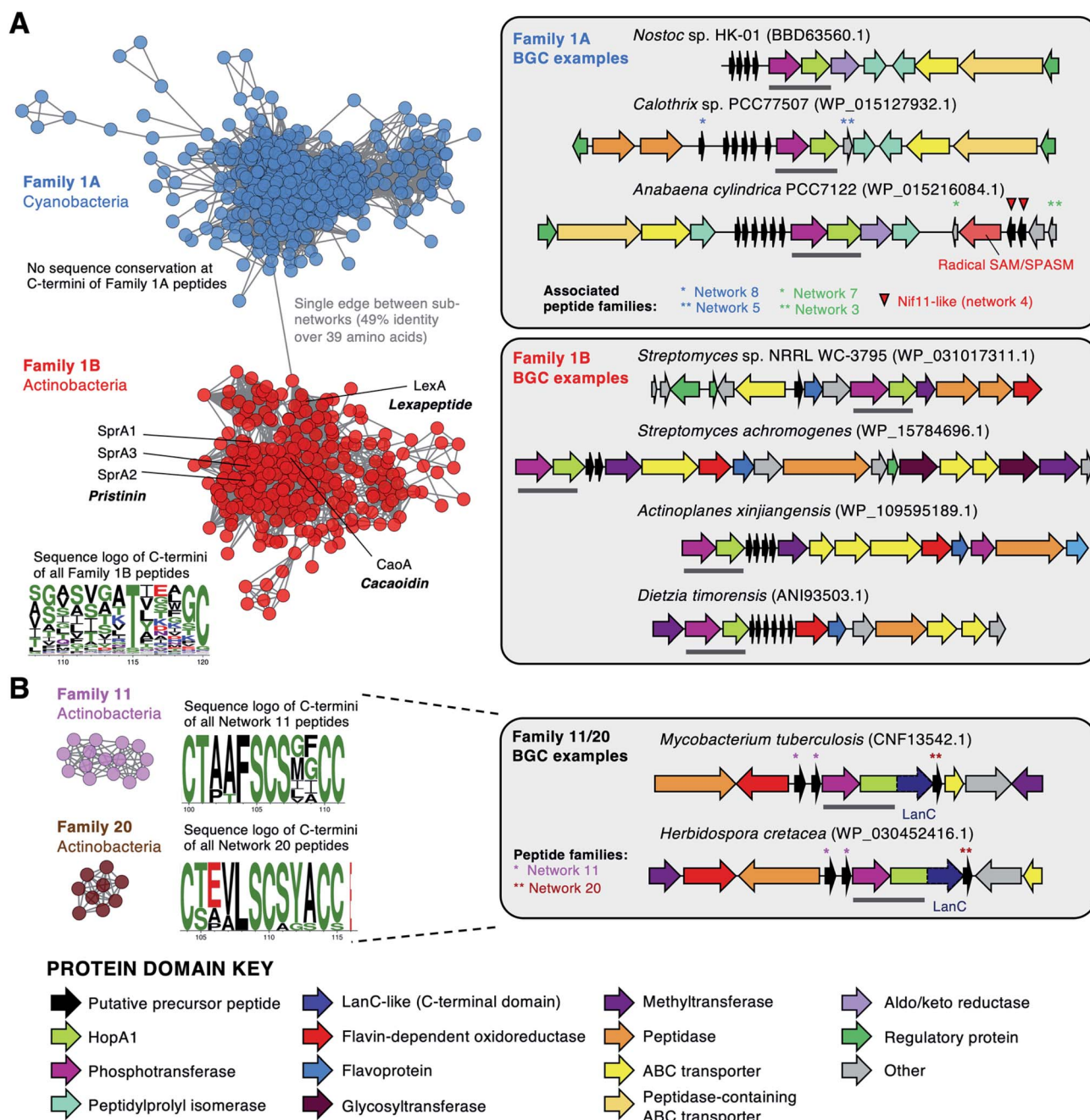


Fig. 7 Selected HopA1-associated precursor peptides and examples of corresponding BGCs. Networks⁵⁷ represent short peptide networking output from RiPPER with a 40% identity cut-off (Fig. S27†). Sequence logos⁵⁸ are shown for selected portions of the C-termini of each family (see Fig. S28 and S29† for full logos). In each BGC, the HopA1/phosphotransferase pair is highlighted with a grey bar and the HopA1-like protein accession is listed. (A) Families 1A and 1B, with precursor peptides of recently identified RiPPs highlighted.^{36,37,56} (B) Family 11/20 peptides, which co-occur in the same HopA1-LanC BGCs. Additional associated peptide networks are highlighted.



lanthipeptides.²⁹ In contrast, their C-terminal regions, which are predicted to correspond to the core peptide regions, are highly variable and do not feature conserved Ser, Thr or Cys residues (Fig. S28†). These BGCs typically encode multiple non-identical precursor peptides, which is common for cyanobacterial RiPP BGCs.³⁹ 20% of HopA1-like proteins are encoded near, or fused to, LanC-like cyclases, such as Family 11/20 (Fig. 7B) and Family 22 (Fig. S30†). The HopA1-phosphotransferase pair could catalyse the dehydration required for LanC-catalysed lanthionine bond formation.²⁹ HopA1-LanC fusions could represent a new uncharacterised lanthionine synthetase, where the lyase and cyclase are fused in a single protein, thereby resembling LanM.²⁹ We also identified additional diverse RiPP-like BGC families (Fig. S31–S32†). Determining the true products of these BGCs represents a significant future effort.

Conclusion

The apoptosis-inducing thioamitides represent some of the most complex RiPPs identified, where thiostreptamide S4 (1) contains 11 post-translational modifications (Fig. 1A). BGC-wide gene deletions can be a powerful method to understand biosynthetic pathways. However, this process can be particularly complicated in RiPP pathways, where a partially modified precursor peptide may rapidly degrade if the pathway stalls in the absence of an essential modification step.¹³ Here, we inactivated every gene in the thiostreptamide S4 BGC and used MS-based untargeted metabolomics and precursor peptide mutations to inform a model of how thioamitides are biosynthesised in the bacterial cell. Our analysis of the metabolites produced by expressing the thiostreptamide S4 BGC in *S. coelicolor* M1146 resulted in the identification of 2–16 (Fig. S17†), mainly by detailed LC-MS/MS characterisation. This LC-MS characterisation was supported by detailed NMR characterisation of 12, a key thioamidated and dehydrated shunt metabolite.

These data include a number of key findings about thioamide biosynthesis, which enables a biosynthetic pathway to be proposed (Fig. 6A). Our work confirms that YcaO and TfuA domain proteins (TsaH and TsaI) are required for iterative thioamidation and this functions as a gatekeeper for all subsequent biosynthetic steps. Prior studies of archaeal YcaO proteins indicate that this is an ATP-dependent process.^{26,60} We define the proteins responsible for histidine hydroxylation and bis-methylation, as well as the reductase required for N-terminal reduction in the thioalbamide pathway. Bis-methylated histidine is currently only found in thioamitides, although 1-*N*-methyl-His is found in archaeal methyl-coenzyme M reductase, which is a protein that intriguingly features a number of other RiPP-like modifications, including thioamidation, methylation, oxidation and hydroxylation.^{61,62} Yeast-mediated assembly provided a route to site-directed mutagenesis of the thiostreptamide S4 precursor peptide, which demonstrated that the pathway is tolerant to precursor peptide mutations, but does stall at an early stage in the biosynthetic pathway with some mutations. This indicates that

macrocyclisation is a bottleneck for engineering thiostreptamide S4 biosynthesis.

We show that a phosphotransferase and a HopA1-like protein (TsaC and TsaD) are required for dehydration, which represents a new route to α,β -dehydroamino acids. Our results contrast with a recent study indicating that lanthipeptide synthetases encoded outside of the BGC catalyse dehydration in thioamide biosynthesis.³⁸ Metabolomic results show that a further phosphotransferase (TsaE) is essential for biosynthesis, where it may have a role in either dehydration or macrocyclisation. A detailed informatic analysis using RIPPER²² shows that the phosphotransferase/HopA1-like protein pair defines multiple new RiPP BGC families, with representatives across over 1,000 sequenced genomes. The variety of tailoring enzymes and precursor peptide sequences indicates that the products will be highly diverse. This is supported by the parallel identification of HopA1-containing BGCs by the decRiPPter algorithm,⁵⁶ which have been recently defined as lanthidins in antiSMASH 5.0.⁶³

Our insights are supported by parallel studies of individual enzymes in other thioamide pathways,^{11,12,64} as well as the recent discoveries of cacaoidin,³⁶ lexapeptide³⁷ and pristinins,⁵⁶ which contain AviMeCys macrocycles, as predicted from our experimental and informatic analyses. We anticipate that the data reported here will inform further experimental work on the thioamitides and related RiPPs to determine key biosynthetic steps, including the true role of HopA1 domain proteins in both RiPP biosynthesis and as a *P. syringae* effector protein,³³ given that this domain does not feature a known catalytic domain. Similarly, it will be important to determine whether bacterial RiPP-associated TfuA proteins function in an equivalent way to the recently characterised archaeal TfuA protein, which hydrolyses thiocarboxylated ThiS to provide a sulphur donor for its cognate YcaO protein.²⁸ A further key goal is to determine the effect that each thioamide post-translational modification has on antiproliferative activity towards cancer cells.⁶⁷ More widely, understanding the diversity of products made by HopA1-like associated RiPP BGCs will be a substantial and exciting research effort, especially given the diversity of pathways identified.

Author contributions

Tom Eyles: investigation, methodology, conceptualisation, visualisation, writing – original draft and review & editing. Natalia Vior: investigation, methodology, formal analysis, data curation, visualisation, writing – review & editing. Rodney Lacret: investigation, validation. Andrew Truman: project administration, supervision, funding acquisition, methodology, conceptualisation, visualisation, writing – original draft and review & editing.

Conflicts of interest

There are no conflicts to declare.



Acknowledgements

This work was funded by a Biotechnology and Biological Sciences Research Council (BBSRC) Norwich Research Park Doctoral Training Partnership grant (BB/M011216/1) for T. H. E., a Royal Society University Research Fellowship (A. W. T.), and BBSRC MET and MfN Institute Strategic Programme grants (BB/J004596/1 and BBS/E/J/000PR9790) for the John Innes Centre (JIC). We are very grateful for the technical assistance at JIC provided by Lionel Hill and Gerhard Saalbach for LC-MS analysis, Sergey Nepogodiev for NMR assistance, and Govind Chandra for assistance with informatics. We thank Vladimir Larionov (National Cancer Institute, NIH, USA) for *S. cerevisiae* VL6-48N and Bradley Moore (Scripps Institution of Oceanography, University of California San Diego, USA) for pCAP03. We are thankful to Javier Santos-Aberturas for helpful discussions.

References

- Y. Hayakawa, K. Sasaki, H. Adachi, K. Furihata, K. Nagai and K. Shin-ya, *J. Antibiot.*, 2006, **59**, 1–5.
- M. Montalbán-López, T. A. Scott, S. Ramesh, I. R. Rahman, A. J. van Heel, J. H. Viel, V. Bandarian, E. Dittmann, O. Genilloud, Y. Goto, M. J. Grande Burgos, C. Hill, S. Kim, J. Koehnke, J. A. Latham, A. J. Link, B. Martínez, S. K. Nair, Y. Nicolet, S. Rebuffat, H.-G. Sahl, D. Sareen, E. W. Schmidt, L. Schmitt, K. Severinov, R. D. Süßmuth, A. W. Truman, H. Wang, J.-K. Weng, G. P. van Wezel, Q. Zhang, J. Zhong, J. Piel, D. A. Mitchell, O. P. Kuipers and W. A. van der Donk, *Nat. Prod. Rep.*, 2021, **38**, 130–239.
- L. Kjaerulff, A. Sikandar, N. Zaburannyi, S. Adam, J. Herrmann, J. Koehnke and R. Müller, *ACS Chem. Biol.*, 2017, **12**, 2837–2841.
- T. Kawahara, M. Izumikawa, I. Kozono, J. Hashimoto, N. Kagaya, H. Koiwai, M. Komatsu, M. Fujie, N. Sato, H. Ikeda and K. Shin-ya, *J. Nat. Prod.*, 2018, **81**, 264–269.
- L. Frattaruolo, R. Lacret, A. R. Cappello and A. W. Truman, *ACS Chem. Biol.*, 2017, **12**, 2815–2822.
- S. Takase, R. Kurokawa, Y. Kondoh, K. Honda, T. Suzuki, T. Kawahara, H. Ikeda, N. Dohmae, H. Osada, K. Shin-ya, T. Kushiro, M. Yoshida and K. Matsumoto, *ACS Chem. Biol.*, 2019, **14**, 1819–1828.
- L. Frattaruolo, M. Fiorillo, M. Brindisi, R. Curcio, V. Dolce, R. Lacret, A. W. Truman, F. Sotgia, M. P. Lisanti and A. R. Cappello, *Cells*, 2019, **8**, 1408.
- C. S. Sit, S. Yoganathan and J. C. Vederas, *Acc. Chem. Res.*, 2011, **44**, 261–268.
- J. Tang, J. Lu, Q. Luo and H. Wang, *Chin. Chem. Lett.*, 2018, **29**, 1022–1028.
- G. E. Kenney, L. M. K. Dassama, M.-E. Pandelia, A. S. Gizzi, R. J. Martinie, P. Gao, C. J. DeHart, L. F. Schachner, O. S. Skinner, S. Y. Ro, X. Zhu, M. Sadek, P. M. Thomas, S. C. Almo, J. M. Bollinger, C. Krebs, N. L. Kelleher and A. C. Rosenzweig, *Science*, 2018, **359**, 1411–1416.
- J. Lu, J. Li, Y. Wu, X. Fang, J. Zhu and H. Wang, *Org. Lett.*, 2019, **21**, 4676–4679.
- A. Sikandar, M. Lopatniuk, A. Luzhetskyy and J. Koehnke, *ACS Chem. Biol.*, 2020, **15**, 2815–2819.
- W. J. K. Crone, N. M. Vior, J. Santos-Aberturas, L. G. Schmitz, F. J. Leeper and A. W. Truman, *Angew. Chem., Int. Ed.*, 2016, **55**, 9639–9643.
- V. Larionov, N. Kouprina, J. Graves, X. N. Chen, J. R. Korenberg and M. A. Resnick, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 491–496.
- X. Tang, J. Li, N. Millán-Aguinaga, J. J. Zhang, E. C. O'Neill, J. A. Ugalde, P. R. Jensen, S. M. Mantovani and B. S. Moore, *ACS Chem. Biol.*, 2015, **10**, 2841–2849.
- J. P. Gomez-Escribano and M. J. Bibb, *J. Microbiol. Biotechnol.*, 2011, **4**, 207–215.
- B. Gust, G. L. Challis, K. Fowler, T. Kieser and K. F. Chater, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 1541–1546.
- M. J. Bibb, J. White, J. M. Ward and G. R. Janssen, *Mol. Microbiol.*, 1994, **14**, 533–545.
- H.-J. Hong, M. I. Hutchings, L. M. Hill and M. J. Buttner, *J. Biol. Chem.*, 2005, **280**, 13055–13061.
- C. Wang, V. Hantke, R. J. Cox and E. Skellam, *Org. Lett.*, 2019, **21**, 4163–4167.
- X. Chen, E. G. Mietlicki-Baase, T. M. Barrett, L. E. McGrath, K. Koch-Laskowski, J. J. Ferrie, M. R. Hayes and E. J. Petersson, *J. Am. Chem. Soc.*, 2017, **139**, 16688–16695.
- J. Santos-Aberturas, G. Chandra, L. Frattaruolo, R. Lacret, T. H. Pham, N. M. Vior, T. H. Eyles and A. W. Truman, *Nucleic Acids Res.*, 2019, **47**, 4624–4637.
- R. P. Grese, R. L. Cerny and M. L. Gross, *J. Am. Chem. Soc.*, 1989, **111**, 2835–2842.
- K. A. Newton and S. A. McLuckey, *J. Am. Soc. Mass Spectrom.*, 2004, **15**, 607–615.
- D. D. Nayak, N. Mahanta, D. A. Mitchell and W. W. Metcalf, *eLife*, 2017, **6**, e29218.
- N. Mahanta, A. Liu, S. Dong, S. K. Nair and D. A. Mitchell, *Proc. Natl. Acad. Sci. U. S. A.*, 2018, **115**, 3030–3035.
- C. J. Schwalen, G. A. Hudson, B. Kille and D. A. Mitchell, *J. Am. Chem. Soc.*, 2018, **140**, 9494–9501.
- A. Liu, Y. Si, S.-H. Dong, N. Mahanta, H. N. Penkala, S. K. Nair and D. A. Mitchell, *Nat. Chem. Biol.*, 2021, DOI: 10.1038/s41589-021-00771-0.
- L. M. Repka, J. R. Chekan, S. K. Nair and W. A. van der Donk, *Chem. Rev.*, 2017, **117**, 5457–5520.
- W. C. Hon, G. A. McKay, P. R. Thompson, R. M. Sweet, D. S. Yang, G. D. Wright and A. M. Berghuis, *Cell*, 1997, **89**, 887–895.
- D. M. Daigle, G. A. McKay, P. R. Thompson and G. D. Wright, *Chem. Biol.*, 1999, **6**, 11–18.
- S. H. Kim, S. I. Kwon, D. Saha, N. C. Anyanwu and W. Gassmann, *Plant Physiol.*, 2009, **150**, 1723–1732.
- Y. Park, I. Shin and S. Rhee, *J. Struct. Biol.*, 2015, **189**, 276–280.
- H. Li, H. Xu, Y. Zhou, J. Zhang, C. Long, S. Li, S. Chen, J.-M. Zhou and F. Shao, *Science*, 2007, **315**, 1000–1003.
- Y. Goto, A. Ökesli and W. A. van der Donk, *Biochemistry*, 2011, **50**, 891–898.
- F. J. Ortiz-López, D. Carretero-Molina, M. Sánchez-Hidalgo, J. Martín, I. González, F. Román-Hurtado, M. de la Cruz,



- S. García-Fernández, F. Reyes, J. P. Deisinger, A. Müller, T. Schneider and O. Genilloud, *Angew. Chem., Int. Ed.*, 2020, **59**, 12654–12658.
- 37 M. Xu, F. Zhang, Z. Cheng, G. Bashiri, J. Wang, J. Hong, Y. Wang, L. Xu, X. Chen, S.-X. Huang, S. Lin, Z. Deng and M. Tao, *Angew. Chem., Int. Ed.*, 2020, **59**, 18029–18035.
- 38 J. Lu, Y. Wu, Y. Li and H. Wang, *Angew. Chem., Int. Ed.*, 2021, **60**, 1951–1958.
- 39 M. Blaesse, T. Kupke, R. Huber and S. Steinbacher, *EMBO J.*, 2000, **19**, 6299–6310.
- 40 V. Wiebach, A. Mainz, M.-A. J. Siegert, N. A. Jungmann, G. Lesquame, S. Tirat, A. Dreux-Zigha, J. Aszodi, D. Le Beller and R. D. Süßmuth, *Nat. Chem. Biol.*, 2018, **14**, 652–654.
- 41 M. Blaesse, T. Kupke, R. Huber and S. Steinbacher, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2003, **59**, 1414–1421.
- 42 T. Kupke and F. Götz, *J. Biol. Chem.*, 1997, **272**, 4759–4762.
- 43 J. Claesen and M. Bibb, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 16297–16302.
- 44 W. Ding, N. Yuan, D. Mandalapu, T. Mo, S. Dong and Q. Zhang, *Org. Lett.*, 2018, **20**, 7670–7673.
- 45 S. Ma and Q. Zhang, *Nat. Prod. Rep.*, 2020, **37**, 1152–1163.
- 46 F. Katzen, D. U. Ferreira, C. G. Oddo, M. V. Ielmini, A. Becker, A. Pühler and L. Ielpi, *J. Bacteriol.*, 1998, **180**, 1607–1617.
- 47 J. E. Velásquez, X. Zhang and W. A. van der Donk, *Chem. Biol.*, 2011, **18**, 857–867.
- 48 M. F. Freeman, C. Gurgui, M. J. Helf, B. I. Morinaka, A. R. Uria, N. J. Oldham, H. G. Sahl, S. Matsunaga and J. Piel, *Science*, 2012, **338**, 387–390.
- 49 B. Kalyanaraman, G. Cheng, M. Hardy, O. Ouari, M. Lopez, J. Joseph, J. Zielonka and M. B. Dwinell, *Redox Biol.*, 2018, **14**, 316–327.
- 50 M. J. Helf, M. F. Freeman and J. Piel, *J. Ind. Microbiol. Biotechnol.*, 2019, **46**, 551–563.
- 51 G. H. Völler, J. M. Krawczyk, A. Pesic, B. Krawczyk, J. Nachtigall and R. D. Süßmuth, *ChemBioChem*, 2012, **13**, 1174–1183.
- 52 T. H. Eyles, N. M. Vior and A. W. Truman, *ACS Synth. Biol.*, 2018, **7**, 1211–1218.
- 53 K. Kudo, H. Koiwai, N. Kagaya, M. Nishiyama, T. Kuzuyama, K. Shin-ya and H. Ikeda, *ACS Chem. Biol.*, 2019, **14**, 1135–1140.
- 54 M. A. Ortega, J. E. Velásquez, N. Garg, Q. Zhang, R. E. Joyce, S. K. Nair and W. A. van der Donk, *ACS Chem. Biol.*, 2014, **9**, 1718–1725.
- 55 S. Halary, J. O. McInerney, P. Lopez and E. Bapteste, *BMC Evol. Biol.*, 2013, **13**, 146.
- 56 A. M. Kloosterman, P. Cimermanic, S. S. Elsayed, C. Du, M. Hadjithomas, M. S. Donia, M. A. Fischbach, G. P. van Wezel and M. H. Medema, *PLoS Biol.*, 2020, **18**, e3001026.
- 57 P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Res.*, 2003, **13**, 2498–2504.
- 58 G. E. Crooks, G. Hon, J.-M. Chandonia and S. E. Brenner, *Genome Res.*, 2004, **14**, 1188–1190.
- 59 B. Li, D. Sher, L. Kelly, Y. Shi, K. Huang, P. J. Knerr, I. Joewono, D. Rusch, S. W. Chisholm and W. A. van der Donk, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 10430–10435.
- 60 S.-H. Dong, A. Liu, N. Mahanta, D. A. Mitchell and S. K. Nair, *ACS Cent. Sci.*, 2019, **5**, 842–851.
- 61 H. Chen, Q. Gan and C. Fan, *Front. Microbiol.*, 2020, **11**, 578356.
- 62 D. D. Nayak, A. Liu, N. Agrawal, R. Rodriguez-Carerro, S.-H. Dong, D. A. Mitchell, S. K. Nair and W. W. Metcalf, *PLoS Biol.*, 2020, **18**, e3000507.
- 63 K. Blin, S. Shaw, K. Steinke, R. Villebro, N. Ziemert, S. Y. Lee, M. H. Medema and T. Weber, *Nucleic Acids Res.*, 2019, **47**, W81–W87.
- 64 Y. Qiu, J. Liu, Y. Li, Y. Xue and W. Liu, *Cell Chem. Biol.*, 2021, DOI: 10.1016/j.chembiol.2020.12.016.

