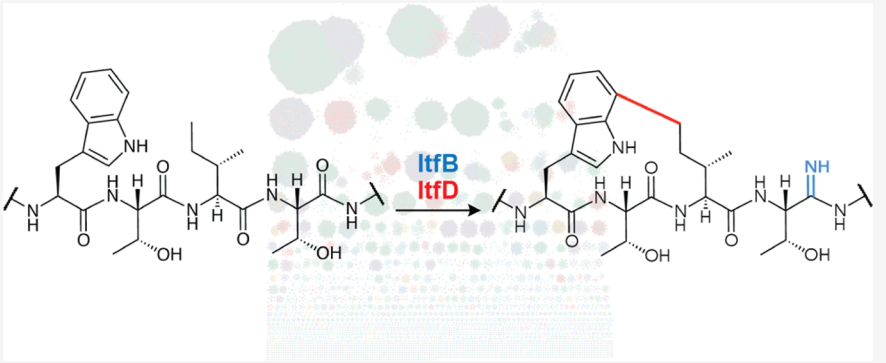
JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Article

Bioinformatic Atlas of Radical SAM Enzyme-Modified RiPP Natural Products Reveals an Isoleucine−Tryptophan Crosslink

Kenzie A. Clark and Mohammad R. Seyedsayamdost*

Cite This: J. Am. Chem. Soc. 2022, 144, 17876−17888
Read Online



ABSTRACT: Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a growing family of natural products with diverse activities and structures. RiPP classes are defined by the tailoring enzyme, which can introduce a narrow range of modifications or a diverse set of alterations. In the latter category, RiPPs synthesized by radical *S*-adenosylmethionine (SAM) enzymes, known as RaS-RiPPs, have emerged as especially divergent. A map of all RaS-RiPP gene clusters does not yet exist. Moreover, precursor peptides remain difficult to predict using computational methods. Herein, we have addressed these challenges and report a bioinformatic atlas of RaS-RiPP gene clusters in available microbial genome sequences. Using co-occurrence of RaS enzymes and transporters from varied families as a bioinformatic hook in conjunction with an in-house code to identify precursor peptides, we generated a map of ∼15,500 RaS-RiPP gene clusters, which reveal a remarkable diversity of syntenies pointing to a tremendous range of enzymatic and natural product chemistries that remain to be explored. To assess its utility, we examined one family of gene clusters encoding a YcaO enzyme and a RaS enzyme. We find the former is noncanonical, contains an iron−sulfur cluster, and installs a novel modification, a backbone amidine into the precursor peptide. The RaS enzyme was also found to install a new modification, a C−C crosslink between the unactivated terminal δ-methyl group of Ile and a Trp side chain. The co-occurrence search can be applied to other families of RiPPs, as we demonstrate with the emerging DUF692 di-iron enzyme superfamily.

INTRODUCTION

Microbial natural products have offered an immense source of therapeutic agents, including some of our most celebrated cures, such as penicillin or vancomycin.[1,2] But their impact has gone beyond medicine, and one of the disciplines that has benefitted from their continued interrogation is metalloenzymology. Investigations into natural product biosynthesis have revealed numerous new enzymatic transformations. The enzymes, however, have been identified rather by accident. Penicillin and vancomycin provide illustrative cases: both antibiotics were discovered in bioactivity-guided screens, and biosynthetic studies revealed, only decades later, the remarkable reactions carried out by the mononuclear iron-dependent isopenicillin N synthase and the cytochrome P450 "Oxy" enzymes.[3−5] Enzyme-first or gene-first discovery methods have only recently become possible with the advent of DNA sequencing technologies and the ensuing explosion in microbial genome sequences, which have ushered in a new paradigm in natural product research.[6−9]

Although uncharacterized metalloenzymes can now be easily spotted in genome sequences, it remains difficult to predict the substrate for new enzyme families. Moreover, most natural product gene clusters are at best sparingly expressed under standard growth conditions and therefore necessitate alternative approaches to access the encoded metabolites. To overcome these challenges, we have focused on ribosomally synthesized and post-translationally modified peptide (RiPPs)

Received: June 20, 2022
Published: September 21, 2022

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

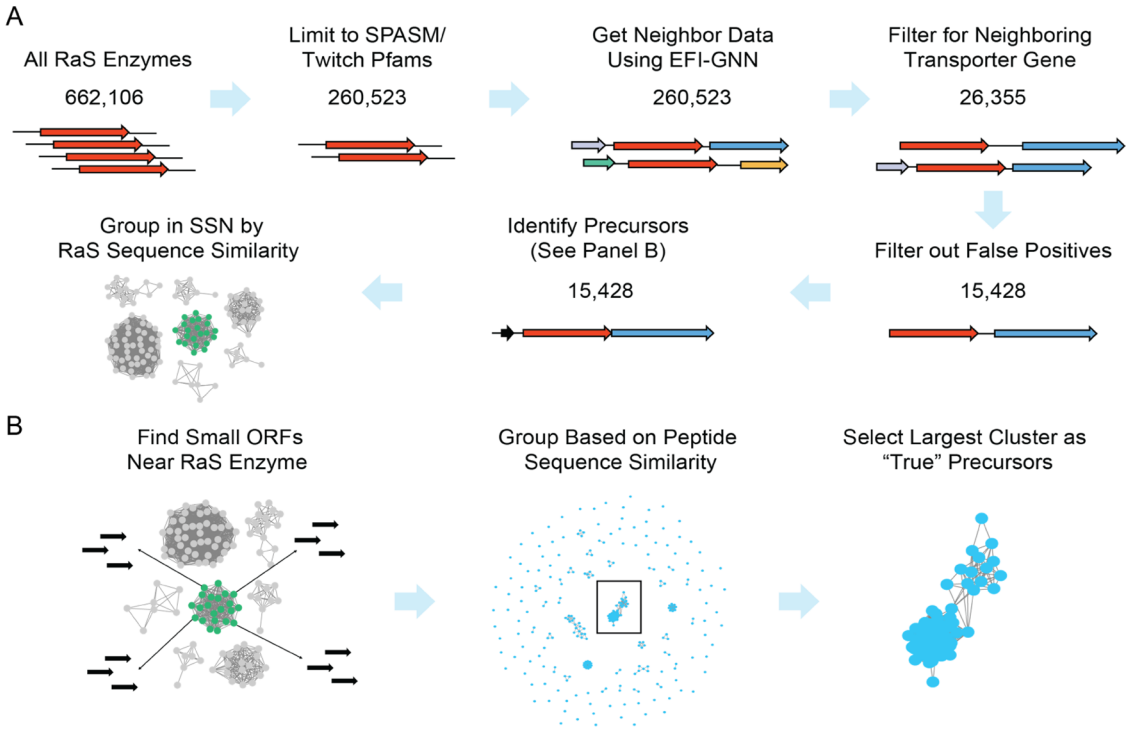


Figure 1. Workflow for generation of a global RaS-RiPPs network. (A) Steps in creating the RaS-RiPP SSN included limiting RaS enzymes to SPASM/Twitch Pfams, searching for neighboring transporter genes, and removing RaS enzymes from primary metabolism and cofactor biosynthesis. (B) Automated detection of precursor peptides was achieved by searching for all possible small open reading frames (ORFs) around each RaS enzyme within a subfamily and then grouping these possible precursors into their own SSN to determine the most likely precursors. LogoPlots were generated from the list of predicted precursors.

natural products that are quorum-sensing (QS)-regulated and generated by radical *S*-adenosylmethionine (RaS) enzymes.[10−13] RiPPs have the unique advantage of a genetically encoded substrate, which enables easy identification of the substrate for the tailoring enzyme of interest. We selected RaS tailoring enzymes, reasoning that the versatile chemistry that members of this superfamily carry out would ensure new classes of RiPPs, rather than variants of known RiPPs.[14−17] Finally, to avoid working on silent biosynthetic gene clusters (BGCs), we focused on those that are regulated by a QS operon.[18,19] Application of these parameters to streptococcal bacteria led to the identification of ∼600 unique RiPP BGCs that are QS-regulated and synthesized by RaS enzymes.[20] Contrary to expectation, our bioinformatic search suggested that streptococci are a deep source of novel metalloenzyme chemistry and RiPP natural products.

We have subsequently validated both propositions by elucidating nearly a dozen new RaS enzyme-catalyzed reactions as well as five classes of RiPP natural products. The transformations include carbon−carbon, carbon−oxygen, and carbon−sulfur bond formation at unactivated positions leading to unusual heterocycles or macrocycles.[20−26] The mature RiPPs that we have discovered—streptide,[26] streptosactin,[24] tryglysin,[27] bicyclostreptin,[28] and enteropeptin[29]—show there is much to be learned regarding the secondary metabolome of streptococci, most of which are associated with human oral microbiomes.

Bioinformatic searches based on gene co-occurrence offer a productive method for categorizing and subsequently analyzing the stores of RaS enzymes that can be found in streptococcal genomes. The QS regulatory operon mentioned above is an effective bioinformatic hook for streptococcal RaS-RiPPs. However, it is not conserved beyond this genus, necessitating other bioinformatic strategies to glean a picture of the "universe" of RaS-RiPPs that are available in sequence databases. In the current work, we have implemented one such strategy to map RaS-RiPP BGCs regardless of microbial origin. Using a co-occurrence search based on the RaS enzyme superfamily and transporters of diverse classes, we generate an atlas of all such instances in available microbial genomes. To locate the precursor peptide, we developed a new algorithm based on sequence similarity networking. Our subsequent map reveals ∼15,500 largely uncharacterized BGCs that organize into 800 subfamilies in diverse microbial phyla. Analysis of one such subfamily that encodes a RaS enzyme and a YcaO homolog revealed two unprecedented transformations, verifying that the encoded enzymes are active and catalyze interesting reactions. The bioinformatic strategy we describe can be applied to any desired enzyme Pfam, and our network of RaS-RiPPs will provide a source of novel enzymology, natural product chemistry, and biology for years to come.

RESULTS AND DISCUSSION

Bioinformatic Rationale. Initial work by Haft and Basu showed that the family of RaS-RiPPs is diverse and widely encoded.[30,31] In recent years, sequence similarity networks (SSNs) and genomic enzymology pioneered by Gerlt and coworkers have become popular tools for identifying new RiPP BGCs.[32−34] As mentioned above, we previously used a streptococcal SSN of RaS-RiPPs to uncover novel RaS enzyme reactions and natural products. More recently, manual inspection of clusters from an EFI-EST-based network of RaS enzymes led to the discovery of an enzyme that catalyzes the repeated formation of cyclopropylglycine from valine, a

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

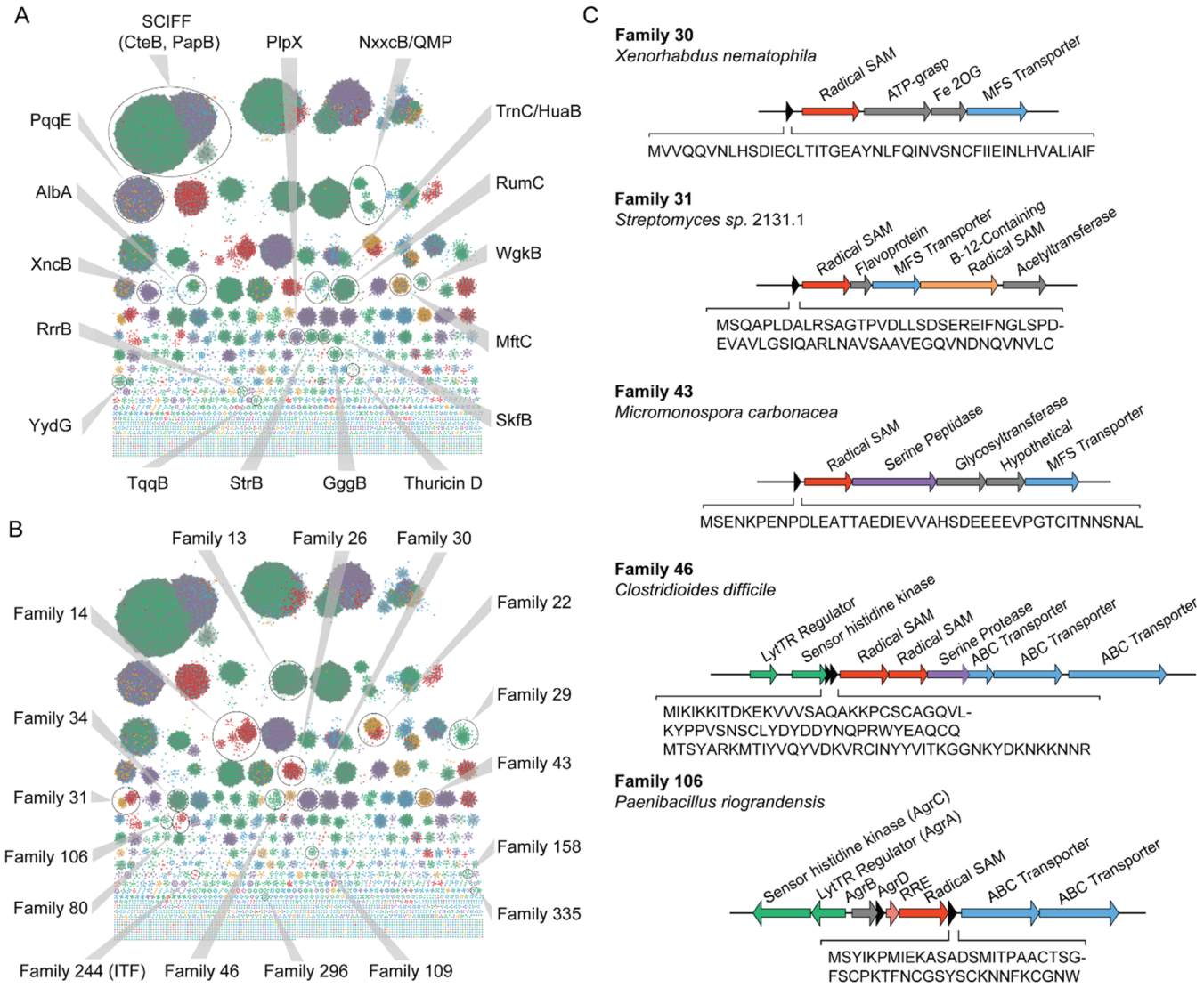


Figure 2. SSN of RaS-RiPPs with neighboring transporters. (A, B) Each node represents an individual RaS enzyme, and nodes connecting them indicate sequence similarity. Nodes are colored based on phyla showing Firmicutes (green), Proteobacteria (purple), Bacteroidetes (red), Actinobacteria (orange), and other phyla (blue). The network contains 15,428 nodes organized as 1294 singletons and 853 families with at least two nodes (see Data S1). Known RaS-RiPP families (A) or interesting ones discussed in the main text or SI (B) are labeled. (C) Representative BGCs from a selection of subfamilies. Biosynthetic proteins are labeled based on annotations; precursor peptide sequences are shown.

new reaction for RaS enzymes.[35] Additionally, an SSN approach has been used to map the biosynthetic landscape of cyclophane-forming RaS enzymes, revealing new cyclophane macrocycles.[36] To make these sorts of analyses easier, RadicalSAM.org was created as a web-based tool for the analysis of SSNs for the entire RaS family.[34]

We set out to develop a bioinformatics platform that could benefit from the use of SSNs but would also include computational filtering steps to limit the amount of manual analysis required. Importantly, we did not want the method to rely on a high level of homology to a single protein but rather co-occurrence of two proteins to enhance the fidelity of finding RiPP BGCs. Examination of known RaS-RiPP clusters revealed that, aside from the RaS enzyme, transporters are the most common element, specifically seven Pfams describing ABC, MFS, MATE, and SecDF transporters (Table S1).[37−40] While a previous study has used the LanT transporter to identify additional examples of lanthipeptide gene clusters, we planned to use a broadly defined search for transporters as one hook in a co-occurrence search.[41] The RaS enzyme superfamily Pfam (PF04055) contained 662,106 members at the time of the most recent search in December 2021, which was too large for the analyses we envisioned. Given that the majority of RaS enzymes found in RaS-RiPP clusters so far belong to either the SPASM or Twitch subfamilies, we narrowed our search to these, reducing the total number of enzyme sequences to 260,523 (Table S2). These RaS enzymes carry an extended C-terminal domain that can bind additional [4Fe-4S] clusters, which can participate in electron transfer reactions and/or directly bind substrates.[42]

All SPASM/Twitch RaS enzymes along with their neighborhood regions were obtained using the EFI-EST and EFI-GNN tools (Figure 1A).[33] The neighborhood information was used to filter the list of RaS enzymes for instances in which a transporter belonging to one of seven transporter Pfams (PF00005, PF00664, PF07690, PF01554, PF01061, PF02355, or PF13346) occurred within five genes from the RaS enzyme on the same strand. Cases in which the RaS enzyme and

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

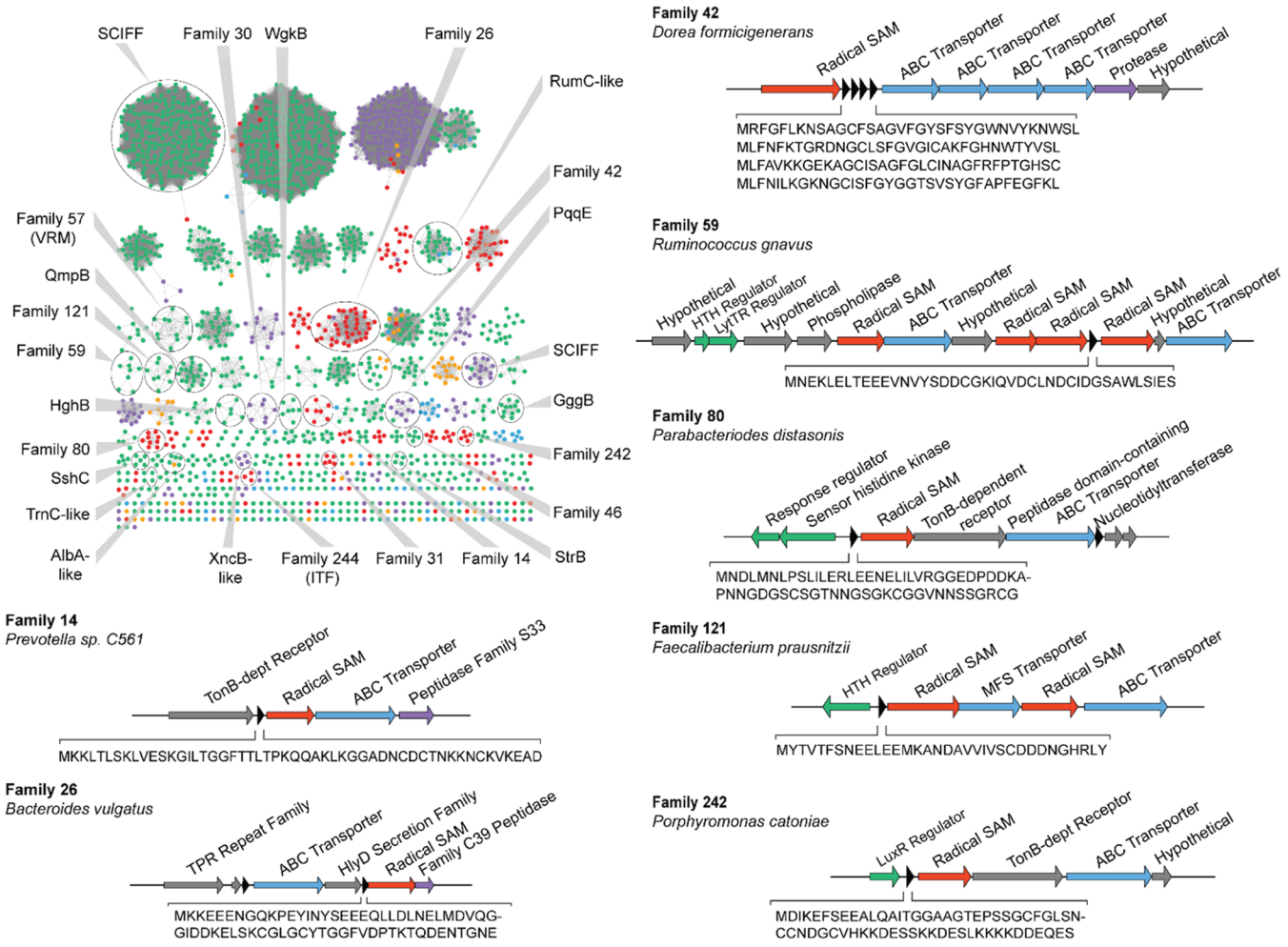


Figure 3. SSN of RaS-RiPPs from the human microbiome. The network contains 1,905 BGCs with 138 singletons and 150 subfamilies with at least two nodes. Representative BGCs from selected subfamilies are shown. Family names are based on the full SSN (Figure 2).

transporter occur on opposite strands are far less common and focusing only on same-strand instances wherein the RaS enzyme and transporter are within five genes of each other provided an optimal balance between a low number of false positives and capture of most known RaS-RiPPs BGCs. This selection step returned 26,355 instances, which were further filtered to remove RaS enzymes that are involved in cofactor biosynthesis and primary metabolism (Table S3). The remaining 15,428 RaS enzymes were submitted to EFI-EST to create a final sequence similarity network (SSN), which was visualized in Cytoscape.[43] The final network contained 15,428 nodes, each representing a RaS enzyme, with 1,757,639 edges connecting them based on an alignment score of 70. The RaS enzymes clustered into 853 families with at least two nodes and 1294 singleton nodes.

Automated Detection of Precursor Peptides. A major challenge in the analysis of new RiPP BGCs is the accurate identification of precursor peptides. In previous bioinformatics studies, they have been located using RODEO for large-scale analysis of BGCs of known RiPP families,[44] RiPPER for precursor peptides of unknown RiPP classes,[45] or, alternatively, manually for novel families of RiPPs in a much smaller set of BGCs.[20] Methods using machine learning predictions have also been developed including NeuRiPP and DeepRiPP, which rely on training neural networks.[46,47] We opted to generate our own simplified approach, which shares some similarities with the RiPPER workflow (Figure 1B). It takes advantage of the fact that the RaS enzymes are already organized in an SSN. We surmised that RaS enzymes sharing a high level of sequence similarity would carry out a similar reaction on their substrates; these peptides therefore would also contain significant sequence similarity. Based on this idea, we developed an algorithm that identifies all possible intergenic open reading frames (ORFs) or annotated genes that are 20−85 amino acids long within a 4000 base pair window upstream and downstream from each RaS enzyme. We found that this length was optimal for correctly identifying known precursor peptides, although we acknowledge that longer precursor peptides such as those belonging to the Nif11/NHLP families will not be detected. For each subfamily in the RaS enzyme network with more than four nodes, the algorithm creates an SSN containing all of the identified ORFs representing potential precursor peptides. From each of these networks, the subfamily carrying the maximum connected components was identified. Peptide sequences contained in this subfamily were mapped back to the corresponding BGC and designated as the precursor peptide. The strategy was applied to the network to identify precursor peptides (Figure S1).

While this approach is imperfect and could not identify precursor peptides for all families in our network, it correctly identified precursor peptides for nearly all previously characterized RaS-RiPP BGCs (11 out of 15 clusters, Table

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

S4). In some cases, no precursors were identified, perhaps because the BGC is not a true RiPP cluster. In other cases, the algorithm incorrectly identified precursor peptides when all members of the family are highly similar, so the intergenic regions contain many random ORFs with high sequence similarity. Conversely, in cases where there was a large amount of sequence diversity in the RaS enzymes, particularly in the largest families, there can be several identifiable subfamilies and the precursors are only located for the largest of these. The effectiveness of the algorithm was highly dependent on the alignment score used in the RaS enzyme network and the parameters used for the generation of the precursor network, both of which were optimized based on their ability to predict precursors for the known RaS-RiPP families. Identification of a precursor peptide was not required for inclusion in the final network. These challenges notwithstanding, the algorithm predicted precursor peptides for most subfamilies in the SSN. The combined approaches allowed us to generate an atlas of all available RaS-RiPP BGCs in microbial genomes within the selected criteria.

Analysis of the RaS-RiPPs Atlas. The network is rich with novel and interesting BGCs. Most previously characterized RaS-RiPP BGCs, including those from our *Streptococcus* network, were identified and are highlighted (Figure 2A, Data S1, and Table S4). Sequences from Firmicutes make up nearly half of the network (45%), consistent with the previous identification of many RaS-RiPPs from this phylum (Table S5). The network also shows many RaS-RiPPs from Proteobacteria (21%) and Bacteroidetes (9%), both prevalent in the human microbiome. Many families are made up of sequences from a single phylum, but several contain sequences that range across multiple phyla, such as Family 22 and 31.

Analysis of individual families revealed that the network contains many clusters that do not seem to belong to known RaS-RiPP classes. Many contain regulators, peptidases, RiPP recognition element (RRE) proteins, and other biosynthetic enzymes like lanthipeptide synthases, YcaO-like proteins, and epimerases within their gene clusters (Figure 2B,C, Data S1, and Table S6). Family 30, primarily from *Pseudomonas* and *Xenorhabdus*, contains an interesting set of BGCs, which harbor a combination of a RaS enzyme, ATP-Grasp enzyme, and a member of the iron/2-oxoglutarate (Fe/2OG) family along with an MFS transporter.[48] RaS-RiPP BGCs from actinomycetes were not common, but several were identified including the one shown from Family 31 that contains a RaS enzyme along with a flavoprotein and B-12-dependent RaS enzyme. Clusters containing glycosyltransferases, indicative of a possible glycosylated RaS-RiPP, were relatively common (e.g., Family 43). Uncommon regulatory genes were observed as well, e.g., in Family 106, which encodes an upstream accessory gene regulator (Agr) quorum-sensing system.[49]

RaS-RiPPs from the Human Microbiome. Many of the streptococci that we focused on in our previous network are important members of the human microbiome. We wondered how many additional RaS-RiPPs might be found from this broader network. The data obtained from the EFI-EST webserver for each RaS enzyme included information on its source and could thus be filtered for those from the Human Microbiome Project (HMP),[32,50,51] giving rise to an SSN of 1905 RaS-RiPP BGCs across numerous phyla (Figure 3).

Many of the clusters from the original streptococcal network were found in this network including members of the *HGH*, *WGK*, *STR*, *QMP*, *GGG*, and *SSH* subfamilies, highlighting the importance of future investigations into the functions of these RiPP classes.[20] Additionally, gene clusters related to subtilosin, SCIFF, ruminococcin, PQQ, thuricin CD, and cyclophanes could be identified. Beyond these known clusters, there are many interesting families that do not share similarities with known clusters. One example containing a RaS enzyme, ABC transporter, and peptidase from Family 26 comes from *Bacteroides vulgatus*, one of the most common bacterial species in healthy adults.[52] Family 42 contains a single RaS enzyme but has four precursor peptides with varying sequences and four ABC transporters. One of the most complex BGCs is found in Family 59 from *Ruminococcus gnavus* and contains four RaS enzymes. Additional examples of interesting RaS-RiPP BGCs from the human microbiome are depicted (Figure 3). The analysis shows that RaS-RiPPs are prevalent in the human microbiome that could encode health-relevant molecules.

RaS-RiPPs from *Clostridium* spp. *Clostridium* is an especially important genus in the human gut with established roles in health and disease. We therefore generated a subnetwork focusing on RaS-RiPPs from this genus, resulting in 936 BGCs grouped into 73 families (Figure S2). Here again, we could easily locate clusters with interesting gene composition. Family 29 from *Clostridium hiranonis*, for example, is a complex cluster with multiple RaS enzymes, transporters, and precursors as well as a regulator and peptidase. Two examples from *Clostridium perfringens*, a common cause of intestinal disease in humans and animals,[53] are found in Families 296 and 335. Another interesting example from *Clostridium butyricum* in Family 158 contains four copies of an identical precursor peptide in addition to a RaS enzyme, transporter, and protease. Similar analyses can be carried out to generate subsets from the original network by selecting any feature of interest. For example, networks can easily be generated for RaS enzymes with a neighboring RRE, glycosyltransferase, or regulator as well as for RaS enzymes from any specific phylum or genus. These analyses are effective starting points for prioritizing BGCs for further interrogation.

Bioinformatic Search for DUF692-RiPP Gene Clusters. The co-occurrence search strategy is not limited to RaS enzymes and can be applied to other enzyme families as well. One emerging class is the multinuclear iron enzymes belonging to the DUF692 Pfam. Two members have been characterized thus far, MbnB, which generates an oxazolone−thioamide moiety in an unprecedented four-electron oxidation reaction in methonobactin synthesis,[54] and TglH, which also catalyzes an unusual four-electron oxidative rearrangement that results in excision of the Cβ atom from a Cys residue in pearlin biosynthesis.[55] These remarkable reactions left us wondering whether additional DUF692-RiPP BGCs could be uncovered using our bioinformatic strategy. To generate a DUF692-RiPPs network, we followed a similar workflow as with RaS-RiPPs: we found all DUF692 enzymes adjacent to the same family of transporters as above and subsequently located precursor peptides (Figure S3). Starting with a total of 12,342 DUF692 enzymes, we identified 1102 RiPP BGCs, which arranged in a network consisting of 36 families with at least two homologous members and 36 singletons (Figure S4). Unlike the RaS-RiPP network, this SSN is dominated by sequences from Proteobacteria, which account for 85% of BGCs (Table S7). There are two very large subfamilies from proteobacteria that do not obviously resemble RiPP gene clusters, though many of the smaller subfamilies resemble genuine RiPP BGCs. Some of

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

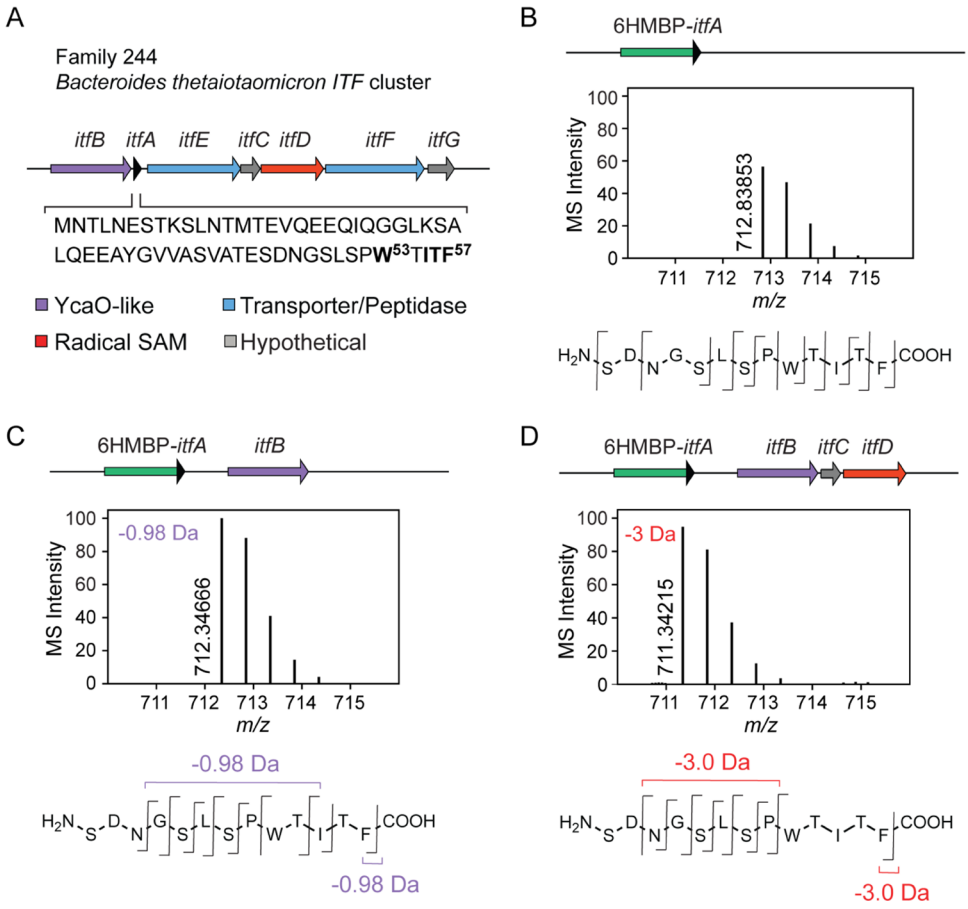


Figure 4. Reactions of ItfBCD characterized by HR-MS and HR-MS/MS. (A) *itf* BGC from *B. thetaiotaomicron*. Genes are color-coded, and the sequence of the precursor peptide is shown. (B–D) HR-MS and HR-MS/MS analyses of heterologous expression constructs indicated. Shown are data for ItfA after coexpression with (B) no other enzymes as control, (C) with ItfB, or (D) with ItfBCD. The construct is shown above each spectrum, which is zoomed in on the $[M + 2H]^{2+}$ ion of the GluC-cleaved peptide. Coexpression with ItfB and ItfBCD gives a product that is −0.98 and −3 Da relative to the substrate, respectively.

the most common neighbors were PF01370 (NAD-dependent epimerase/dehydratase family), PF02219 (methylenetetrahydrofolate reductase), PF07690 (MFS), PF12833 (HTH), PF12849 (periplasmic binding protein), and PF01494 (FAD-dependent enzyme) (Table S8).

ITF Subfamily. With an SSN of available RaS-RiPPs at hand, we next sought to test its ability in identifying new families of modifications. We selected Family 244 from the RaS-RiPP SSN as it (i) is encoded by *Bacteroides thetaiotaomicron*, a prominent member and model strain in the human microbiome,[56,57] (ii) contains an unusual combination of a RaS enzyme and a YcaO, an enzyme involved in the ATP-dependent formation of diverse hetero- and macrocycles,[58] and (iii) is associated with a precursor peptide that lacks Cys residues eliminating the possibility of thioether bond formation, a common modification in RaS-RiPPs. We have named this cluster *itf* based on the conserved C-terminal Ile-Thr-Phe triad in the precursor peptide. This BGC can be found in six strains with an identical 57mer precursor peptide (ItfA) (Table S9), a YcaO enzyme (ItfB),[58] a RaS enzyme (ItfD), an ABC transporter/C39 peptidase (ItfE), and a TonB-dependent transporter (ItfF) (Figure 4A). Additionally, two small hypothetical proteins (ItfC and ItfG) were found in the gene cluster, but neither encoded a *bona fide* RiPP recognition element.[59]

Characterization of the Reactions Catalyzed by ItfB, ItfC, and ItfD. Due to the length of the precursor peptide and the presence of multiple tailoring enzymes in the gene cluster, we chose to pursue a heterologous coexpression strategy to characterize the reactions carried out by each biosynthetic enzyme. The precursor peptide, ItfA, was expressed in *Escherichia coli* with an N-terminal hexaHis maltose-binding protein (6HMBP) tag with a 3HRV3C protease cleavage site between the tag and the ItfA sequence. The 6HMBP-tagged ItfA was inserted into the first multiple cloning site of a pRSFDuet vector to allow for coexpression with ItfB or ItfBCD inserted into the second multiple cloning site (Tables S10 and S11). While ItfC was not bioinformatically identified as an RRE,[59] it was included in case it was necessary for the activity of the RaS enzyme (ItfD). After expression, the biosynthetic enzymes can modify the precursor peptide in *E. coli* followed by purification of the product by metal affinity chromatography. The tag can be removed through proteolysis and the structure of the product is determined using high-resolution mass spectrometry (HR-MS), tandem HR-MS (HR-MS/MS), and one-dimensional/two-dimensional (1D/2D) NMR spectroscopy.

Expression and purification of the ItfA precursor alone, followed by digestion with the GluC protease gave the expected unmodified linear m/z and fragmentation pattern (Figure 4B, Tables S12, and S13). Following coexpression with

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

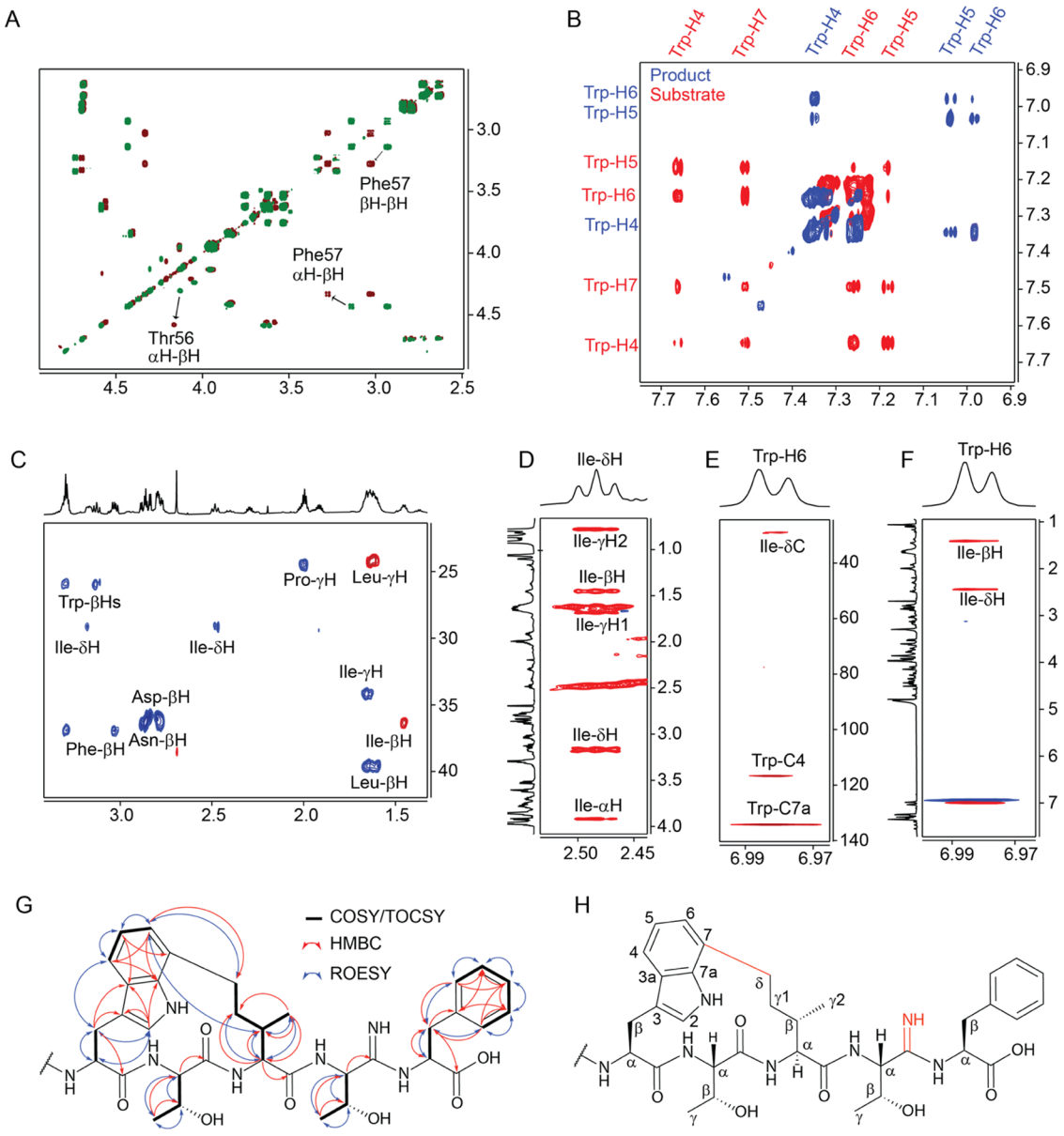


Figure 5. Structural elucidation of the products of ItfB and ItfBCD. (A) Comparative analysis of COSY spectra of unmodified ItfA and ItfB-modified ItfA (i.e., ItfB product). (B) Comparative analysis of TOCSY spectra of unmodified ItfA and ItfBCD product focusing on Trp53 correlations. (C) HSQC spectra of ItfBCD product focusing on Ile55 correlations. (D) TOCSY correlations of the δ-H of Ile55 in the ItfBCD product. (E) HMBC correlations of Trp53-H6 in the ItfBCD product. (F) ROESY correlations of Trp53-H6 in the ItfBCD product. Crosspeaks are labeled in the 2D NMR spectra. (G) Summary of relevant NMR correlations used to solve the structure of the ItfBCD product. (H) Structure of the relevant portion of the ItfBCD product; modifications are shown in red.

the YcaO protein (ItfB) alone, a mass loss of 0.98 Da and a reduction in retention time was observed (Table S12 and Figure S5). This mass loss was intriguing and suggested a possible conversion of an oxygen to a nitrogen in the peptide. HR-MS/MS analysis located the modification to the terminal Thr56-Phe57 dyad: all b-ions generated in the collision-induced dissociation of the ItfB product before Thr57 were either unmodified or not observed (Table S14). Tellingly, the final b-ion following Phe57 and all y-ion fragments observed beginning after the C-terminal ITF were 0.98 Da lighter than the equivalent fragments in the unmodified substrate (Figure 4C).

When ItfB, ItfC, and ItfD were coexpressed with ItfA, a product that was 3 Da lighter than the substrate with another reduction in retention time was observed suggesting that an additional −2 Da modification had occurred in the presence of the RaS enzyme (ItfD) and hypothetical protein (ItfC) (Table S12 and Figure S5). HR-MS/MS showed that all b-ions before b8 were unmodified and the final b13 contained the same −3 Da mass loss as the final product (Table S15). All y-ions observed beginning with y5 were 3 Da lighter than the substrate. This fragmentation pattern suggests that ItfD installs a second modification in the region containing Trp53-Phe57 (Figure 4D).

Structural Elucidation of the ItfB and ItfBCD Products. To gain more detailed structural information, the unmodified peptide, the ItfB product, and the ItfBCD product were generated on a large scale, proteolyzed with GluC to reduce the length of the peptide, and subsequently subjected to 1D/2D NMR spectral analysis (Table S16 and Figures S6 and

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

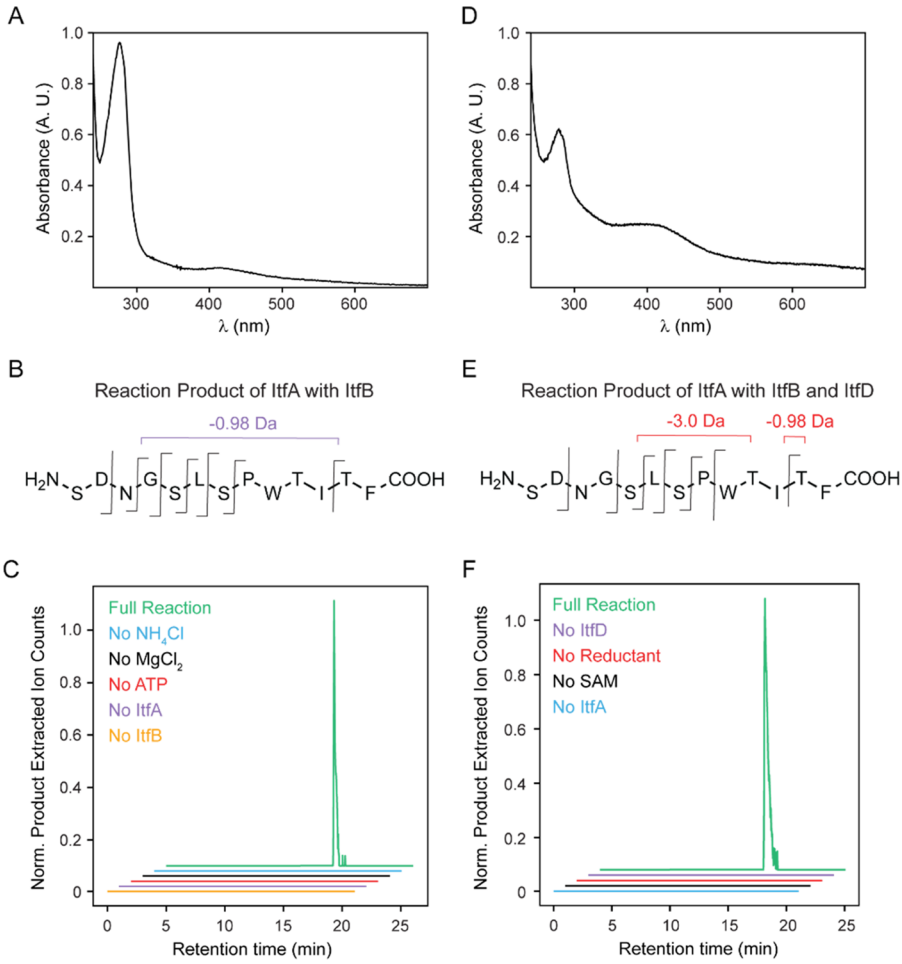


Figure 6. Characterization of ItfB and ItfD *in vitro*. (A) UV−vis absorption spectrum of purified ItfB. (B) HR-MS/MS analysis of the ItfB product; detected fragments are marked. (C) Enzymatic activity assays of ItfB via HPLC-Qtof-MS analysis. Shown are the extracted ion chromatograms for the −0.98 Da product of ItfB. (D) UV−vis absorption spectrum of purified ItfD. (E) HR-MS/MS analysis of the ItfD product; detected fragments are marked. (F) Enzymatic activity assays of ItfD via HPLC-Qtof-MS analysis. Shown are the extracted ion chromatograms for the −3 Da product of ItfD. In panels (C) and (F), traces are offset in both axes for clarity and color coded as indicated.

S8). A DEPT-edited HSQC spectrum of the ItfB product revealed no new peaks but a significant change in the chemical shifts assigned to Thr56 and Phe57. Specifically, the α-H of Thr56 shifted from 4.31 to 4.56 ppm, and the peaks assigned to the α-H and β-Hs of Phe57 shifted from 4.43 to 4.32 and 2.92/3.15 to 3.02/3.26, respectively (Figure 5A). All other peaks in the NMR spectrum remained unchanged between the substrate and the ItfB product. Together with the HR-MS data, these results suggest that ItfB converts the backbone amide between Thr56 and Phe57 to an amidine. Additionally, previously reported NMR data on peptides containing backbone amidines show similar changes to what we observed in which the N-terminal α-H shifts significantly downfield (as for Thr56), the C-terminal α-H shifts slightly upfield (as for Phe57), and the signal for the adjacent carbonyl group is split.[60] Members of the YcaO family have been previously shown to install backbone thioamides as well as cyclize peptides through amidine linkages. In this case, however, ItfB installs a backbone amidine, a new type of modification for the YcaO family but one that is analogous to the reactions that this enzyme family is known to catalyze.

We next investigated the structure of the ItfBCD product. Correlations related to the amidine were still present, indicating that it remained unchanged after reaction with ItfD. However, peaks assigned to Trp53 and Ile55 were significantly altered. In the HSQC spectrum, an aromatic proton of Trp53 was missing and the remaining peaks had shifted upfield. The TOCSY spectrum showed that only three of the aromatic protons were correlated with each other, identifying the missing proton as either C4 or C7 (Figure 5B). Comparison to the NMR data obtained for C-4- and C-7-modified indoles, such as streptide, pinpointed C7 as the site of modification. The peaks assigned to Ile55 in the substrate data had also shifted. Importantly, one of the peaks assigned as the terminal methyl proton on Ile55 was absent in the DEPT-edited HSQC spectrum, and a new methylene group appeared at 29.1 ppm (^{13}C) associated with shifts at 2.48/2.18 ppm (^{1}H) (Figure 5C). This methylene showed TOCSY correlations to the α-, β-, and γ-Hs of Ile, identifying it as the δ-Hs and implicating it as the site of the new crosslink to Trp53 (Figure 5D). Additionally, the Trp53 H-6 showed an HMBC correlation to the Ile δ-C and through-space ROSEY correlations to the Ile β-Hs and δ-Hs, thus confirming the C−C crosslink joining the Trp53 C-7 to the Ile55 δ-C (Figure 5E,F). Together these data show that ItfD installs a carbon−carbon crosslink between Trp53 and Ile55, giving rise to a 3-residue, 15-membered macrocycle (Figure 5G,H). While C−C crosslinks involving Trp are known to be installed by RaS

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

enzymes, previous reports have always involved crosslinks at more activated positions such as the Lys β-C, as was observed in streptide.[7] The methyl group of Ile is one of the least activated positions found in peptides, making this an impressive modification that functionalizes Ile for macrocycle formation. It represents a new reaction for RaS enzymes, further expanding their catalytic capabilities.

In Vitro Characterization of ItfB. We explored these unprecedented reactions further in vitro. N-terminally hexaHis-tagged ItfB was expressed recombinantly in *E. coli* and purified aerobically through metal affinity chromatography. The pure protein exhibited an orange color; a UV−visible absorption spectrum revealed a shallow, broad feature with λ_{max} at 415 nm, which is diagnostic for $[4Fe\text{-}4S]^{2+}$ clusters and is routinely observed for RaS enzymes (Figure 6A). Quantification of Fe and labile sulfide indeed showed 2.1 ± 0.1 Fe and 3.7 ± 0.4 S^{2-} per ItfB protomer (Table S17), consistent with the presence of a Fe−S cluster, although its exact nature remains to be determined.

A Fe−S cluster has not been observed in YcaO proteins before. We conducted a multiple sequence alignment of ItfB to other YcaO family members and found an additional C-terminal domain rich in Cys residues that was not present in previously characterized members of the family (Figure S9).[58,61−68] The Cys residues are conserved in a subset of the top ItfB BLAST hits including examples in what appear to be RiPP gene clusters not resembling the ITF subfamily as well as other examples of the ITF subfamily (Table S18 and Figure S10). An SSN of the entire Pfam describing YcaO-like proteins showed ItfB to belong to a unique subfamily that contains this C-terminal extension (Figure S11). ItfB is the first characterized member of this group of unique YcaO proteins, but more examples are likely to be assessed in the future. We conducted site-directed mutagenesis to change the four conserved Cys residues in the C-terminal domain to Ala; however, the mutant protein was not soluble and could not be isolated.

Because full-length ItfA is a 57mer, a truncated 30mer version of the substrate (ItfA*) was synthesized through solid-phase peptide synthesis to use for in vitro reactions with purified ItfB. The reactions were carried out in an inert atmosphere to avoid oxidative damage to the Fe−S cluster, although the enzyme was active aerobically for short periods immediately after purification. Based on previously reported YcaO reaction conditions, the 30mer peptide was incubated with ItfB, ATP, $MgCl_2$ as well as NH_4Cl to serve as a reduced nitrogen source for amidine formation. Under these conditions, we observed the time-dependent formation of a product with a mass loss of 0.98 Da, as determined by HR-MS (Table S19 and Figure S12). HR-MS/MS yielded a very similar fragmentation pattern as observed in the in vivo analysis of the reaction above (Figure 6B and Table S20). This product was not formed in the absence of ItfB, substrate, ATP, $MgCl_2$, or NH_4Cl (Figure 6C). Additionally, reaction with $^{15}NH_4Cl$ showed incorporation of ^{15}N into the product, confirming that NH_4^+ is the source of nitrogen for amidine formation (Table S21 and Figure S13). Together, these data establish that ItfB converts the ultimate amide bond between Thr56 and Phe57 into an amidine. The mechanism of this transformation is likely analogous to that of backbone thioamide formation previously reported for YcaO,[69] commencing with the activation of the amide-carbonyl via reaction with ATP followed by nucleophilic attack by NH_3, which could be activated through Lewis acid catalysis by the Fe−S cluster (Figure S14).

In Vitro Characterization of ItfD. We next investigated the RaS enzyme, ItfD, in vitro. It was expressed recombinantly as an N-terminally hexaHis-tagged construct in *E. coli* and subsequently purified anaerobically through metal affinity chromatography. As expected, ItfD exhibited distinctive UV−visible absorption features at 325 and 420 nm that is characteristic of Fe−S clusters (Figure 6D). It contained 14.1 ± 0.4 Fe and 19.0 ± 0.7 labile sulfide per protomer (Table S17). These data, along with a multiple sequence alignment to the RaS enzyme SuiB for which the crystal structure has been determined,[70] suggest that ItfD has two auxiliary [4Fe-4S] clusters (Figure S15). Alignment with SuiB also suggests that ItfD contains an RRE as an N-terminal fusion (Figure S16).

Reactions were carried out anaerobically by incubating ItfA* with *S*-adenosylmethionine (SAM), reductant, ItfD alone, or in combination with the small hypothetical protein, MBP-ItfC (recombinantly expressed as a 6HMBP-tagged protein in *E. coli*) and ItfB along with ATP, $MgCl_2$, and NH_4Cl. No products were observed when ItfD was incubated with ItfA* alone (in the presence of SAM and reductant). However, when ItfD was incubated with ItfA* in combination with ItfB (and the necessary small molecules), the −3 Da product was observed, suggesting that ItfD reacts only with the amidine-containing peptide (Table S19). HR-MS/MS analysis showed that this product had a very similar fragmentation pattern as observed from coexpression of the peptide and enzymes in vivo (Figure 6E and Table S22). Furthermore, this −3 Da product was formed in the presence or absence of MBP-ItfC; no additional products were observed when MBP-ItfC was included. The role of ItfC remains unknown at this time, but it does not seem to function as an RRE.

Since ItfD was only active with the amidine-containing peptide, we used the peptide produced from the coexpression of 6HMBP-ItfA and ItfB as a substrate for additional reactions with ItfD. These reactions showed that product was only formed in the presence of SAM, ItfD, substrate, and reductant (Figure 6F), and that both the product and 5′-dA appeared in a time-dependent manner with similar kinetics (Figure S17). We hypothesized that specific recognition of the amidine-containing peptide over the unmodified substrate may be achieved through the Lewis basicity of the former via ligation to auxiliary Fe−S clusters. Flühe et al have previously shown that ligation of an auxiliary Fe−S cluster by Cys-thiolate leads to suppression of the extinction coefficient and thereby to lowered absorption at the 400 nm feature.[71] Indeed, incubation of ItfD with varying concentrations of the amidine-containing peptide substrate (in the absence of SAM and reductant) resulted in concentration-dependent suppression of the absorption bands at 310 and 420 nm, which are associated with the $[4Fe\text{-}4S]^{2+}$ cluster (Figure S18). Importantly, the addition of similar concentrations of unmodified ItfA* did not lead to changes in these features. These data provide interesting initial clues regarding substrate recognition by ItfD.

Mechanistic Investigations of ItfD. To gain further mechanistic insights, we synthesized ItfA* with uniformly deuterated Ile ($^2H_{10}$-Ile) at the position of the crosslink and used it in a reaction with ItfB and ItfD. Analysis by HPLC-Qtof-MS showed a −4 Da shift in the product, indicating that deuterium is lost from Ile during the reaction (Tables S23 and S24). Additionally, the reaction generated deuterated 5′-dA as

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

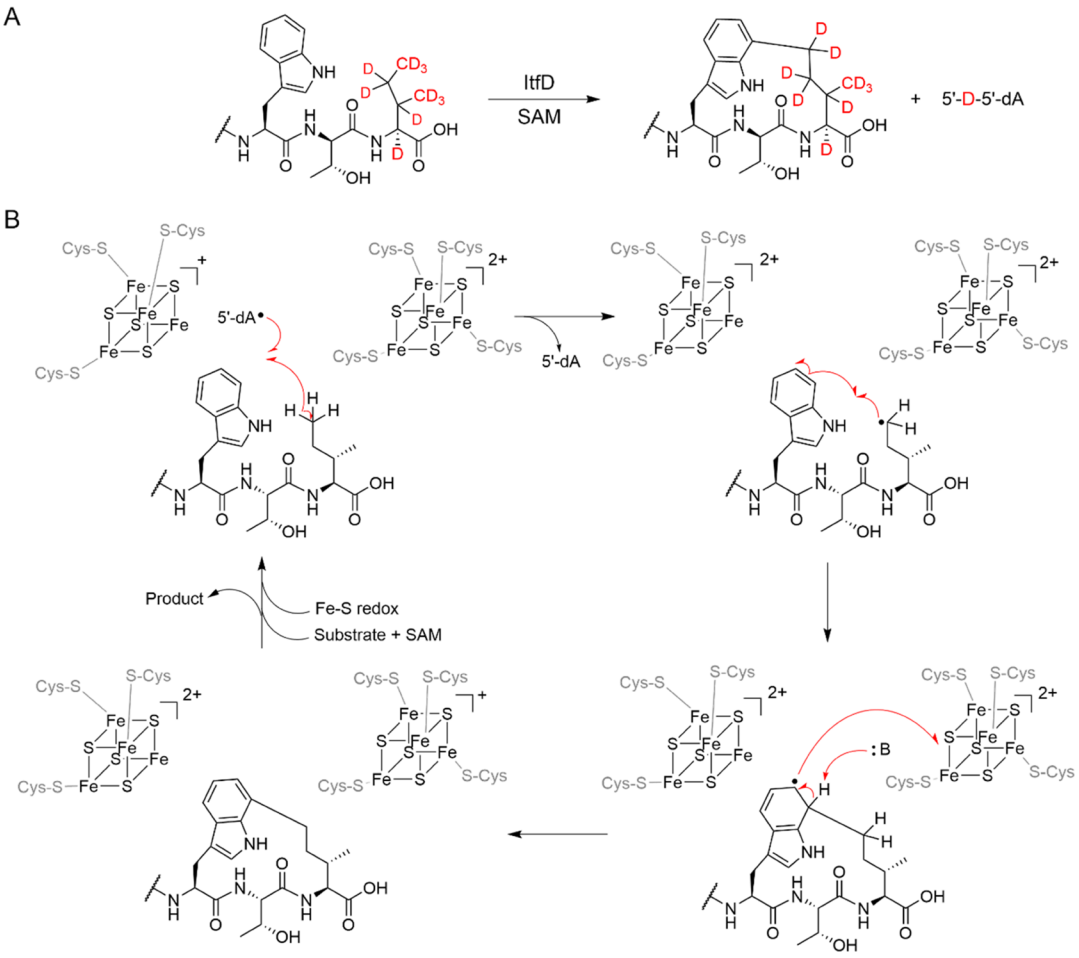


Figure 7. Mechanism of Ile-Trp crosslink formation by ItfD. (A) Reaction of ItfD with D_{10}-Ile55-ItfA yields a product that is 4 Da lighter than substrate and formation of 5′-D-5′-dA. (B) Proposed mechanism for Ile-Trp crosslink formation by ItfD. See text for details.

a product consistent with the transfer of the deuterium atom from Ile55 to 5′-dA•, implicating Ile55 as the site of H-atom abstraction (Figures 7A and S19). We propose that ItfD follows a mechanism analogous to that of SuiB (Figure 7B),[26,72,73] wherein the peptide is recognized by ligation of the amidine to an auxiliary Fe−S cluster in ItfD. Reductive cleavage of SAM leads to the formation of 5′-dA•, which abstracts a δ-H atom from Ile. This substrate radical then adds into the LUMO of the Trp53-indole to generate a cross-linked indoyl radical. Loss of a proton followed by rearomatization concomitant with reduction of an auxiliary Fe−S cluster completes the reaction. Resetting the redox states of the Fe−S clusters then allows the enzyme to carry out multiple turnovers.

CONCLUSIONS

With the rising number of available microbial genome sequences and the need for unexplored sources of natural products, developing new tools for RiPP discovery is paramount. While previously developed bioinformatics methods have performed exceptionally well in finding additional examples of RiPPs from known classes, the detection of new structural motifs has proved more difficult. Using a RaS enzyme−transporter co-occurrence approach, we present thousands of BGCs that can accelerate the discovery of entirely new classes of RiPPs. The approach does not rely on a high level of homology to a single tailoring enzyme family but instead focuses on co-occurrence of two broad protein families. Actinomycetes have been a prolific source of bioactive compounds from NRPS and PKS BGCs;[74] our results suggest that Firmicutes, Proteobacteria, and Bacteriodetes could prove to be rich sources of RiPP natural products that have thus far eluded detection. Moreover, the approach can be applied to other enzyme classes, as the DUF692 network demonstrates.

Characterization of a BGC within the ITF family led to the identification of two novel reactions and an intriguing combination of modifications, thereby providing proof of concept for the utility of the network. We show that ItfD can functionalize the unactivated δ-methyl group of an Ile side chain for macrocyclization, expanding the scope of reactivity beyond the terminal methyl group of Thr as in thermocellin biosynthesis.[44] Crosslinking to isoleucine is highly unusual, with a C−N crosslink from tryptophan to the β-carbon of isoleucine by a BURP-domain cyclase serving as the only example thus far.[75] Unexpectedly, in addition to Ile-Trp crosslink, we also identified backbone amidine formation as a new reaction for the YcaO family. ItfB proved divergent from YcaO family members examined thus far with a Fe−S cluster as its defining feature. The role for the Fe−S cluster remains to be determined. One possibility is the activation of ammonia for the nucleophilic attack via Lewis acid catalysis.

Our search for RaS-RiPPs in streptococci has already identified nearly a dozen new transformations in natural product biosynthesis. The work herein expands the original

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888

network to include all bacterial phyla and identifies hundreds of families of RaS-RiPPs, which surely will reveal exciting new RaS enzyme reactions and RiPP natural products for many years to come. We have only begun to scratch the surface of the diversity of RaS-RiPPs and the full scope of reactions that the incredible superfamily of RaS enzymes is able to catalyze.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.2c06497.

Complete materials & methods; sequence-similarity networks for RaS-RiPPs in the human microbiome, for *Clostridium* spp., and for the DUF692 superfamily; diagrams for selected biosynthetic gene clusters; HR-MS data, HR-MS/MS data, and NMR spectra for substrates/products of ItfB and ItfD; assays of ItfB with substrate isotopologs; and kinetic, spectroscopic, and substrate isotopolog assays of ItfD (PDF)

Cytoscape data (ZIP)

■ AUTHOR INFORMATION

Corresponding Author

Mohammad R. Seyedsayamdost − *Department of Chemistry and Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, United States;* orcid.org/0000-0003-2707-4854; Email: mrseyed@princeton.edu

Author

Kenzie A. Clark − *Department of Chemistry, Princeton University, Princeton, New Jersey 08544, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.2c06497

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Eli Lilly-Edward C. Taylor Fellowship in Chemistry (to K.A.C.), the National Science Foundation (NSF CAREER Award No. 1847932 to M.R.S.), and the National Institutes of Health (GM129496 to M.R.S.) for financial support.

■ REFERENCES

(1) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770−803.

(2) Bérdy, J. Bioactive Microbial Metabolites. *J. Antibiot.* **2005**, *58*, 1−26.

(3) Baldwin, J. E.; Abraham, E. The biosynthesis of Penicillins and Cephalosporins. *Nat. Prod. Rep.* **1988**, *5*, 29−45.

(4) Walsh, C. T.; Moore, B. S. Enzymatic Cascade Reactions in Biosynthesis. *Angew. Chem., Int. Ed.* **2019**, *58*, 6846−6879.

(5) Hubbard, B. K.; Walsh, C. T. Vancomycin Assembly: Nature's Way. *Angew. Chem., Int. Ed.* **2003**, *42*, 730−765.

(6) Katz, L.; Baltz, R. H. Natural Product Discovery: Past, Present, and Future. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 155−176.

(7) Bachmann, B. O.; Van Lanen, S. G.; Baltz, R. H. Microbial Genome Mining for Accelerated Natural Products Discovery: Is a Renaissance in the Making? *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 175−184.

(8) Clark, K. A.; Bushin, L. B.; Seyedsayamdost, M. R. RaS-RiPPs in Streptococci and the Human Microbiome. *ACS Bio Med. Chem. Au* **2022**, *2*, 328−339.

(9) Scott, T. A.; Piel, J. The Hidden Enzymology of Bacterial Natural Product Biosynthesis. *Nat. Rev. Chem.* **2019**, *3*, 404−425.

(10) Arnison, P. G.; Bibb, M. J.; Bierbaum, G.; Bowers, A. A.; Bugni, T. S.; Bulaj, G.; Camarero, J. A.; Campopiano, D. J.; Challis, G. L.; Clardy, J.; Cotter, P. D.; Craik, D. J.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P. C.; Entian, K.-D.; Fischbach, M. A.; Garavelli, J. S.; Göransson, U.; Gruber, C. W.; Haft, D. H.; Hemscheidt, T. K.; Hertweck, C.; Hill, C.; Horswill, A. R.; Jaspars, M.; Kelly, W. L.; Klinman, J. P.; Kuipers, O. P.; Link, A. J.; Liu, W.; Marahiel, M. A.; Mitchell, D. A.; Moll, G. N.; Moore, B. S.; Müller, R.; Nair, S. K.; Nes, I. F.; Norris, G. E.; Olivera, B. M.; Onaka, H.; Patchett, M. L.; Piel, J.; Reaney, M. J. T.; Rebuffat, S.; Ross, R. P.; Sahl, H.-G.; Schmidt, E. W.; Selsted, M. E.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R. D.; Tagg, J. R.; Tang, G.-L.; Truman, A. W.; Vederas, J. C.; Walsh, C. T.; Walton, J. D.; Wenzel, S. C.; Willey, J. M.; van der Donk, W. A. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* **2013**, *30*, 108−160.

(11) Montalbán-López, M.; Scott, T. A.; Ramesh, S.; Rahman, I. R.; van Heel, A.; Viel, J. H.; Bandarian, V.; Dittmann, E.; Genilloud, O.; Goto, Y.; Burgos, M. J. G.; Hill, C.; Kim, S.; Koehnke, J.; Latham, J. A.; Link, A. J.; Martínez, B.; Nair, S. K.; Nicolet, Y.; Rebuffat, S.; Sahl, H.-G.; Sareen, D.; Schmidt, E. W.; Schmitt, L.; Severinov, K.; Süssmuth, R. D.; Truman, A. W.; Wang, H.; Weng, J.-K.; van Wezel, G. P.; Zhang, Q.; Zhong, J.; Piel, J.; Mitchell, D. A.; Kuipers, O. P.; van der Donk, W. A. New Developments in RiPP Discovery, Enzymology and Engineering. *Nat. Prod. Rep.* **2021**, *38*, 130−239.

(12) McIntosh, J. A.; Donia, M. S.; Schmidt, E. W. Ribosomal Peptide Natural Products: Bridging the Ribosomal and Nonribosomal Worlds. *Nat. Prod. Rep.* **2009**, *26*, 537−559.

(13) Velásquez, J. E.; van der Donk, W. A. Genome Mining for Ribosomally Synthesized Natural Products. *Curr. Opin. Chem. Biol.* **2011**, *15*, 11−21.

(14) Frey, P. A.; Booker, S. J.Radical Mechanisms of S-Adenosylmethionine-Dependent Enzymes. In *Advances in Protein Chemistry*; Academic Press, 2001; Vol. *58*; pp 1−45.

(15) Landgraf, B. J.; McCarthy, E. L.; Booker, S. J. Radical S-Adenosylmethionine Enzymes in Human Health and Disease. *Annu. Rev. Biochem.* **2016**, *85*, 485−514.

(16) Broderick, J. B.; Duffus, B. R.; Duschene, K. S.; Shepard, E. M. Radical S-Adenosylmethionine Enzymes. *Chem. Rev.* **2014**, *114*, 4229−4317.

(17) Mehta, A. P.; Abdelwahed, S. H.; Mahanta, N.; Fedoseyenko, D.; Philmus, B.; Cooper, L. E.; Liu, Y.; Jhulki, I.; Ealick, S. E.; Begley, T. P. Radical S-Adenosylmethionine (SAM) Enzymes in Cofactor Biosynthesis: A Treasure Trove of Complex Organic Radical Rearrangement Reactions. *J. Biol. Chem.* **2015**, *290*, 3980−3986.

(18) Bassler, B. L.; Losick, R. Bacterially Speaking. *Cell* **2006**, *125*, 237−246.

(19) Whiteley, M.; Diggle, S. P.; Greenberg, E. P. Progress in and Promise of Bacterial Quorum Sensing Research. *Nature* **2017**, *551*, 313−320.

(20) Bushin, L. B.; Clark, K. A.; Pelczer, I.; Seyedsayamdost, M. R. Charting an Unexplored Streptococcal Biosynthetic Landscape Reveals a Unique Peptide Cyclization Motif. *J. Am. Chem. Soc.* **2018**, *140*, 17674−17684.

(21) Clark, K. A.; Bushin, L. B.; Seyedsayamdost, M. R. Aliphatic Ether Bond Formation Expands the Scope of Radical SAM Enzymes in Natural Product Biosynthesis. *J. Am. Chem. Soc.* **2019**, *141*, 10610−10615.

(22) Caruso, A.; Bushin, L. B.; Clark, K. A.; Martinie, R. J.; Seyedsayamdost, M. R. Radical Approach to Enzymatic β-Thioether Bond Formation. *J. Am. Chem. Soc.* **2019**, *141*, 990−997.

(23) Caruso, A.; Martinie, R. J.; Bushin, L. B.; Seyedsayamdost, M. R. Macrocyclization via an Arginine-Tyrosine Crosslink Broadens the

Reaction Scope of Radical S-Adenosylmethionine Enzymes. *J. Am. Chem. Soc.* **2019**, *141*, 16610−16614.
(24) Bushin, L. B.; Covington, B. C.; Rued, B. E.; Federle, M. J.; Seyedsayamdost, M. R. Discovery and Biosynthesis of Streptosactin, a Sactipeptide with an Alternative Topology Encoded by Commensal Bacteria in the Human Microbiome. *J. Am. Chem. Soc.* **2020**, *142*, 16265−16275.
(25) Caruso, A.; Seyedsayamdost, M. R. Radical SAM Enzyme QmpB Installs Two 9-Membered Ring Sactionine Macrocycles during Biogenesis of a Ribosomal Peptide Natural Product. *J. Org. Chem.* **2021**, *86*, 11284−11289.
(26) Schramma, K. R.; Bushin, L. B.; Seyedsayamdost, M. R. Structure and Biosynthesis of a Macrocyclic Peptide Containing an Unprecedented Lysine-to-Tryptophan Crosslink. *Nat. Chem.* **2015**, *7*, 431−437.
(27) Rued, B. E.; Covington, B. C.; Bushin, L. B.; Szewczyk, G.; Laczkovich, I.; Seyedsayamdost, M. R.; Federle, M. J. Quorum Sensing in Streptococcus Mutans Regulates Production of Tryglysin, a Novel RaS-RiPP Antimicrobial Compound. *mBio* **2021**, *12*, No. e02688-20.
(28) Bushin, L. B.; Covington, B. C.; Clark, K. A.; Caruso, A.; Seyedsayamdost, M. R. Bicyclostreptins Are Radical SAM Enzyme-Modified Peptide Natural Products with Unique Cyclization Motifs. *Nat. Chem. Biol.* **2022**, No. 8.
(29) Clark, K. A.; Covington, B. C.; Seyedsayamdost, M. R.*Biosynthesis-Guided Discovery of Enteropeptins, Unusual Sactipeptides Containing an N-Methylornithine*; ChemRxiv, 2021.
(30) Haft, D. H. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. *BMC Genomics* **2011**, *12*, No. 21.
(31) Haft, D. H.; Basu, M. K. Biological systems discovery in silico: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. *J. Bacteriol.* **2011**, *193*, 2745−2755.
(32) Zallot, R.; Oberg, N.; Gerlt, J. A. The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. *Biochemistry* **2019**, *58*, 4169−4182.
(33) Gerlt, J. A. Genomic Enzymology: Web Tools for Leveraging Protein Family Sequence−Function Space and Genome Context to Discover Novel Functions. *Biochemistry* **2017**, *56*, 4293−4308.
(34) Oberg, N.; Precord, T. W.; Mitchell, D. A.; Gerlt, J. A. RadicalSAM.org: A Resource to Interpret Sequence-Function Space and Discover New Radical SAM Enzyme Chemistry. *ACS Bio Med. Chem. Au* **2022**, *2*, 22−35.
(35) Kostenko, A.; Lien, Y.; Mendauletova, A.; Hgendahimana, T.; Novitskiy, I. M.; Eaton, S. S.; Latham, J. A. Identification of a polycyclopropylglycine-containing peptide via bioinformatic mapping of radical S-adenosylmethionine enzymes. *J. Biol. Chem.* **2022**, *298*, No. 101881.
(36) Sugiyama, R.; Suarez, A. F. L.; Morishita, Y.; Nguyen, T. Q. N.; Tooh, Y. W.; Roslan, M. N. H. B.; Choy, J. L.; Su, Q.; Goh, W. Y.; Gunawan, G. A.; Wong, F. T.; Morinaka, B. I. The Biosynthetic Landscape of Triceptides Reveals Radical SAM Enzymes That Catalyze Cyclophane Formation on Tyr- and His-Containing Motifs. *J. Am. Chem. Soc.* **2022**, *144*, 11580−11593.
(37) Tsukazaki, T. Structure-Based Working Model of SecDF, a Proton-Driven Bacterial Protein Translocation Factor. *FEMS Microbiol. Lett.* **2018**, *365*, No. fny112.
(38) Håvarstein, L. S.; Diep, D. B.; Nes, I. F. A Family of Bacteriocin ABC Transporters Carry out Proteolytic Processing of Their Substrates Concomitant with Export. *Mol. Microbiol.* **1995**, *16*, 229−240.
(39) Quistgaard, E. M.; Löw, C.; Guettou, F.; Nordlund, P. Understanding Transport by the Major Facilitator Superfamily (MFS): Structures Pave the Way. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 123−132.
(40) Kuroda, T.; Tsuchiya, T. Multidrug Efflux Transporters in the MATE Family. *Biochim. Biophys. Acta, Proteins Proteomics* **2009**, *1794*, 763−768.
(41) Singh, M.; Sareen, D. Novel LanT associated lantibiotic clusters identified by genome database mining. *PLoS One* **2014**, *9*, No. e91352.
(42) Grell, T. A. J.; Goldman, P. J.; Drennan, C. L. SPASM and Twitch Domains in S-Adenosylmethionine (SAM) Radical Enzymes. *J. Biol. Chem.* **2015**, *290*, 3964−3971.
(43) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498−2504.
(44) Hudson, G. A.; Burkhart, B. J.; DiCaprio, A. J.; Schwalen, C. J.; Kille, B.; Pogorelov, T. V.; Mitchell, D. A. Bioinformatic Mapping of Radical S-Adenosylmethionine-Dependent Ribosomally Synthesized and Post-Translationally Modified Peptides Identifies New Cα, Cβ, and Cγ-Linked Thioether-Containing Peptides. *J. Am. Chem. Soc.* **2019**, *141*, 8228−8238.
(45) Santos-Aberturas, J.; Chandra, G.; Frattaruolo, L.; Lacret, R.; Pham, T. H.; Vior, N. M.; Eyles, T. H.; Truman, A. W. Uncovering the Unexplored Diversity of Thioamidated Ribosomal Peptides in Actinobacteria Using the RiPPER Genome Mining Tool. *Nucleic Acids Res.* **2019**, *47*, 4624−4637.
(46) De Los Santos, E. L. C. NeuRiPP: Neural Network Identification of RiPP Precursor Peptides. *Sci. Rep.* **2019**, *9*, No. 13406.
(47) Merwin, N. J.; Mousa, W. K.; Dejong, C. A.; Skinnider, M. A.; Cannon, M. J.; Li, H.; Dial, K.; Gunabalasingam, M.; Johnston, C.; Magarvey, N. A. DeepRiPP Integrates Multiomics Data to Automate Discovery of Novel Ribosomally Synthesized Natural Products. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 371−380.
(48) Krebs, C.; Galonić Fujimori, D.; Walsh, C. T.; Bollinger, J. M. Non-Heme Fe(IV)−Oxo Intermediates. *Acc. Chem. Res.* **2007**, *40*, 484−492.
(49) Le, K. Y.; Otto, M. Quorum-Sensing Regulation in Staphylococci—an Overview. *Front. Microbiol.* **2015**, *6*, No. 1174.
(50) Proctor, L. M.; Creasy, H. H.; Fettweis, J. M.; Lloyd-Price, J.; Mahurkar, A.; Zhou, W.; Buck, G. A.; Snyder, M. P.; Strauss, J. F.; Weinstock, G. M.; White, O.; Huttenhower, C. The Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project. *Nature* **2019**, *569*, 641−648.
(51) Turnbaugh, P. J.; Ley, R. E.; Hamady, M.; Fraser-Liggett, C. M.; Knight, R.; Gordon, J. I. The Human Microbiome Project. *Nature* **2007**, *449*, 804−810.
(52) Kulagina, E. V.; Efimov, B. A.; Maximov, P. Y.; Kafarskaia, L. I.; Chaplin, A. V.; Shkoporov, A. N. Species Composition of Bacteroidales Order Bacteria in the Feces of Healthy People of Various Ages. *Biosci., Biotechnol., Biochem.* **2012**, *76*, 169−171.
(53) Yadav, J. P.; Das, S. C.; Dhaka, P.; Vijay, D.; Kumar, M.; Mukhopadhyay, A. K.; Chowdhury, G.; Chauhan, P.; Singh, R.; Dhama, K.; Malik, S. V. S.; Kumar, A. Molecular Characterization and Antimicrobial Resistance Profile of *Clostridium perfringens* Type A Isolates from Humans, Animals, Fish and Their Environment. *Anaerobe* **2017**, *47*, 120−124.
(54) Kenney, G. E.; Dassama, L. M. K.; Pandelia, M.-E.; Gizzi, A. S.; Martinie, R. J.; Gao, P.; DeHart, C. J.; Schachner, L. F.; Skinner, O. S.; Ro, S. Y.; Zhu, X.; Sadek, M.; Thomas, P. M.; Almo, S. C.; Bollinger, J. M.; Krebs, C.; Kelleher, N. L.; Rosenzweig, A. C. The Biosynthesis of Methanobactin. *Science* **2018**, *359*, 1411−1416.
(55) Ting, C. P.; Funk, M. A.; Halaby, S. L.; Zhang, Z.; Gonen, T.; van der Donk, W. A. Use of a Scaffold Peptide in the Biosynthesis of Amino Acid Derived Natural Products. *Science* **2019**, *365*, 280−284.
(56) Wexler, H. M. Bacteroides: The Good, the Bad, and the Nitty-Gritty. *Clin. Microbiol. Rev.* **2007**, *20*, 593−621.
(57) Eckburg, P. B.; Bik, E. M.; Bernstein, C. N.; Purdom, E.; Dethlefsen, L.; Sargent, M.; Gill, S. R.; Nelson, K. E.; Relman, D. A. Diversity of the Human Intestinal Microbial Flora. *Science* **2005**, *308*, 1635−1638.

(58) Burkhart, B. J.; Schwalen, C. J.; Mann, G.; Naismith, J. H.; Mitchell, D. A. YcaO-Dependent Posttranslational Amide Activation: Biosynthesis, Structure, and Function. *Chem. Rev.* **2017**, *117*, 5389−5456.
(59) Kloosterman, A. M.; Shelton, K. E.; van Wezel, G. P.; Medema, M. H.; Mitchell, D. A. RRE-Finder: A Genome-Mining Tool for Class-Independent RiPP Discovery. *mSystems* **2020**, *5*, No. e00267-20.
(60) Foden, C. S.; Islam, S.; Fernández-García, C.; Maugeri, L.; Sheppard, T. D.; Powner, M. W. Prebiotic Synthesis of Cysteine Peptides That Catalyze Peptide Ligation in Neutral Water. *Science* **2020**, *370*, 865−869.
(61) Liu, A.; Si, Y.; Dong, S.-H.; Mahanta, N.; Penkala, H. N.; Nair, S. K.; Mitchell, D. A. Functional Elucidation of TfuA in Peptide Backbone Thioamidation. *Nat. Chem. Biol.* **2021**, *17*, 585−592.
(62) Mahanta, N.; Liu, A.; Dong, S.; Nair, S. K.; Mitchell, D. A. Enzymatic Reconstitution of Ribosomal Peptide Backbone Thioamidation. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115*, 3030−3035.
(63) Franz, L.; Adam, S.; Santos-Aberturas, J.; Truman, A. W.; Koehnke, J. Macroamidine Formation in Bottromycins Is Catalyzed by a Divergent YcaO Enzyme. *J. Am. Chem. Soc.* **2017**, *139*, 18158−18161.
(64) Franz, L.; Kazmaier, U.; Truman, A. W.; Koehnke, J. Bottromycins - Biosynthesis, Synthesis and Activity. *Nat. Prod. Rep.* **2021**, *38*, 1659−1683.
(65) Travin, D. Y.; Metelev, M.; Serebryakova, M.; Komarova, E. S.; Osterman, I. A.; Ghilarov, D.; Severinov, K. Biosynthesis of Translation Inhibitor Klebsazolicin Proceeds through Heterocyclization and N-Terminal Amidine Formation Catalyzed by a Single YcaO Enzyme. *J. Am. Chem. Soc.* **2018**, *140*, 5625−5633.
(66) Russell, A. H.; M Vior, N.; S Hems, E.; Lacret, R.; W Truman, A. Discovery and Characterisation of an Amidine-Containing Ribosomally-Synthesised Peptide That Is Widely Distributed in Nature. *Chem. Sci.* **2021**, *12*, 11769−11778.
(67) Schwalen, C. J.; Hudson, G. A.; Kille, B.; Mitchell, D. A. Bioinformatic Expansion and Discovery of Thiopeptide Antibiotics. *J. Am. Chem. Soc.* **2018**, *140*, 9494−9501.
(68) Frattaruolo, L.; Lacret, R.; Cappello, A. R.; Truman, A. W. A Genomics-Based Approach Identifies a Thioviridamide-Like Compound with Selective Anticancer Activity. *ACS Chem. Biol.* **2017**, *12*, 2815−2822.
(69) Dong, S.-H.; Liu, A.; Mahanta, N.; Mitchell, D. A.; Nair, S. K. Mechanistic Basis for Ribosomal Peptide Backbone Modifications. *ACS Cent. Sci.* **2019**, *5*, 842−851.
(70) Davis, K. M.; Schramma, K. R.; Hansen, W. A.; Bacik, J. P.; Khare, S. D.; Seyedsayamdost, M. R.; Ando, N. Structures of the Peptide-modifying Radical SAM Enzyme SuiB Elucidate the Basis of Substrate Recognition. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 10420−10425.
(71) Flühe, L.; Knappe, T. A.; Gattner, M. J.; Schäfer, A.; Burghaus, O.; Linne, U.; Marahiel, M. A. The Radical SAM Enzyme AlbA Catalyzes Thioether Bond Formation in Subtilosin A. *Nat. Chem. Biol.* **2012**, *8*, 350−357.
(72) Schramma, K. R.; Forneris, C. C.; Caruso, A.; Seyedsayamdost, M. R. Mechanistic Investigations of Lysine−Tryptophan Cross-Link Formation Catalyzed by Streptococcal Radical S-Adenosylmethionine Enzymes. *Biochemistry* **2018**, *57*, 461−468.
(73) Balo, A. R.; Caruso, A.; Tao, L.; Tantillo, D. J.; Seyedsayamdost, M. R.; Britt, R. D. Trapping a Cross-Linked Lysine−Tryptophan Radical in the Catalytic Cycle of the Radical SAM Enzyme SuiB. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2101571118.
(74) Fischbach, M. A.; Walsh, C. T. Assembly-line Enzymology for Polyketide and Nonribosomal Peptide Antibiotics: Logic, Machinery, and Mechanism. *Chem. Rev.* **2006**, *106*, 3468−3496.
(75) Chigumba, D. N.; Mydy, L. S.; de Waal, F.; Li, W.; Shafiq, K.; Wotring, J. W.; Mohamed, O. G.; Mladenovic, T.; Tripathi, A.; Sexton, J. Z.; Kautsar, S.; Medema, M. H.; Kersten, R. D. Discovery and Biosynthesis of Cyclic Plant Peptides via Autocatalytic Cyclases. *Nat. Chem. Biol.* **2022**, *18*, 18−28.

https://doi.org/10.1021/jacs.2c06497
J. Am. Chem. Soc. 2022, 144, 17876−17888