HHS Public Access

Author manuscript

J Am Chem Soc. Author manuscript; available in PMC 2020 May 22.

Published in final edited form as:

J Am Chem Soc. 2019 May 22; 141(20): 8228–8238. doi:10.1021/jacs.9b01519.

Bioinformatic mapping of radical SAM-dependent RiPPs identifies new Cα, Cβ, and Cγ-linked thioether-containing peptides

Graham A. Hudson[†,§], **Brandon J. Burkhart**[†,‡,§], **Adam J. DiCaprio**[†,‡], **Christopher J. Schwalen**[†], **Bryce Kille**[†], **Taras V. Pogorelov**[†,¶,§,⊥,‖], and **Douglas A. Mitchell**[†,‡,*]

[†]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

[‡]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

[¶]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

[§]School of Chemical Sciences, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

[⊥]National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

[‖]Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA.

Abstract

Recently developed bioinformatic tools have bolstered the discovery of ribosomally synthesized and posttranslationally modified peptides (RiPPs). Using an improved version of Rapid ORF Description & Evaluation Online (RODEO 2.0), a biosynthetic gene cluster mining algorithm, we bioinformatically mapped the sactipeptide RiPP class via the radical *S*-adenosylmethionine (SAM) enzymes that form the characteristic sactionine (sulfur-to-alpha carbon) crosslinks between cysteine and acceptor residues. Hundreds of new sactipeptide biosynthetic gene clusters were uncovered and a novel sactipeptide "huazacin" with growth-suppressive activity against *Listeria monocytogenes* was characterized. Bioinformatic analysis further suggested that a group of sactipeptide-like peptides heretofore referred to as SCIFFs (six cysteines in forty-five residues) might not be sactipeptides as previously thought. Indeed, the bioinformatically-identified SCIFF

[*]corresponding author, Phone: 1-217-333-1345; Fax: 1-217-333-0508; douglasm@illinois.edu, Address: 600 South Mathews Avenue, Urbana, Illinois 61801, USA.

[§]These authors contributed equally.

The authors declare no competing financial interest.

ASSOCIATED CONTENT

The Supporting Information is available free of charge on the ACS Publications website at DOI: [TO BE INSERTED] and contains a description of bioinformatic and experimental methods, and the supporting spectroscopic data collected for freyrasin and huazacin (Tables S1–S7 and Figures S1–S25. Additionally, bioinformatics output from RODEO may be found in Supplemental Dataset 1 and predicted precursor peptides in Supplemental Dataset 2.

peptide “freyrasin” was demonstrated to contain six thioethers linked the beta carbons of six aspartate residues. Another SCIFF, thermocellin, was shown to contain a thioether crosslinked to the gamma carbon of threonine. SCIFFs feature a different paradigm of non-alpha carbon thioether linkages and they are exclusively formed by radical SAM enzymes, as opposed to the polar chemistry employed during lanthipeptide biosynthesis. Therefore, we propose the renaming of the SCIFF family as radical non-alpha thioether peptides (ranthipeptides) to better distinguish them from the sactipeptide and lanthipeptide RiPP classes.

Introduction

With the growing availability of genome sequences, bioinformatics has become an increasingly popular and powerful technique for natural product discovery.[1,2] The biosynthetic gene clusters (BGCs) of known natural product classes are readily identified by the presence of conserved genes and the structure of their products can be predicted to varying degrees of accuracy based on the number and type of enzymes locally encoded.[3] This approach has been particularly useful in the discovery of ribosomally synthesized and posttranslationally modified peptides (RiPPs).[4] RiPPs have no universally conserved gene but instead are united by a common biosynthetic logic: with few known exceptions, the biosynthetic enzymes bind their respective precursor peptides through a recognition sequence in the N-terminal region of the peptide, referred to as the leader peptide,[5] and the posttranslational modifications are installed on the C-terminal region, referred to as the core peptide. During biosynthetic maturation, the leader peptide is eventually removed, and frequently additional enzymatic tailoring of the peptide occurs to yield the final RiPP product. Several posttranslational chemistries are known and are used to categorize RiPPs into their respective classes.[4,6]

Despite the utility of genome mining to discover new natural products, the identification of RiPP BGCs remains challenging. First, RiPPs require analysis of a local genomic context for the presence of encoded regulators and transporters, or better yet, clear supporting genetic markers of RiPP biosynthetic pathways because RiPP enzymes are often members of larger protein superfamilies that bear homology to proteins not associated with natural product biosynthesis. Second, RiPP precursor peptides are often short and hypervariable, which means their coding sequences are often unannotated when deposited into public databases and they are often not identified by sequence homology search tools (e.g., BLAST[7]). To address these challenges, we developed a bioinformatics program, Rapid ORF Description & Evaluation Online (RODEO), which automates the genome mining process.[8] A typical RODEO input will be a list of NCBI gene accession identifiers for a biosynthetic protein of interest. The program then fetches genomic records from GenBank and uses the PFAM[9] and/or TIGRFAM[10] databases to predict the function of neighboring genes. Next, RODEO identifies all possible precursor peptides and scores their likelihood of being a true RiPP precursor using a scoring function based on motif analysis and supervised machine learning. Given the disparate nature of RiPP precursor peptides, the scoring function is optimized for each distinct class. In our initial report, RODEO was leveraged to identify and classify lasso peptides which led to an order of magnitude expansion of the RiPP class. RODEO has also been reconfigured to aid in the discovery of thiopeptides[11] and was incorporated into

antiSMASH (version 4.0),[12] a tool which mines single genome inputs for natural products. Once the requisite sequence attributes and machine learning classifications are established, performing updates for a RiPP class via RODEO is straightforward, as demonstrated very recently with the lasso peptides.[13]

Given the success of RODEO in defining other RiPP classes, we sought to expand its capabilities to aid in the discovery of new sactipeptides. Sactipeptides have garnered interest due to their unique hairpin structures and potent, narrow-spectrum activity towards several human pathogenic bacteria.[14–16] They are defined by a radical SAM (rSAM) enzyme that forms "sactionine" crosslinks between a donor Cys sulfur and the alpha carbon of an acceptor amino acid (Figure 1).[15,17,18] Prevalent in RiPP pathways, rSAM enzymes are versatile catalysts that employ [4Fe-4S] clusters to reductively cleave SAM to generate a 5′-deoxyadenosyl radical.[19–23] This reactive intermediate often will abstract a hydrogen atom from the substrate, which leads to a variety of different outcomes based on the specifics of a given biosynthetic pathway. The precise mechanistic details of sactipeptide crosslink formation remains poorly characterized, although several groups are taking on this challenge.[24,25] Studies of the sactipeptide rSAM enzyme itself have revealed that sactisynthases contain three [4Fe-4S] clusters: one which reductively cleaves SAM and two auxiliary clusters in what has been called dubbed a SPASM domain.[26–29] Only five peptides with the hallmark S–Cα crosslinks have been isolated thus far: subtilosin A,[30, 31] thurincin H,[30] thuricin CD (two peptide products),[16] and sporulation killing factor.[31,32] A sixth sactipeptide, hyicin 4244, was recently reported but was not isolated for detailed characterization.[33] All structurally characterized sactipeptides contain ring-within-a-ring topologies and thus form rigid, hairpin-like structures (Figure 1). However, in 2011, Haft reported the bioinformatic discovery of a related class of putative thioether-crosslinked peptides, which he referred to as the six Cys in forty-five residue peptides (SCIFFs).[34] No bona fide SCIFFs have been isolated, but in vitro enzymatic reactions with the purified rSAM and precursor peptide indicated that a crosslink is formed between a donor Cys and an acceptor Thr residue.[24,25]

Similar to our earlier work that redefined the genomic landscape of the lasso peptides and thiopeptides, we herein conducted a comprehensive analysis of the sactipeptide RiPP class using a re-written and improved edition of RODEO, version 2.0. This release of RODEO features numerous improvements, including reduction of redundant calculations, more robust handling of internet connectivity and genomic record formatting errors, and enhanced multithreading that enables parallelization of data fetching from GenBank and subsequent processing, collectively leading to a large increase in processing speed. Although previous bioinformatic surveys have been conducted on the sactipeptide class,[35–39] none have employed machine learning to cast a wider net for identifying candidate precursor peptides. In this study, we have conducted a comprehensive analysis and connected 3,156 rSAM enzymes to their cognate RiPP precursor peptide and found new varieties of thioether-linked natural products. Guided by this dataset, we selected a new sactipeptide, huazacin, for characterization that contains a thioether ring architecture larger than any sactipeptide previously characterized. Additionally, our bioinformatic analysis indicated that SCIFF-associated rSAM maturases were much more closely related to QhpD proteins than characterized sactisynthases. QhpD is a rSAM enzyme which is known to form β- and γ-

linked thioethers,[40] thus calling into question the linkage architecture within a SCIFF natural product. We pursued this line of inquiry resulting in the bioinformatic identification and structural characterization of two SCIFF natural products. The first, freyrasin, features six S-Cβ thioether linkages. The second, thermocellin, has previously been reported in the literature and characterized only after partial in vitro enzymatic reconstitution. We have confirmed thermocellin to feature a Cys-Thr thioether as reported, but rather than harboring a S–Cα linkage, the peptide contains a S–Cγ thioether linkage. To distinguish this group from the structurally distinct sactipeptides, we propose the renaming of this RiPP class as the ranthipeptides, for radical non-alpha thioether peptides.

Results and Discussion

RODEO2.0-enabled sactipeptide discovery

With >450,000 rSAM proteins listed in the InterPro 72.0 database (IPR007197),[41] a more focused bioinformatic search was performed to winnow down candidates more likely to be involved in sactipeptide biosynthesis. Therefore, four rounds of iterative PSI-BLAST (Position-Specific Iterated Basic Local Alignment Search Tool[42]) searches using rSAM sequences from four known sactipeptides (AlbA, SkfB, ThnB, and TrnC)[15] and two partially characterized SCIFFs (CteB[25] and Tte1186[24]) were performed to identify candidate proteins. A stringent expectation value of 1e-70 was used as a cut-off to assist in keeping the retrieved candidate proteins more likely to be involved in RiPP biosynthesis in addition to filtering sequences <300 amino acids. This procedure yielded ~4,600 protein sequences, which were subsequently analyzed by RODEO2.0 (see Supplementary Methods; output available in Supplementary Dataset 1), to annotate the local genomic context and score potential sactipeptide precursor sequences (Supplementary Dataset 2).

Sactipeptide precursor peptides were identified using a series of heuristics in conjunction with support vector machine classification as previously reported resulting in 504 candidate precursor sequences.[12] The top scoring sequences were then manually inspected for similarity to known sactipeptides. In general, a peptide was considered to be a strong candidate sactipeptide if it had 3 or more Cys separated by equal spacing of 1 to 3 residues. Additionally, peptides were scored more favorably if the Cys residues were clustered within the N-terminal half of the core peptide, similar to that observed for other sactipeptides (i.e. the donor residues line one side of the hairpin while the receptor residues line the other side, Figure 1). Precursor peptides were further analyzed by a sequence similarity network (SSN) generated using the Enzyme Function Initiative Enzyme Similarity Tool (EFI-EST)[43] which revealed that clustering of precursors appears to be largely driven by conservation of donor Cys residues (Figure S1), which is predicted to govern the location of crosslinks and thus the overall structure of the sactipeptide. This bioinformatic analysis identified four significantly populated groups of uncharacterized sactipeptides. Two groups, named hypervariable groups 1 and 2, possess precursor peptides that are more than double the length of any known sactipeptide. While the core sequences within these two groups are overall “hypervariable”, there are 6–8 locations that display highly conserved Cys residues. A third uncharacterized group of sactipeptides are concentrated in the *Lachnospiraceae* family, thus we have termed these predicted compounds the “lachnocins” (Figure S1).

We elected, based on the above bioinformatics results, to pursue the fourth uncharacterized sactipeptide group. The target producer chosen was *Bacillus thuringiensis* serovar *huazhongensis* (Figure 2), owing to strain availability from the Bacillus Genetic Stock Center (BGSC 4BD1) and the fact that the predicted structure contained a unique crosslinked structure (4 crosslinks separated by 3 residues each). Although bioinformatically predicted in a previous study,[39] the mature natural product has never been reported. Upon screening various culture extracts by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS), we observed a mass consistent with the predicted structure (Figure S2). The metabolite, hereafter huazacin, was then subjected to high-resolution and tandem MS (HR-MS/MS) analysis using an Orbitrap instrument. The observed *m/z* value ($[M+3H]^{3+}$ = 1241.5396) agreed with the theoretical mass of the expected molecular formula of $C_{160}H_{241}N_{43}O_{50}S_5^{3+}$ ($[M+3H]^{3+}$ = 1241.5393; error = 0.2 ppm, Figures S2 **and** S3). This formula contains four thioether linkages (8 Da lighter than the unmodified core peptide). Upon subjecting the parent ion to collision-induced dissociation (CID) conditions, we observed robust fragmentation of the amide and thioether linkages (Figure S3). Similar to other characterized sactipeptides, this gas-phase dissociation of thioethers enabled the assignment of acceptor residues and was consistent with the thioether linkages of huazacin being S-Cα linked (i.e., sactionines), as predicted.[44]

Multidimensional NMR spectroscopy (^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY) were collected on HPLC-purified huazacin, which corroborated the sactionine (S-Cα) linkages of huazacin (Figures S4–S5, Table S3). We observed NOESY correlations for the Cys-CαH (and Cys-CβH) to the recipient site backbone NH (Figure S6). Further, no TOCSY correlations were observed from NH to CαH, which was consistent with a quaternary carbon at the alpha position as expected for a sactionine linkage (Figure S7). Ring topology was also established by NOESY analysis of the Cys12-Thr30 and Cys8-Tyr34 rings.

We next assessed the antibacterial activity of huazacin by testing against a diverse panel of bacteria for growth suppression using the microbroth dilution method. Reported sactipeptides typically have a narrow spectrum of activity towards Gram-positive organisms, most notably *Clostridium difficile*, *Listeria monocytogenes*, and *Bacillus cereus*.[16,17,46–49] Consistent with previously characterized sactipeptides, huazacin inhibited the growth of *L. monocytogenes* 4b F2365 with a minimum inhibitory concentration (MIC) of 4 μM (Table S4). These results with huazacin reinforce the concept that RODEO is a useful tool to accelerate RiPP discovery. While many additional, predicted sactipeptides are found in less populated clusters (with many being singletons, Figure S1), these data provide a roadmap to guide for future sactipeptide discovery efforts.

Mining the rSAM-modified RiPP thioether genomic landscape

Having shown the utility of RODEO for the discovery of a new sactipeptide, we sought to more broadly survey the genomic landscape of rSAM-modified RiPPs that were likely to contain thioether linkages. We began this effort by visualizing the sequence similarity of our list of rSAM enzymes (from PSI-BLAST) in the form of a SSN (Figure 3). Sequences with >70% amino acid identity were conflated to a single node and connected by an edge if significant sequence similarity was shared between the two sequences (expectation value <

1e-50). Thus, only highly similar rSAM proteins form clusters by this analysis. Given that the similarly of a RiPP modifying enzyme strongly correlates with the similarity of the cognate precursor peptides (i.e. similar rSAM enzymes produce similar products),[8] a properly annotated SSN provides a simplified but informative view of the peptide substrate diversity.

To provide context for the rSAM SSN, the nodes were colored based on the sequence characteristics of the cognate precursor peptide identified by RODEO (Figure 3). Several of the clusters were readily annotated by similarity to known sactipeptide precursor peptides, such as subtilosin (SboA), thurincin H (ThnA), thuricin CD (TrnA), and sporulation killing factor (SkfA). The novel sactipeptide huazacin (HuaA) described above is also indicated on the network. Two outlier sequences, the thermocellins (CteA and Tte1186a), belong to the so-called SCIFF family and are contained in a much larger separate group. Additionally, two large groupings of uncharacterized rSAM proteins with low-level similarity to TrnC were readily identified. The RODEO-identified precursor peptides encoded next to these rSAM proteins comprise the above-mentioned hypervariable groups 1 and 2 (Supplementary Dataset 2).

Several trends for the number and position of Cys residues within the peptides were further analyzed to gain deeper insight into the RiPP thioether landscape (Figure S8). The positioning of the Cys residues indicated that many groups had the potential to form the traditional sactipeptide hairpin-like topology wherein Cys residues are concentrated near the N-terminus and crosslink to acceptor residues concentrated near the C-terminus. Only the SkfA and SCIFF groups seemed capable of adopting an alternative structural topology based on their Cys residues being more evenly distributed (Figure S8). However, SKF is known to be N-to-C macrocyclized and it is the only known sactipeptide to contain an internal disulfide bond. No mature SCIFFs have been reported in the literature, thus their topologies have remained undetermined. The positioning of the Cys residues would suggest SCIFFs display non-hairpin topologies (Figures 1 and S9).

The majority of the identified thioether-forming rSAM enzymes reside within the large SCIFF cluster. Despite their greater numbers, the cognate precursor peptides display low sequence diversity. The RiPP products of these BGCs have been termed the thermocellins, although the final structure of thermocellin has not been reported. We identified ~2600 SCIFF precursor peptides (1,023 unique sequences), which outnumber the RODEO-identified sactipeptides by three-fold (Figure S9 and Supplemental Dataset 1). While the majority of these precursor peptides are virtually identical to thermocellin (CteA/Tte1186a), numerous smaller groupings display distinct core peptides (Figure S9). Several of these smaller groups display conserved Asp residues along with the putative Cys donors, indicating that the Asp residues may often serve as an acceptor for thioether formation.

To uncover additional phylogenetic trends, we inspected the taxonomic distribution of all rSAMs predicted form thioether-containing RiPPs. The vast majority (~95%) are encoded by Firmicutes, consistent with *Clostridium* sequencing bias as well as the fact that all previously isolated sactipeptides originate from *Bacillus* (Figure S10). Nonetheless, divergent examples occur in numerous other phyla, including Proteobacteria, Actinobacteria,

and Bacteroidetes, albeit with much lower frequency. We note that some of the genes encoding the rSAM enzyme from *Clostridium* showed significantly higher % GC content when compared to the entire genome, which may be indicative of horizontal gene transfer (Figure S11). Additionally, based on local genomic context (Table S1), many unusual gene cluster architectures were identified including cases where the RiPP may be further elaborated by locally encoded enzymes (Figure S12). Co-occurrence with genes related to tRNA metabolism were also highly represented in the genomic context of SCIFFs, which may hint towards a function for these RiPPs, although this will require additional investigation (Table S2 and Figure S12).

Not every rSAM enzyme in our dataset could be confidently assigned a substrate peptide. Therefore, the precursor peptides in such cases remain "unidentified" due to one of several possible reasons, such as *(i)* poor quality genome assembly, *(ii)* the precursor peptide is encoded elsewhere in the genome, *(iii)* the BGC has been deactivated (a relic) and there is no longer a coding sequence for the precursor peptide, or finally, (*iv*) the rSAM is not involved in RiPP biosynthesis. Indeed, some members within the unidentified group do not appear to be involved in RiPP biosynthesis; however, our investigation found that only a few examples could be confidently placed in "non-RiPP" category based on local genomic context. A final group of rSAM enzymes were identified as orthologs of QhpD. These proteins are not involved in RiPP biosynthesis but rather they are involved in the posttranslational modification of a large protein.[52–54] Nonetheless, this indicates that our targeted BLAST searching was sufficient in identifying many thioether-installing rSAM enzymes involved in RiPP biosynthesis.

Insights from relatedness to QhpD

Given our targeted BLAST searching, it was initially surprising to retrieve such a large number of enzymes annotated as QhpD (n = 674, Supplemental Dataset 2). QhpD is not involved in RiPP biosynthesis but rather the posttranslational maturation of quinohemoprotein amine dehydrogenase (QHNDH) which catalyzes the oxidative deamination of aliphatic and aromatic amines.[55–57] QHNDH is composed of three subunits and contains several unusual posttranslational modifications, including a cysteinyl tryptophylquinone cofactor as well as three catalytically essential thioether linkages. Two are between Cys-Asp (S-Cβ) while the third thioether linkage is between Cys and Glu (S-Cγ).[52–54] As illustrated on the SSN, QhpD and orthologous rSAM enzymes share considerable sequence similarity to rSAM enzymes involved in SCIFF biosynthesis (Figure 3). This finding is significant because sactipeptides are defined based on a thioether crosslink to the alpha carbon (S-Cα) of an acceptor residue. Therefore, in addition to modifying a protein instead of a RiPP precursor peptide, the QhpD group should be carefully distinguished from rSAM enzymes that form sactipeptides. However, considering the stringency used during our bioinformatics survey and the similarity score threshold in the SSN, we predicted that any rSAM enzyme connected to the QhpD (i.e. the SCIFF cluster) would be more likely to catalyze S-Cβ and/or S-Cγ thioether linkges rather than a sactionine linkage. While an SSN provides a useful and simplified illustration of sequence similarity, the relatedness of the separate clusters is not depicted. Thus, the same set of rSAM enzymes were analyzed after constructing a maximum likelihood tree (Figure S13).

Corroborating the previous analysis, rSAM enzymes implicated in SCIFF biosynthesis are more related to QhpD than those involved in sactionine formation (Table S7). A subset of SCIFFs with precursor peptides distinct from those previously reported (Tte1186a and CteA) were also identified that are more similar to NxxcB, a rSAM enzyme recently described to form a S-Cβ thioether linkage to an Asn acceptor in a RiPP that is neither a sactipeptide nor a SCIFF.[58] It should be noted that this target dataset was aggregated using BLAST queries of biosynthetic enzymes known to be involved in sactipeptide and SCIFF biosynthesis, rather than using an all-inclusive list of rSAM proteins. This resulted in a more manageable dataset size but one that also lacks many proteins that may or may not be involved in peptidic thioether formation. Thus, the relationship of the rSAM proteins involved in sactipeptide versus SCIFF biosynthesis appear here artificially more related to one another. To illustrate this, a BLAST query using AlbA (subtilosin, sacipeptide) will return nearly 40,000 protein sequences before retrieving the less-similar CteB (thermocellin, SCIFF).

Freyrasin is a S–Cβ crosslinked peptide: MS evidence

With bioinformatic evidence that rSAM proteins annotated as SCIFF maturases might not contain S-Cα thioether linkages, and that previous studies never confirmed the nature of the SCIFF crosslink in the SCIFFs,[24,25] we next sought to determine the connectivity of the SCIFF thioether crosslink. To obtain the milligram quantities of peptide that would be required for spectroscopic characterization, we first targeted the predicted SCIFF from *Paenibacillus polymyxa* ATCC 842 (Figure 4). This BGC was selected because it displayed a unique sequence variation of six CX_3D repeats which suggested that each of the six Cys donors would be linked to an Asp acceptor (similar to QhpD, Figure S9). After MS-based screening of various culturing conditions, we did not detect any ions consistent with the expected natural product. Therefore, we heterologously coexpressed the precursor peptide (PapA) and rSAM enzyme (PapB) in *Escherichia coli* following a previously published method.[44,59] After purification, the peptide product obtained was analyzed by MALDI-TOF-MS, which indicated that the precursor peptide had undergone processing with a mass loss of 12 Da (consistent with six thioether linkages, Figure S14). Controls where the precursor peptide was expressed in the absence of the rSAM enzyme resulted in a product with a mass loss of 6 Da from the starting material, consistent with the formation of three disulfide bonds). In contrast to the single crosslinks reported from in vitro studies on Tte1186a and CteA (thermocellin),[24,25] we observed a single product where all of the six Cys were modified. We have termed this group the freyrasins.

The structure of freyrasin obtained through heterologous expression was further examined by HR-MS/MS (Figure S15). The observed *m/z* value ($[M+4H]^{4+}$ = 1397.6113) agreed with the molecular formula for the predicted structure ($C_{233}H_{262}N_{66}O_{80}S_7^{4+}$; $[M+4H]^{4+}$ = 1397.6094; error = 1.4 ppm). The tandem MS data were also consistent with the freyrasin precursor peptide, although the fragmentation pattern observed showed a limited number of ions for a peptide of this size. Significant ions were only detected for sites that dissociated either directly before or after non-overlapping CX_3D motifs. Such fragmentation behavior was reminiscent of a lanthipeptide-like structure as opposed to a sactipeptide-like structure.[60–62] Indeed, prior work from our lab and others have shown that the sactionine linkage

itself (S-Cα thioether) readily dissociate under standard CID condition whereas S-Cβ thioether linkages do not, as is well-documented for lanthipeptides.[16,59,63] These observations are further consistent with the lack of inter-ring fragmentation data recently reported from in vitro work on SCIFF peptides that contain a single thioether.[24,25] By virtue of not undergoing this characteristic fragmentation, we speculated that freyrasin might not contain S-Cα linkages.

To better understand this difference in CID fragmentation behavior, we initiated quantum mechanical calculations of two model tripeptides (Gly-Cys-Gly and Gly-Asp-Gly) that were Cys-Asp thioether linked to the Cα or Cβ of Asp (Supplemental Methods). These calculations revealed that the difference in zero-point energy between the two model peptides was comparable to thermal energies (1 kcal/mol) and thus negligible. However, the electronic energy of the S-Cβ linked model peptide was found to be 12 kcal/mol more stable than the corresponding S-Cα linked isomer (Figure S16). We attribute the weaker bond energy of the S-Cα linkage to the greater stability of the alpha-centered radical that arises after homolytic bond dissociation. Taken with the bioinformatic predictions, these calculations provide an explanation for the differential CID behavior between the two linkage types and further suggests that SCIFF thioethers are not alpha linked.

Freyrasin is a S–Cβ crosslinked peptide: NMR evidence

To rigorously establish the atomic connectivity of the freyrasin thioether linkages, various multidimensional NMR experiments were performed. Proteolytic digestion yielded a peptide cleaved between Tyr/Gly that was employed for NMR analysis (Figure S17). Using the known precursor peptide sequence (Figure 4) and ^{1}H-^{1}H TOCSY spectrum, residues were assigned based on the cross-peak patterns observed in the amide region from 7.70–8.60 ppm (Figure S18). The ^{1}H-^{1}H NOESY spectrum was used to confirm the linear freyrasin sequence (Figure S19). All residues were confidently assigned with the exception of Cys17 (Table S5). Through-space correlations were visible for all residues with the exception of Asn5 and Asn16. This is likely due to a flexible structure, as demonstrated by a narrow amide chemical shift dispersion.[64] As revealed in the TOCSY spectrum, freyrasin displayed NH-CαH-CβH correlations for five of the six Asp residues believed to be thioether crosslinked, confirming a continuous spin system. This correlation would not be possible for a S-Cα thioether (i.e. sactionine). Additionally, each of the six Asp residues generated NH-CαH-CβH TOCSY correlations similar to that of Ser, with two deshielded resonances occurring in the CαH region (3.6–4.7 ppm), again supportive of a Cys-Asp (S-Cβ) linked thioether (Figure S20). To assign the ring topology of freyrasin, NOESY correlations between Cys-Asp were assigned based on the chemical shift values obtained by TOCSY (Figure 5). Two of the six thioether rings were assigned unambiguously. The redundancy of the freyrasin primary sequence unfortunately led to extensive chemical shift overlap compounded with line broadening, presumably arising from the flexibility of the peptide, so complete NOESY correlations were not evident for the remaining rings in freyrasin. Nonetheless, when coupled with the bioinformatics, MS data, and the definitive NMR assignments, freyrasin has been determined to contain exclusively Cys-Asp (S-Cβ) linkages. With the structure confirmed, we next tested for antibiotic activity of the heterologously obtained freyrasin, but it was not significantly active against any strain tested (MIC ≥32

μM). It is possible that additional modifications are present in the native product, but also possible are that freyrasin harbors very narrow-spectrum activity towards an untested strain or instead functions in a capacity unrelated to growth suppression (Table S4).

Thermocellin contains a Cys-Thr (S-Cγ) linked thioether

Given that freyrasin contains exclusively S-Cβ linkages in agreement with our bioinformatic predictions, we sought to determine if other "SCIFFs" also lacked sactionine linkages. The only two (partially) characterized SCIFFs thus far reported (i.e. Tte1186a and thermocellin) were previously investigated after in vitro enzymatic reconstitution, which resulted in the formation of only one Cys-Thr thioether.[24,25] Owing to the precursor containing six cysteines, the single thioether-containing species is likely a biosynthetic intermediate and not reflective of fully mature thermocellin. Despite these studies drawing on homology to sactionine-forming rSAM enzymes to predict linkage type, the presumed thioether installed by these enzymes did not undergo the gas-phase dissociation reaction that is characteristic of sactionine linkages. When further considering *(i)* our tandem MS data that show huazacin, a Cys-Thr (S-Cα)-containing sactipeptide, behaved as expected under CID conditions (Figure S3) and *(ii)* the thermocellin-forming rSAM enzyme, CteB, is more related to QhpD than any known sactionine-forming rSAM enzyme (Figure S13), we questioned whether thermocellin actually contains S-Cα linkages, as previously reported.

The possible thioethers for the Cys-Thr crosslink in thermocellin could theoretically include either the alpha, beta, or gamma carbon of Thr. We believed that in addition to an unlikely Cys-Thr (S-Cα) linkage for the above-mentioned reasons, a (S-Cβ) linkage would also be unlikely given the instability of resulting thiohemiacetal. To confirm or refute if thermocellin possessed a Cys-Thr (S-Cγ) crosslink, we prepared an isotopically labeled version of the thermocellin precursor peptide, CteA, that included ^{2}H at the α and β positions of Thr while retaining ^{1}H at the γ-methyl group. Isotopically labeled Thr was incorporated by overexpression of CteA in a Thr auxotroph (Δ*thrC*) of *E. coli* in defined minimal media where the only source of Thr was the supplied L-[4-^{13}C,2,3-$^{2}H_2$]-Thr. Initial protein preparations led to scrambling of the ^{13}C and ^{2}H labels into other amino acids, resulting in a complex series of masses (Figure S21). Isotopic scrambling was readily suppressed by the addition of natural abundance Ile and Gly to the expression media, resulting in production of fully labeled CteA with no detectable under- or over-labeling (Figure S21).[65] The location of the isotopic labels was verified to be confined to Thr residues by HR-MS/MS (Figure S22).

Having devised a means to probe the chemical nature of the thioether linkage in thermocellin, we next coexpressed CteA with CteB (the rSAM enzyme) using the same conditions that led to exclusive Thr isotope labeling. Consistent with a Cys-Thr (S-Cγ) linkage, a two Da mass loss was observed relative to the starting material. This result is consistent only with the abstraction of ^{1}H from Thr and the loss of the thiol hydrogen during thioether formation. Abstraction of ^{2}H from either the α or β carbons of Thr would have yielded a three Da mass loss (Figure S23). HR-MS/MS analysis of the product confirmed a Cys-Thr (S-Cγ) thioether species that did not undergo gas-phase dissociation associated with S-Cα thioethers. Unexpectedly, the tandem MS data unequivocally showed that the Cys-Thr linkage is to the Thr that resides two positions away to the donor Cys, which is in

contrast to previous reports that show localization to a Thr five residues away from the donor Cys (Figure 6). This localization was further corroborated by the observation of fragmentation within the new assignment of the macrocycle in unmodified peptide reacted with *N*-ethylmaleimide (Figure S24). To evaluate if the incorporation of stable isotopes altered the in vivo activity of CteB, we analyzed the coexpression product with natural abundance Thr and verified the same location for the thioether linkage (Figure S25). While it cannot be ruled out that the difference in localization is due to previous work being done in vitro with purified enzymes, our data conclusively show that ^{1}H was abstracted from the γ-methyl group of Thr during thioether formation which is consistent only with a (S-Cγ) linkage.

Proposal to alter nomenclature for thioether-containing natural products

Both freyrasin and thermocellin belong to the "SCIFF" family owing to both containing six Cys residues in a span of ~45 total precursor residues. As we have shown that freyrasin and thermocellin lack S-Cα thioethers but instead harbor S-Cβ and S-Cγ linkages, respectively. Therefore, we propose to rename this class the ranthipeptides (radical non-alpha thioether peptides) to distinguish them both structurally and biosynthetically from other thioether-containing RiPP classes, namely the sactipeptides and lanthipeptides.

Conclusion

In this work, we revealed that SCIFFs are not sactipeptides but instead feature rSAM-installed thioether crosslinks between Cys donor residues and the β- or γ-carbon of an acceptor residue. To better reflect their rSAM-dependent biosynthesis and to draw a correlation to lanthipeptides, which are S-Cβ linked, we propose the name ranthipeptides for this RiPP class. Our assignment of ranthipeptides as a RiPP class distinct from sactipeptides is supported by several independent lines of evidence including: (*i*) rSAM enzymes associated with ranthipeptide precursor peptides are considerably more sequence similar to QhpD, a known S-Cβ and S-Cγ thioether-forming enzyme, than any sactipeptide-associated rSAM enzyme, (*ii*) unlike sactionines, ranthipeptide thioethers do not fragment under mild CID conditions, an observation now supported by quantum mechanical bond dissociation energy calculations, (*iii*) multidimensional NMR spectroscopy, including ^{1}H-^{1}H TOCSY and ^{1}H-^{1}H NOESY that support a Cys-Asp S-Cβ linkage for freyrasin, (*iv*) direct tandem MS and NMR comparison of freyrasin to the sactipeptide huazacin, which contains a Cys-Asp S-Cα linkage, and (*v*) CteB (thermocellin-forming rSAM enzyme) abstracting hydrogen from the γ-carbon of the acceptor Thr during thioether formation. Collectively, these data provide conclusive evidence that both freyrasin and thermocellin are not alpha-linked. With previously reported beta-linked thioether-containing RiPPs being restricted to the lanthipeptides and the recently reported NxxcA, the ranthipeptide linkage chemistry is now considerably much richer as it is theoretically compatible with any non-Gly acceptor residue. This feature renders ranthipeptides an attractive target for engineering conformationally restrained peptides.

Over the past few years, several new peptide reaction chemistries catalyzed by RiPP rSAM enzymes[23] have been discovered and include the proteusins,[66] epipeptides,[67] streptide and

structurally similar doubly-crosslinked RiPPs,[68,69] mycofactocin,[70,71] and the alpha-keto beta-amino acid-containing peptides.[45] The discovery of S-Cβ and S-Cγ linked thioethers adds ranthipeptides to this growing list of unique chemistry found in RiPP biosynthetic pathways. We propose that a combination of sequence similarity and tandem MS analysis will provide valuable insight in discerning the various linkage chemistries of yet-uncharacterized ranthipeptide products.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENT

We are grateful to L. Zhu (Univ. of Illinois) for assistance with NMR data acquisition. We further thank R. Zallot and T. Stack (J. Gerlt lab, Univ. of Illinois) for assisting with anaerobic culturing. Lastly, we acknowledge R. Gennis (Univ. of Illinois) and Toshio Iwasaki (Nippon Medical School) for creating the threonine auxotroph strain, *E. coli* RF2.

Funding Sources

This work was supported in part by the Robert C. and Carolyn J. Springborn Endowment (to B.J.B), the National Science Foundation (DGE-1144245 to B.J.B.), the Chemistry-Biology Interface Research Training Program (2T32 GM070421 to B.J.B. and C.J.S.), and the National Institutes of Health (Grant GM123998 to D.A.M.).

References

(1). Ziemert N; Alanjary M; Weber T The evolution of genome mining in microbes - a review. Nat. Prod. Rep. 2016, 33 (8), 988. [PubMed: 27272205]

(2). Medema MH; Fischbach MA Computational approaches to natural product discovery. Nat. Chem. Biol. 2015, 11 (9), 639. [PubMed: 26284671]

(3). Tietz JI; Mitchell DA Using genomics for natural product structure elucidation. Curr. Top. Med. Chem. 2016, 16 (15), 1645. [PubMed: 26456468]

(4). Arnison PG; Bibb MJ; Bierbaum G; Bowers AA; Bugni TS; Bulaj G; Camarero JA; Campopiano DJ; Challis GL; Clardy J; Cotter PD; Craik DJ; Dawson M; Dittmann E; Donadio S; Dorrestein PC; Entian KD; Fischbach MA; Garavelli JS; Göransson U; Gruber CW; Haft DH; Hemscheidt TK; Hertweck C; Hill C; Horswill AR; Jaspars M; Kelly WL; Klinman JP; Kuipers OP; Link AJ; Liu W; Marahiel MA; Mitchell DA; Moll GN; Moore BS; Müller R; Nair SK; Nes IF; Norris GE; Olivera BM; Onaka H; Patchett ML; Piel J; Reaney MJ; Rebuffat S; Ross RP; Sahl HG; Schmidt EW; Selsted ME; Severinov K; Shen B; Sivonen K; Smith L; Stein T; Süssmuth RD; Tagg JR; Tang GL; Truman AW; Vederas JC; Walsh CT; Walton JD; Wenzel SC; Willey JM; van der Donk WA Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. 2013, 30 (1), 108. [PubMed: 23165928]

(5). Oman TJ; van der Donk WA Follow the leader: The use of leader peptides to guide natural product biosynthesis. Nat. Chem. Biol. 2010, 6 (1), 9. [PubMed: 20016494]

(6). Hudson GA; Mitchell DA RiPP antibiotics: biosynthesis and engineering potential. Curr. Opin. Microbiol. 2018, 45, 61. [PubMed: 29533845]

(7). Johnson M; Zaretskaya I; Raytselis Y; Merezhuk Y; McGinnis S; Madden TL NCBI BLAST: a better web interface. Nucleic Acids Res. 2008, 36, W5. [PubMed: 18440982]

(8). Tietz JI; Schwalen CJ; Patel PS; Maxson T; Blair PM; Tai H-C; Zakai UI; Mitchell DA A new genome-mining tool redefines the lasso peptide biosynthetic landscape. Nat. Chem. Biol. 2017, 13 (5), 470. [PubMed: 28244986]

(9). Finn RD; Coggill P; Eberhardt RY; Eddy SR; Mistry J; Mitchell AL; Potter SC; Punta M; Qureshi M; Sangrador-Vegas A; Salazar GA; Tate J; Bateman A The Pfam protein families database:

towards a more sustainable future. Nucleic Acids Res. 2016, 44 (D1), D279. [PubMed: 26673716]

(10). Haft DH; Selengut JD; Richter RA; Harkins D; Basu MK; Beck E TIGRFAMs and Genome Properties in 2013. Nucleic Acids Res. 2013, 41 (D1), D387. [PubMed: 23197656]

(11). Schwalen CJ; Hudson GA; Kille B; Mitchell DA Bioinformatic expansion and discovery of thiopeptide antibiotics. J. Am. Chem. Soc. 2018, 140 (30), 9494. [PubMed: 29983054]

(12). Blin K; Wolf T; Chevrette MG; Lu X; Schwalen CJ; Kautsar SA; Suarez Duran H. G.; de Los Santos E. L. C.; Kim HU; Nave M; Dickschat JS; Mitchell DA; Shelest E; Breitling R; Takano E; Lee SY; Weber T; Medema MH antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. Nucleic Acids Res. 2017, 45 (W1), W36. [PubMed: 28460038]

(13). DiCaprio AJ; Firouzbakht A; Hudson GA; Mitchell DA Enzymatic Reconstitution and Biosynthetic Investigation of the Lasso Peptide Fusilassin. J. Am. Chem. Soc. 2019, 141 (1), 290. [PubMed: 30589265]

(14). Mathur H; Fallico V; O'Connor PM; Rea MC; Cotter PD; Hill C; Ross RP Insights into the Mode of Action of the Sactibiotic Thuricin CD. Front. Microbiol. 2017, 8, 696. [PubMed: 28473822]

(15). Flühe L; Marahiel MA Radical S-adenosylmethionine enzyme catalyzed thioether bond formation in sactipeptide biosynthesis. Curr. Opin. Chem. Biol. 2013, 17 (4), 605. [PubMed: 23891473]

(16). Rea MC; Sit CS; Clayton E; O'Connor PM; Whittal RM; Zheng J; Vederas JC; Ross RP; Hill C Thuricin CD, a posttranslationally modified bacteriocin with a narrow spectrum of activity against Clostridium difficile. Proc. Natl. Acad. Sci. USA 2010, 107 (20), 9352. [PubMed: 20435915]

(17). Babasaki K; Takao T; Shimonishi Y; Kurahashi K Subtilosin A, a new antibiotic peptide produced by Bacillus subtilis 168: isolation, structural analysis, and biogenesis. J. Biochem. 1985, 98 (3), 585. [PubMed: 3936839]

(18). Flühe L; Knappe TA; Gattner MJ; Schafer A; Burghaus O; Linne U; Marahiel MA The radical SAM enzyme AlbA catalyzes thioether bond formation in subtilosin A. Nat. Chem. Biol. 2012, 8 (4), 350. [PubMed: 22366720]

(19). Broderick JB; Duffus BR; Duschene KS; Shepard EM Radical S-adenosylmethionine enzymes. Chem. Rev. 2014, 114 (8), 4229. [PubMed: 24476342]

(20). Bauerle MR; Schwalm EL; Booker SJ Mechanistic diversity of radical S-adenosylmethionine (SAM)-dependent methylation. J. Biol. Chem. 2015, 290 (7), 3995. [PubMed: 25477520]

(21). Mahanta N; Hudson GA; Mitchell DA Radical S-adenosylmethionine enzymes involved in RiPP biosynthesis. Biochemistry 2017, 56 (40), 5229. [PubMed: 28895719]

(22). Mahanta N; Zhang Z; Hudson GA; van der Donk WA; Mitchell DA Reconstitution and Substrate Specificity of the Radical S-Adenosyl-methionine Thiazole C-Methyltransferase in Thiomuracin Biosynthesis. J. Am. Chem. Soc. 2017, 139 (12), 4310. [PubMed: 28301141]

(23). Benjdia A; Balty C; Berteau O Radical SAM enzymes in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs). Front. Chem. 2017, 5 (87), 87. [PubMed: 29167789]

(24). Bruender NA; Wilcoxen J; Britt RD; Bandarian V Biochemical and spectroscopic characterization of a radical S-adenosyl-l-methionine enzyme involved in the formation of a peptide thioether cross-link. Biochemistry 2016, 55 (14), 2122. [PubMed: 27007615]

(25). Grove TL; Himes PM; Hwang S; Yumerefendi H; Bonanno JB; Kuhlman B; Almo SC; Bowers AA Structural insights into thioether bond formation in the biosynthesis of sactipeptides. J. Am. Chem. Soc. 2017, 139 (34), 11734. [PubMed: 28704043]

(26). Benjdia A; Guillot A; Lefranc B; Vaudry H; Leprince J; Berteau O Thioether bond formation by SPASM domain radical SAM enzymes: Cα H-atom abstraction in subtilosin A biosynthesis. Chem. Commun. 2016, 52 (37), 6249.

(27). Grell TAJ; Goldman PJ; Drennan CL SPASM and Twitch Domains in S-Adenosylmethionine (SAM) Radical Enzymes. J. Biol. Chem. 2015, 290 (7), 3964. [PubMed: 25477505]

(28). Barr I; Stich TA; Gizzi AS; Grove TL; Bonanno JB; Latham JA; Chung T; Wilmot CM; Britt RD; Almo SC; Klinman JP X-ray and EPR characterization of the auxiliary Fe-S clusters in the radical SAM enzyme PqqE. Biochemistry 2018, 57 (8), 1306. [PubMed: 29405700]

(29). Grell TAJ; Kincannon WM; Bruender NA; Blaesi EJ; Krebs C; Bandarian V; Drennan CL Structural and spectroscopic analyses of the sporulation killing factor biosynthetic enzyme SkfB, a bacterial AdoMet radical sactisynthase. J. Biol. Chem. 2018, ASAP.

(30). Sit CS; van Belkum MJ; McKay RT; Worobo RW; Vederas JC The 3D solution structure of thurincin H, a bacteriocin with four sulfur to α-carbon crosslinks. Angew. Chem., Int. Ed. 2011, 50 (37), 8718.

(31). Liu W-T; Yang Y-L; Xu Y; Lamsa A; Haste NM; Yang JY; Ng J; Gonzalez D; Ellermeier CD; Straight PD; Pevzner PA; Pogliano J; Nizet V; Pogliano K; Dorrestein PC Imaging mass spectrometry of intraspecies metabolic exchange revealed the cannibalistic factors of Bacillus subtilis. Proc. Natl. Acad. Sci. USA 2010, 107 (37), 16286. [PubMed: 20805502]

(32). Flühe L; Burghaus O; Wieckowski BM; Giessen TW; Linne U; Marahiel MA Two [4Fe-4S] clusters containing radical SAM enzyme SkfB catalyze thioether bond formation during the maturation of the sporulation killing factor. J. Am. Chem. Soc. 2013, 135 (3), 959. [PubMed: 23282011]

(33). Duarte AFS; Ceotto-Vigoder H; Barrias ES; Souto-Padrón T; Nes IF; Bastos M Hyicin 4244, the first sactibiotic described in staphylococci, exhibits an anti-staphylococcal biofilm activity. Int J Antimicrob. Agents 2018, 51 (3), 349. [PubMed: 28705677]

(34). Haft DH; Basu MK Biological systems discovery in silico: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. J. Bacteriol. 2011, 193 (11), 2745. [PubMed: 21478363]

(35). Azevedo AC; Bento CBP; Ruiz JC; Queiroz MV; Mantovani HC Distribution and genetic diversity of bacteriocin gene clusters in rumen microbial genomes. Appl. Environ. Microbiol. 2015, 81 (20), 7290. [PubMed: 26253660]

(36). Letzel A-C; Pidot SJ; Hertweck C Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. BMC Genomics 2014, 15, 983. [PubMed: 25407095]

(37). Walsh CJ; Guinane CM; Hill C; Ross RP; O'Toole PW; Cotter PD In silico identification of bacteriocin gene clusters in the gastrointestinal tract, based on the Human Microbiome Project's reference genome database. BMC Microbiol. 2015, 15, 183. [PubMed: 26377179]

(38). Skinnider MA; Johnston CW; Edgar RE; Dejong CA; Merwin NJ; Rees PN; Magarvey NA Genomic charting of ribosomally synthesized natural product chemical space facilitates targeted mining. Proc. Natl. Acad. Sci. USA 2016, 113 (42), E6343.

(39). Murphy K; O'Sullivan O; Rea MC; Cotter PD; Ross RP; Hill C Genome mining for radical SAM protein determinants reveals multiple sactibiotic-like gene clusters. PLoS One 2011, 6 (7), e20852.

(40). Nakai T; Ito H; Kobayashi K; Takahashi Y; Hori H; Tsubaki M; Tanizawa K; Okajima T The radical S-adenosyl-L-methionine enzyme QhpD catalyzes sequential formation of intra-protein sulfur-to-methylene carbon thioether bonds. J. Biol. Chem. 2015, 290 (17), 11144. [PubMed: 25778402]

(41). Finn RD; Attwood TK; Babbitt PC; Bateman A; Bork P; Bridge AJ; Chang H-Y; Dosztányi Z; El-Gebali S; Fraser M; Gough J; Haft D; Holliday GL; Huang H; Huang X; Letunic I; Lopez R; Lu S; Marchler-Bauer A; Mi H; Mistry J; Natale DA; Necci M; Nuka G; Orengo CA; Park Y; Pesseat S; Piovesan D; Potter SC; Rawlings ND; Redaschi N; Richardson L; Rivoire C; Sangrador-Vegas A; Sigrist C; Sillitoe I; Smithers B; Squizzato S; Sutton G; Thanki N; Thomas PD; Tosatto Silvio C E; Wu CH.; Xenarios I; Yeh L-S.; Young S-Y.; Mitchell AL. InterPro in 2017—beyond protein family and domain annotations. Nucleic Acids Res. 2017, 45 (D1), D190. [PubMed: 27899635]

(42). Altschul SF; Madden TL; Schaffer AA; Zhang J; Zhang Z; Miller W; Lipman DJ Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997, 25 (17), 3389. [PubMed: 9254694]

(43). Gerlt JA; Bouvier JT; Davidson DB; Imker HJ; Sadkhin B; Slater DR; Whalen KL Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. Biochim. Biophys. Acta 2015, 1854 (8), 1019. [PubMed: 25900361]

(44). Himes PM; Allen SE; Hwang S; Bowers AA Production of sactipeptides in Escherichia coli: Probing the substrate promiscuity of subtilosin A biosynthesis. ACS Chem. Biol. 2016, 11 (6), 1737. [PubMed: 27019323]

(45). Burkhart BJ; Hudson GA; Dunbar KL; Mitchell DA A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. Nat. Chem. Biol. 2015, 11 (8), 564. [PubMed: 26167873]

(46). Mathur H; Rea MC; Cotter PD; Hill C; Ross RP The sactibiotic subclass of bacteriocins: an update. Curr. Protein Pept. Sci. 2015, 16 (6), 549. [PubMed: 26031307]

(47). Shelburne CE; An FY; Dholpe V; Ramamoorthy A; Lopatin DE; Lantz MS The spectrum of antimicrobial activity of the bacteriocin subtilosin A. J. Antimicrob. Chemother. 2007, 59 (2), 297. [PubMed: 17213266]

(48). Wang G; Manns DC; Guron GK; Churey JJ; Worobo RW Large-Scale Purification, Characterization, and Spore Outgrowth Inhibitory Effect of Thurincin H, a Bacteriocin Produced by Bacillus thuringiensis SF361. Probiotics Antimicrob. Proteins 2014, 6 (2), 105. [PubMed: 24798125]

(49). Lee H; Churey JJ; Worobo RW Biosynthesis and transcriptional analysis of thurincin H, a tandem repeated bacteriocin genetic locus, produced by Bacillus thuringiensis SF361. FEMS Microbiol. Lett. 2009, 299 (2), 205. [PubMed: 19732155]

(50). Kohl M; Wiese S; Warscheid B Cytoscape: software for visualization and analysis of biological networks. Methods Mol. Biol. 2011, 696, 291. [PubMed: 21063955]

(51). Su G; Morris JH; Demchak B; Bader GD Biological network exploration with Cytoscape 3. Curr. Protoc. Bioinformatics 2014, 47, 8 13 1.

(52). Satoh A; Kim JK; Miyahara I; Devreese B; Vandenberghe I; Hacisalihoglu A; Okajima T; Kuroda S; Adachi O; Duine JA; Van Beeumen J; Tanizawa K; Hirotsu K Crystal structure of quinohemoprotein amine dehydrogenase from Pseudomonas putida. Identification of a novel quinone cofactor encaged by multiple thioether cross-bridges. J. Biol. Chem. 2002, 277 (4), 2830. [PubMed: 11704672]

(53). Datta S; Mori Y; Takagi K; Kawaguchi K; Chen Z-W; Okajima T; Kuroda S.i.; Ikeda T; Kano K; Tanizawa K; Mathews FS Structure of a quinohemoprotein amine dehydrogenase with an uncommon redox cofactor and highly unusual crosslinking. Proc. Natl. Acad. Sci. USA 2001, 98 (25), 14268. [PubMed: 11717396]

(54). Vandenberghe I; Kim JK; Devreese B; Hacisalihoglu A; Iwabuki H; Okajima T; Kuroda S; Adachi O; Jongejan JA; Duine JA; Tanizawa K; Van Beeumen J The covalent structure of the small subunit from Pseudomonas putida amine dehydrogenase reveals the presence of three novel types of internal cross-linkages, all involving cysteine in a thioether bond. J. Biol. Chem. 2001, 276 (46), 42923. [PubMed: 11555656]

(55). Takagi K; Yamamoto K; Kano K; Ikeda T New pathway of amine oxidation respiratory chain of Paracoccus denitrificans IFO 12442. Eur. J. Biochem. 2001, 268 (2), 470. [PubMed: 11168384]

(56). Takagi K; Torimura M; Kawaguchi K; Kano K; Ikeda T Biochemical and Electrochemical Characterization of Quinohemoprotein Amine Dehydrogenase from Paracoccus denitrificans. Biochemistry 1999, 38 (21), 6935. [PubMed: 10346915]

(57). Adachi O; Kubota T; Hacisalihoglu A; Toyama H; Shinagawa E; Duine JA; Matsushita K Characterization of Quinohemoprotein Amine Dehydrogenase from Pseudomonas putida. Biosci. Biotechnol. Biochem. 1998, 62 (3), 469. [PubMed: 27315927]

(58). Caruso A; Bushin LB; Clark KA; Martinie RJ; Seyedsayamdost MR Radical Approach to Enzymatic β-Thioether Bond Formation. J. Am. Chem. Soc. 2019, 141 (2), 990. [PubMed: 30521328]

(59). Burkhart BJ; Kakkar N; Hudson GA; van der Donk WA; Mitchell DA Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products. ACS Cent. Sci. 2017, 3 (6), 629. [PubMed: 28691075]

(60). Zhang Q; Ortega M; Shi Y; Wang H; Melby JO; Tang W; Mitchell DA; van der Donk WA Structural investigation of ribosomally synthesized natural products by hypothetical structure enumeration and evaluation using tandem MS. Proc. Natl. Acad. Sci. USA 2014, 111 (33), 12031. [PubMed: 25092299]

(61). Wang J; Zhang L; Teng K; Sun S; Sun Z; Zhong J Cerecidins, novel lantibiotics from Bacillus cereus with potent antimicrobial activity. Appl. Environ. Microbiol. 2014, 80 (8), 2633. [PubMed: 24532070]

(62). Voller GH; Krawczyk JM; Pesic A; Krawczyk B; Nachtigall J; Süssmuth RD Characterization of new class III lantibiotics--erythreapeptin, avermipeptin and griseopeptin from Saccharopolyspora erythraea, Streptomyces avermitilis and Streptomyces griseus demonstrates stepwise N-terminal leader processing. ChemBioChem 2012, 13 (8), 1174. [PubMed: 22556031]

(63). Lohans CT; Vederas JC Structural characterization of thioether-bridged bacteriocins. J. Antibiot. 2014, 67 (1), 23. [PubMed: 24022605]

(64). Shenkarev ZO; Finkina EI; Nurmukhamedova EK; Balandin SV; Mineev KS; Nadezhdin KD; Yakimenko ZA; Tagaev AA; Temirov YV; Arseniev AS; Ovchinnikova TV Isolation, structure elucidation, and synergistic antibacterial activity of a novel two-component lantibiotic lichenicidin from Bacillus licheniformis VK21. Biochemistry 2010, 49 (30), 6462. [PubMed: 20578714]

(65). Monneau YR; Ishida Y; Rossi P; Saio T; Tzeng S-R; Inouye M; Kalodimos CG Exploiting E. coli auxotrophs for leucine, valine, and threonine specific methyl labeling of large proteins for NMR applications. J. Biomol. NMR 2016, 65 (2), 99. [PubMed: 27255761]

(66). Freeman MF; Gurgui C; Helf MJ; Morinaka BI; Uria AR; Oldham NJ; Sahl HG; Matsunaga S; Piel J Metagenome mining reveals polytheonamides as posttranslationally modified ribosomal peptides. Science 2012, 338 (6105), 387. [PubMed: 22983711]

(67). Benjdia A; Guillot A; Ruffié P; Leprince J; Berteau O Post-translational modification of ribosomally synthesized peptides by a radical SAM epimerase in Bacillus subtilis. Nat. Chem. Biol. 2017, 9 (7), 698.

(68). Schramma KR; Bushin LB; Seyedsayamdost MR Structure and biosynthesis of a macrocyclic peptide containing an unprecedented lysine-to-tryptophan crosslink. Nat. Chem. 2015, 7 (5), 431. [PubMed: 25901822]

(69). Bushin LB; Clark KA; Pelczer I; Seyedsayamdost MR Charting an unexplored streptococcal biosynthetic landscape reveals a unique peptide cyclization motif. J. Am. Chem. Soc. 2018, 140 (50), 17674. [PubMed: 30398325]

(70). Khaliullin B; Ayikpoe R; Tuttle M; Latham JA Mechanistic elucidation of the mycofactocin-biosynthetic radical S-adenosylmethionine protein, MftC. J. Biol. Chem. 2017, 292 (31), 13022. [PubMed: 28634235]

(71). Khaliullin B; Aggarwal P; Bubas M; Eaton GR; Eaton SS; Latham JA Mycofactocin biosynthesis: modification of the peptide MftA by the radical S-adenosylmethionine protein MftC. FEBS Lett. 2016, 590 (16), 2538. [PubMed: 27312813]

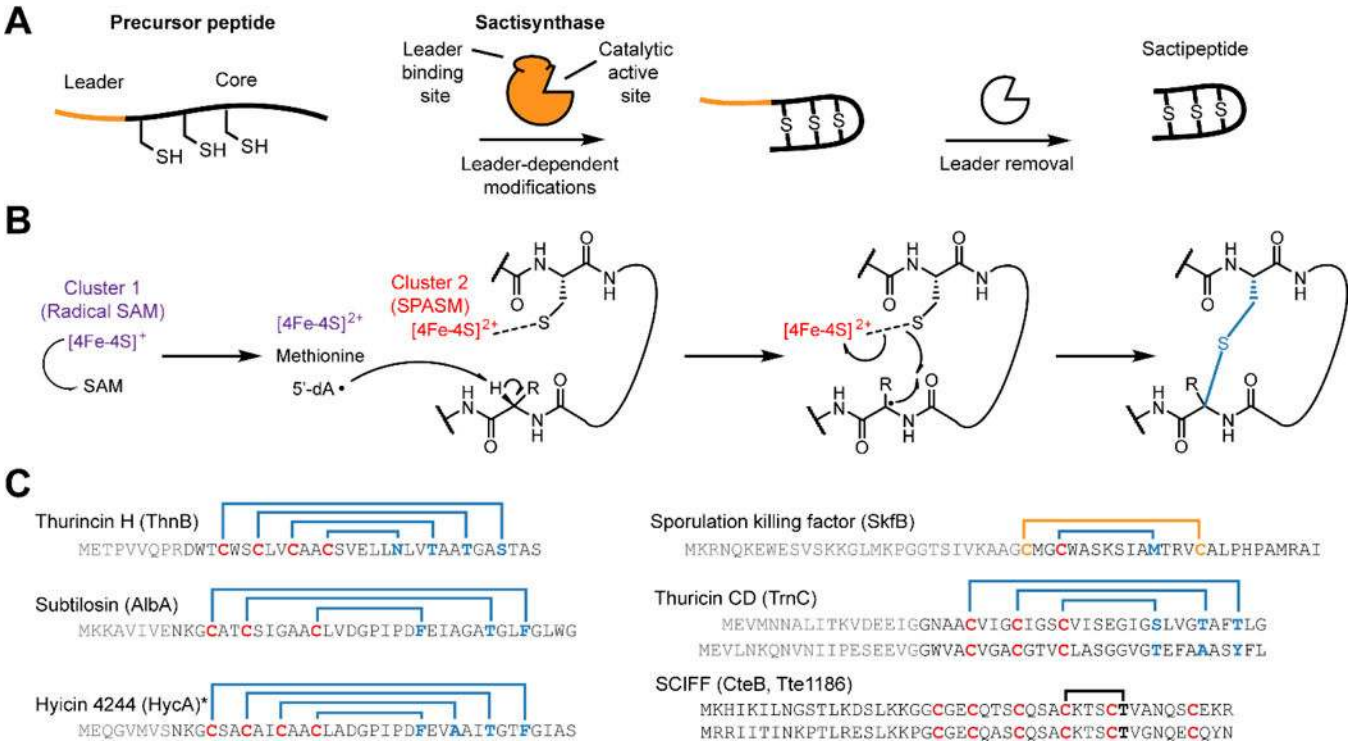

Figure 1. Sactipeptide biosynthesis and ring topologies.

(**A**) Schematic overview of sactipeptide biosynthesis. (**B**) Proposed mechanism for the rSAM-dependent formation of the sactipeptide S–Cα thioether crosslink. (**C**) Precursor peptide sequence and ring topologies of characterized sactipeptides (blue, S–Cα linkage; yellow, S-S linkage; black, uncharacterized). The requisite rSAM enzyme name is given in parentheses. Characterization of the six Cys in forty-five residues (SCIFFs) is limited to in vitro enzymatic assay for which only one thioether linkage was observed. Data shown later in this manuscript support a Cys-Thr linkage different to what has been previously reported and further that SCIFFs are not sactipeptides. * Hyicin 4244 has not been structurally characterized, thus the linkages indicated are speculative.

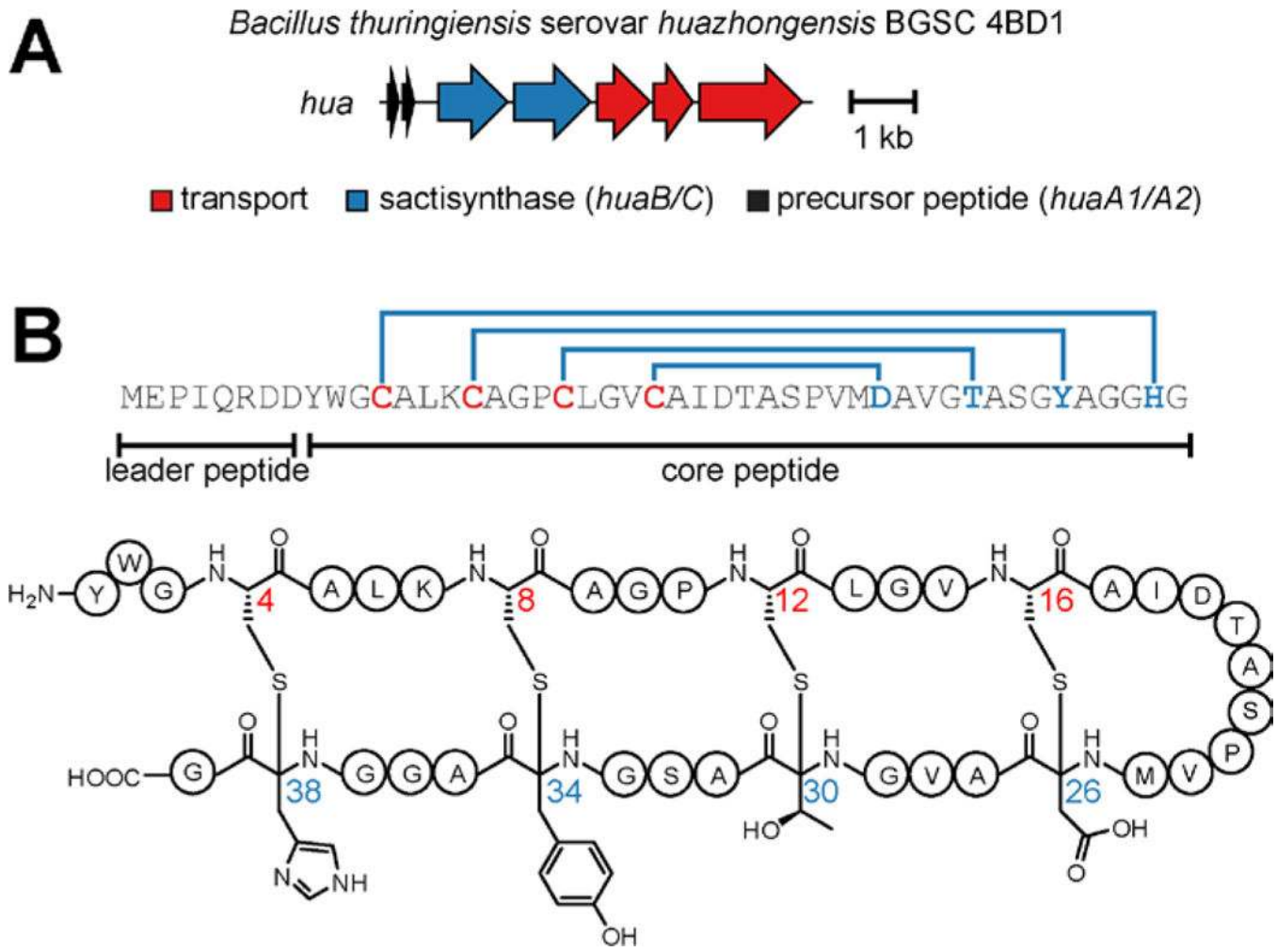

Figure 2. Huazacin from *Bacillus thuringiensis* serovar *huazhongensis*.

(**A**) Biosynthetic gene cluster. The sactipeptide synthases HuaB and HuaC (NCBI accession identifiers: EEM79976.1 and EEM79977.1, respectively) feature an N-terminal RRE domain and C-terminal SPASM domain similar to other sactisynthases.[27,45] (**B**) Precursor peptide sequence and chemical structure of huazacin. The *hua* gene cluster contains two identical copies of the precursor peptide shown.

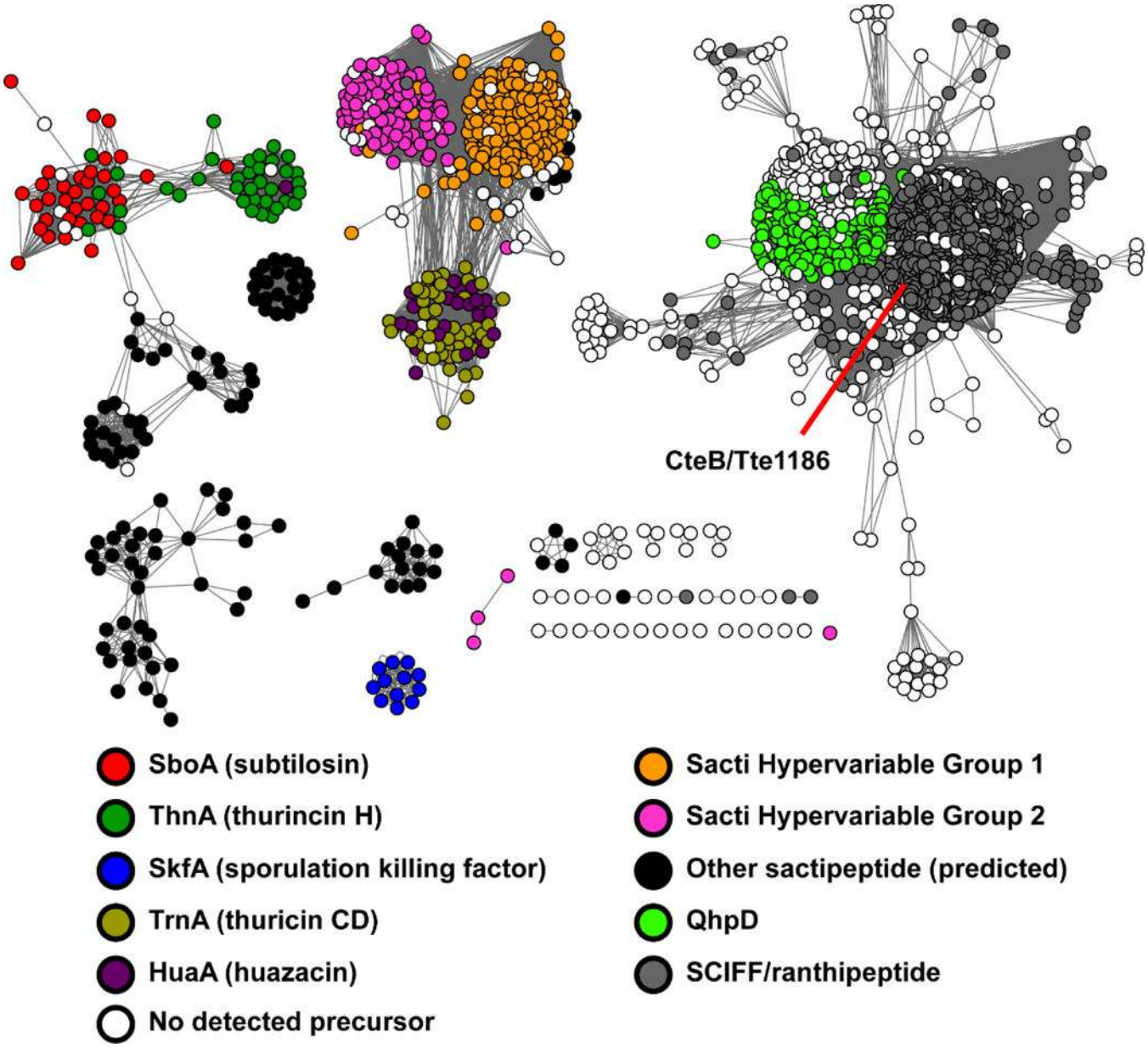

Figure 3. Sequence similarity network of rSAM enzymes.

rSAM enzymes involved in sactipeptide and SCIFF biosynthesis were used as BLAST inputs to gather additional rSAM enzymes predicted to form RiPP thioether linkages (n = 4,693). Sequences were analyzed by EFI-EST[43] and visualized using Cytoscape.[50] Nodes are colored based on the category of its associated precursor peptide (see legend). Sequences sharing >70% identity are conflated as a single node, and connected nodes indicate a similarity score of < 1e-50.[43,50,51] Sequence logos for core peptides associated with these rSAMs are shown in Figures S1 and S9. QhpD, non-RiPP rSAM involved in modifying quinohemoprotein amine dehydrogenase.[40]

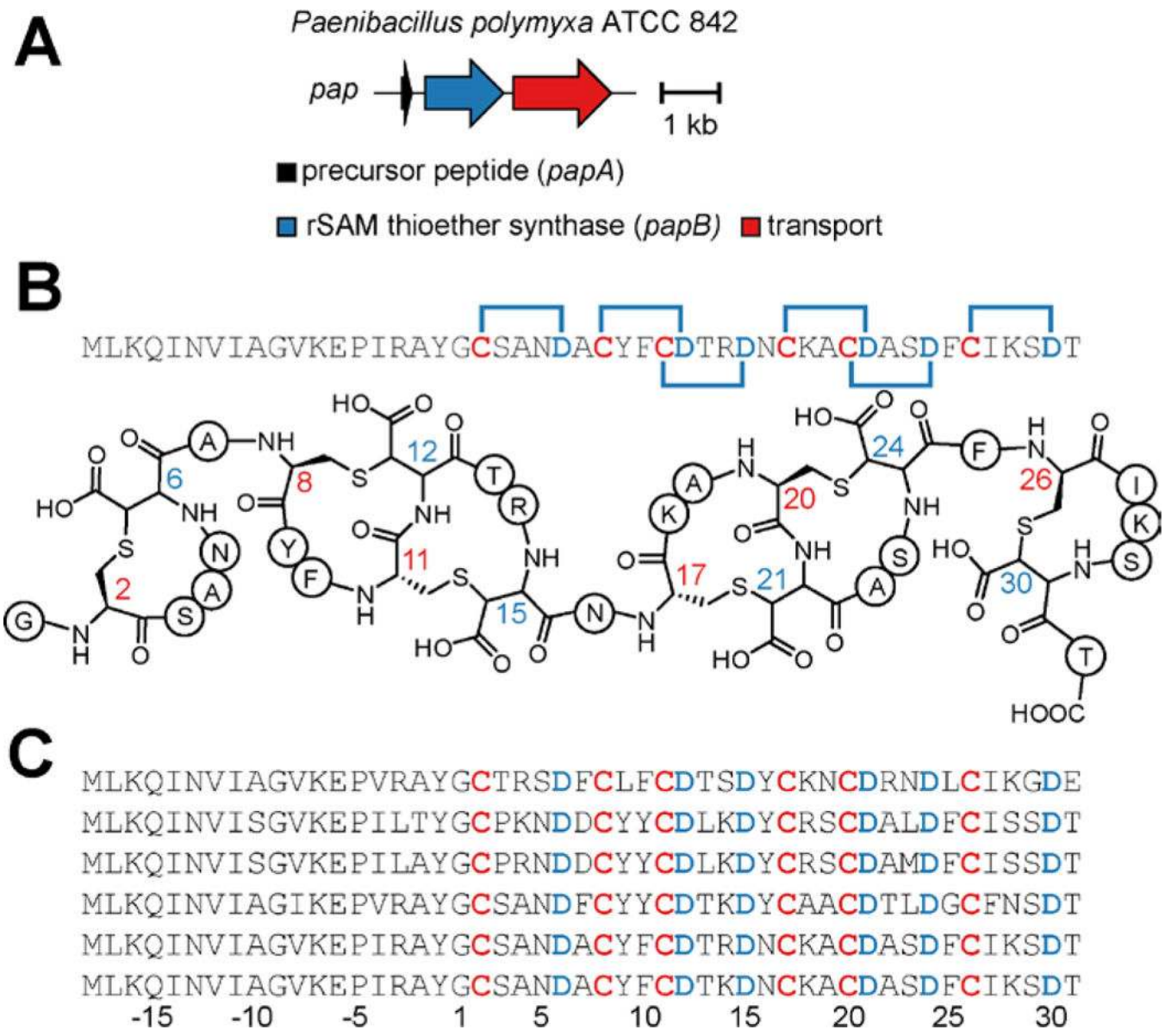

Figure 4. Freyrasin biosynthetic gene cluster and related peptides.

(**A**) Biosynthetic gene cluster diagram. (**B**) Freyrasin precursor peptide sequence with six CX_3D motifs and thioether topologies highlighted. (**C**) Alignment of additional freyrasin members. Leader peptide residues are negatively numbered relative to the core peptide, which we predict begins with Gly.

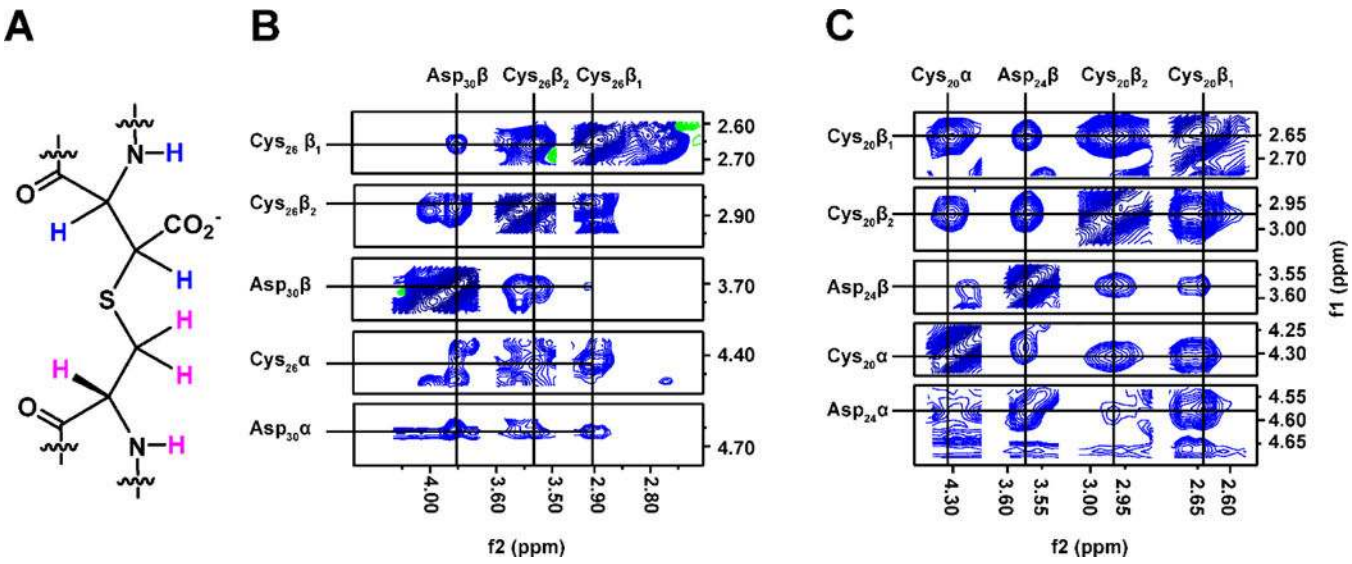

Figure 5. Freyrasin contains beta-linked thioethers.

(**A**) Cys-Asp (S–Cβ) connectivity with hydrogens colored pink (Cys) or blue (Asp). (**B**) ^{1}H-^{1}H NOESY correlations confirming a S–Cβ thioether linkage between Cys26-Asp30. (**C**) Same as panel B but for Cys20-Asp24.

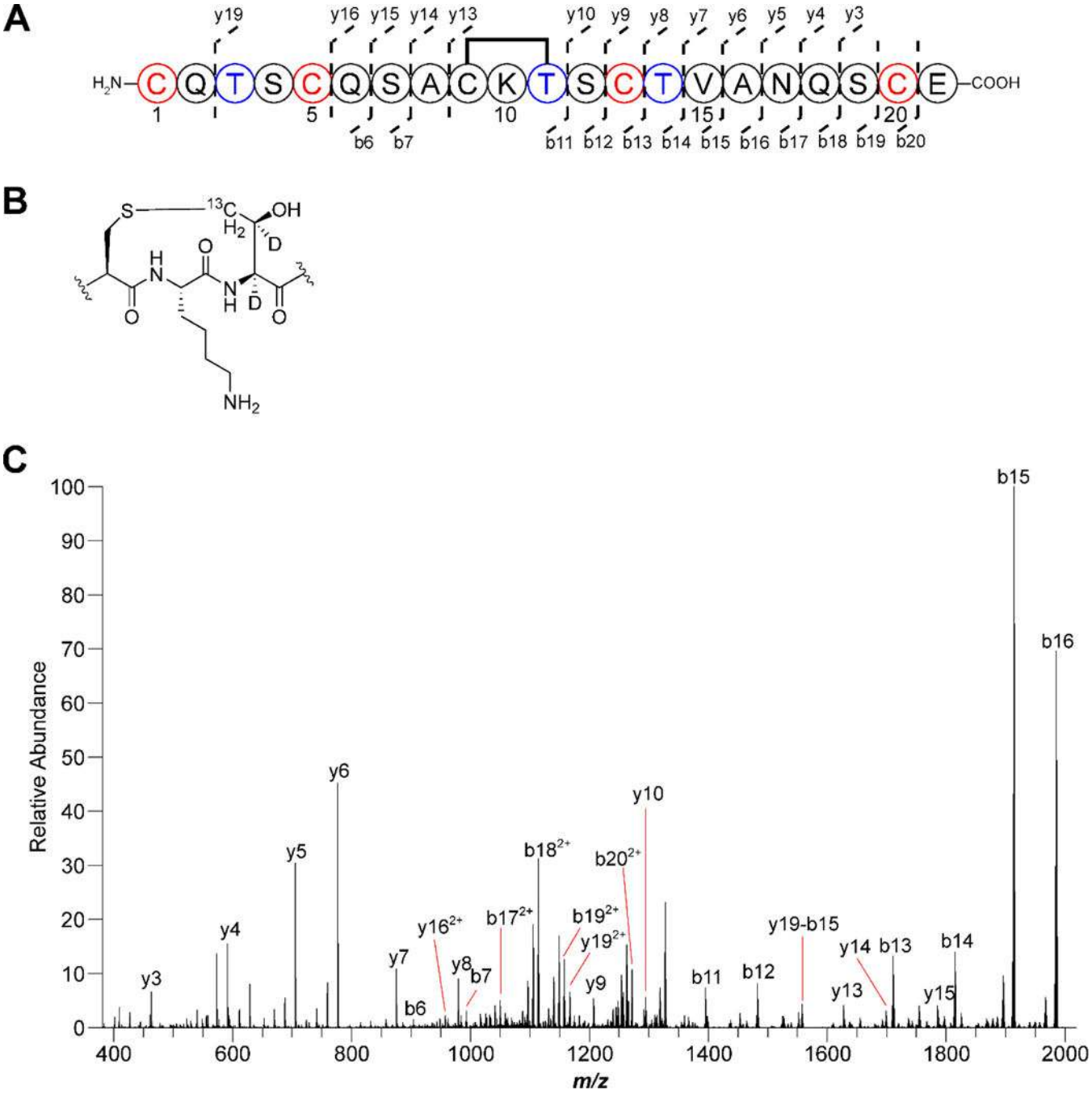

Figure 6. Thermocellin contains a Cys-Thr (S-Cγ) thioether.

(**A**) Coexpression of the thermocellin precursor containing labeled Thr (blue) features a −2 Da shift relative to unmodified peptide, demonstrating a S-Cγ linkage was formed. The modified peptide was treated with endoproteinase GluC and unmodified Cys were alkylated with *N*-ethylmaleimide (red) Residues are numbered based on the sequence after GluC treatment. (**B**) Structure of the isotopically labeled thioether linkage. (**C**) MS/MS analysis

yielded fragmentation consistent with a macrocycle formed between Cys9 and Thr11. Error tables for assigned ions are available in Table S6.