ACS BIO & MED CHEM Au
AN OPEN ACCESS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

pubs.acs.org/biomedchemau

Article

Catalytic Site Proximity Profiling for Functional Unification of Sequence-Diverse Radical *S*-Adenosylmethionine Enzymes

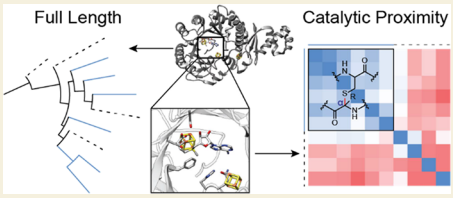
Timothy W. Precord, Sangeetha Ramesh, Shravan R. Dommaraju, Lonnie A. Harris, Bryce L. Kille, and Douglas A. Mitchell*

Cite This: *ACS Bio Med Chem Au* 2023, 3, 240−251

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The radical *S*-adenosylmethionine (rSAM) superfamily has become a wellspring for discovering new enzyme chemistry, especially regarding ribosomally synthesized and post-translationally modified peptides (RiPPs). Here, we report a compendium of nearly 15,000 rSAM proteins with high-confidence involvement in RiPP biosynthesis. While recent bioinformatics advances have unveiled the broad sequence space covered by rSAM proteins, the significant challenge of functional annotation remains unsolved. Through a combination of sequence analysis and protein structural predictions, we identified a set of catalytic site proximity residues with functional predictive power, especially among the diverse rSAM proteins that form sulfur-to-α carbon thioether (sactionine) linkages. As a case study, we report that an rSAM protein from *Streptomyces sparsogenes* (StsB) shares higher full-length similarity with MftC (mycofactocin biosynthesis) than any other characterized enzyme. However, a comparative analysis of StsB to known rSAM proteins using "catalytic site proximity" predicted that StsB would be distinct from MftC and instead form sactionine bonds. The prediction was confirmed by mass spectrometry, targeted mutagenesis, and chemical degradation. We further used "catalytic site proximity" analysis to identify six new sactipeptide groups undetectable by traditional genome-mining strategies. Additional catalytic site proximity profiling of cyclophane-forming rSAM proteins suggests that this approach will be more broadly applicable and enhance, if not outright correct, protein functional predictions based on traditional genomic enzymology principles.



KEYWORDS: *bioinformatics, radical SAM enzyme, sactipeptide, thioether, RiPP, enzyme function, genomic enzymology*

INTRODUCTION

Radical *S*-adenosylmethionine (rSAM) proteins comprise nearly 750,000 entries in UniProt and make up one of the largest and most functionally diverse enzyme superfamilies.[1] rSAM proteins are unified by the common usage of a [4Fe−4S] center to reductively cleave SAM to form a 5′-deoxyadenosine (5′-dA) radical.[2,3] Generally, the 5′-dA radical initiates substrate modification by hydrogen atom abstraction. The products of rSAM-catalyzed reactions are diverse, including methylations, epimerizations, carbon−carbon and carbon−heteroatom cross-linking, and various complex rearrangements.[4,5] Given the size of the rSAM superfamily and the breadth of reactions catalyzed, functional annotation has been particularly challenging. Exhaustive literature mining has identified that only a few hundred distinct rSAM proteins have been characterized. As a result of this knowledge gap, it is virtually impossible to predict the reaction chemistry for any uncharacterized rSAM protein of interest unless it shares an exceptionally high level of sequence identity with a known example. To assist with the bioinformatic analysis of this extensive superfamily, RadicalSAM.org was recently developed to modernize and expand the capabilities of the rSAM portion of the Structure−Function Linkage Database. The Web resource provides a comprehensive catalog of all rSAM proteins in UniProt, along with precomputed sequence similarity networks (SSNs), genome neighborhood diagrams, taxonomic data, useful statistical analyses, and external database link-outs to contextualize the superfamily.[6]

While useful inferences can often be gleaned from full-length sequence comparison and genomic context analysis, they leave several unsolved difficulties with functional prediction in uncharacterized proteins. One such difficulty pertains to assigning isofunctional groups within an SSN or clades within a phylogenetic tree. Generally speaking, SSNs are simplified versions of trees where protein sequences are represented by nodes. An edge drawn between two nodes indicates that the two proteins share a predefined similarity threshold based on the alignment score (e.g., an AS value of 50 is approximately equal to a BLAST expectation value of 1×10^{-50}). Even with experimentally characterized members, confident functional prediction for other proteins within the clade/group of interest can be difficult (Figure S1). Evolutionary theory posits that functional divergence will arise in all protein families and that

Received: December 24, 2022
Revised: February 8, 2023
Accepted: February 10, 2023
Published: March 1, 2023

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251

distinct lineages will exhibit variable rates of sequence divergence. For example, an SSN of the glycyl radical enzyme superfamily (InterPro family IPR004184) viewed at AS = 320 creates several trustworthy isofunctional groups but over-fractionates many other groups with identical functions (homologous proteins with the same function appear in separate groups).[7] Conversely, the same proteins viewed at AS = 120 removes the over-fractionation issue but produces an SSN that fails to discriminate the function. Unfortunately, automatic algorithms, and sometimes manual curation by knowledgeable researchers, will make incorrect assumptions, compounding the well-known misannotation problem.[8] Another difficulty is the potential for multiple functional splits between proteins within a clade through paralogous duplication or acquisition of new function in an ortholog,[9] resulting in nearly identical proteins with different functions (Figure S1). The relationship between proteins of the same function in such a clade can be described as paraphyletic. Such cases are recently illustrated by the flavin-dependent amine oxidase superfamily (IPR002937)[10] and BesD-related radical halogenases.[11] Several methods to overcome the difficulty of paralogous gene separation have been developed, including genome neighborhood analysis and identification of signature motifs to segregate paralogous sequences by function.[7,12] Despite these efforts, significant challenges remain in proper annotation for proteins that reside in multifunctional clades.

We hypothesized that reliance on full-length sequence similarity, which treats every position of an enzyme with equal weight, contributes to the above-stated difficulties. Full-length sequence analysis dilutes the outsized role that residues in and around the catalytic site play in defining reaction chemistry. Previous studies have established that sequence conservation follows a spatial gradient radiating outward from the enzymatic active site, with residues proximal to the catalytic site exhibiting the highest conservation levels.[13] The conservation gradient can be quantified using enzyme "active site conservation (ASC) ratios", defined as the percent similarity of residues within a 10 Å radial sphere around the active site divided by the full-length percent similarity of any two sequence homologous enzymes.[14] Homologous enzymes with the same function generally yield ASC ratios >1, while closely related enzymes catalyzing distinct reactions generally give ASC ratios <1, consistent with the residues near the active site governing the reaction chemistry.[14] Widespread adoption of ASC ratios to enhance enzyme functional prediction has been hindered by the lack of high-resolution structures. However, this former obstacle may have been removed by the recent availability of >200 million AlphaFold predicted structures that are retrievable from UniProt.[1,15,16] Since rSAM enzymes represent an exceptionally large and functionally diverse superfamily, we chose it as a test bed for the broader application of ASC analysis to aid in protein functional prediction. These efforts suggested the need for a streamlined procedure that could be more readily adapted, which we developed and termed catalytic site proximity analysis.

Among the many characterized rSAM enzymes, a significant portion is involved in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs). A large group of rSAM-dependent RiPPs contain thioether linkages.[17,18] The class-defining rSAM enzymes can be broadly divided into two categories: (i) those forming sulfur-to-α carbon thioether-containing peptides (sactipeptides) and (ii) those forming radical non-α thioether-containing peptides (ranthipeptides; Figure 1 and Table S2). While ranthipeptide synthases (ranthisynthases, hereafter) are

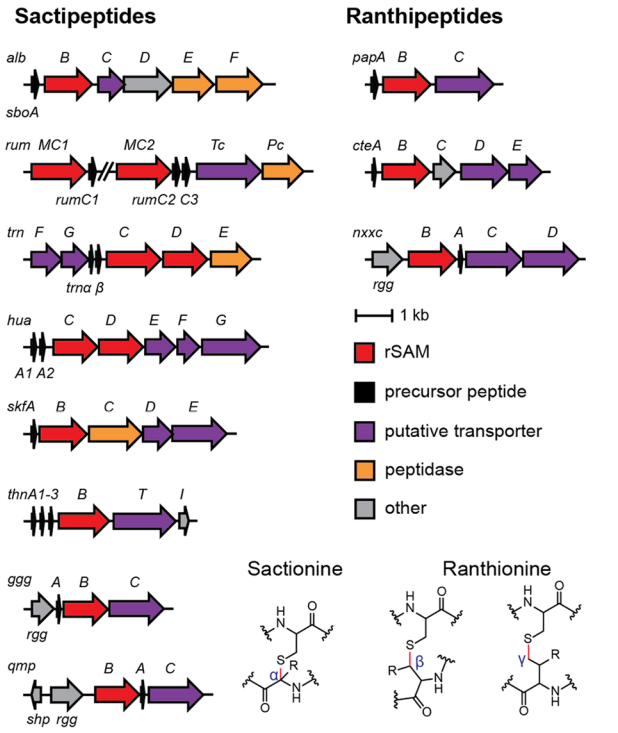


Figure 1. Sactipeptide and ranthipeptide gene cluster diagrams. Simplified BGC diagrams of characterized sactipeptides and ranthipeptides. Class-defining modifications for sactipeptides and ranthipeptides are provided. Precursor peptide sequences are given in Table S2.

highly similar to one another, sactipeptide synthases (sactisynthases, hereafter) can display considerable sequence divergence.[17,19] For example, the sactisynthases from the subtilosin A (AlbA) and thurincin H (ThnB) pathways are highly similar, but they are quite divergent from the sactisynthases from the thuricin CD, huazacin, and ruminococcin C pathways. In fact, the latter three sactisynthases, which form a monophyletic clade, are more sequence similar to known ranthisynthases than to the former two sactisynthases (Figure S2).[17,20−25] The sporulation killing factor, enteropeptins, streptosactin, and suisactin all represent other sactipeptides modified by sequence-divergent sactisynthases (SkfB, KgrC, GggB, and QmpB, respectively).[26−29] Interestingly, AlbA, ThnB, KgrC, and SkfB are paraphyletic, residing in a clade with rSAM enzymes that perform C−O, C−C, and rearrangement chemistry (Figure S2).

We have previously attempted to define the genomic landscape of sactipeptides and ranthipeptides.[17] However, the approach was only effective at identifying biosynthetic gene clusters (BGCs) encoding rSAM proteins closely related to known sactisynthases because sactisynthases form multiple sequence-divergent groups. This high sequence divergence (Figure S2) prompted the hypothesis that additional BGCs would exist wherein the responsible sactisynthases would be insufficiently similar to permit confident retrieval by BLASTP. For most prokaryotic RiPPs, including sactipeptides, the interaction between the precursor peptide and the modifying enzymes is governed by a ∼90-residue RiPP precursor recognition element (RRE).[30] We therefore sought to collect

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251

all RRE-associated rSAM proteins to search for additional occurrences of divergent sactisynthases. Such a dataset might also facilitate the discovery of new RiPPs and rSAM chemistry.

Here, we report a UniProt-derived dataset of all RRE-associated rSAM proteins, with entries classified based on overall similarity to characterized enzymes. We analyzed the dataset to identify likely sactisynthases using catalytic site proximity profiling. As a proof of concept, we describe a new sactipeptide from *Streptomyces sparsogenes*. The responsible sactisynthase StsB has greater full-length sequence similarity to the mycofactocin synthase MftC than to any known sactisynthase. Furthermore, we demonstrate that modified StsA contains three Cys-to-Gly sactionine linkages. These insights were further leveraged to predict six new sactipeptide groups that evaded detection by traditional genome-mining techniques. We further analyzed a group of sequence-diverse cyclophane-forming rSAM proteins and found that their catalytic site proximity profiles are consistent with their similar function despite lower full-length sequence similarity. This strategy offers a means to predict/refine the sequence-function space for rSAM proteins.

RESULTS AND DISCUSSION

Collection and Curation of All RRE-Associated rSAM Proteins

We combined RRE-Finder, a tool that predicts the presence of RRE domains in queried protein sequences,[31] with information available in RadicalSAM.org to generate a set of all RRE-associated rSAM proteins. We considered fused and discrete RRE architectures to compile the dataset. Fused RRE rSAM proteins were cataloged by collecting all rSAM proteins listed in RadicalSAM.org and the RRE-Finder datasets. A tolerant bitscore was used, and false positives were manually removed (Methods). Radical SAM proteins associated with discrete RRE domains were cataloged using a twofold approach to ensure maximum coverage. First, sequences in UniProt with RRE-Finder bitscores of 20 and higher were analyzed by the Enzyme Function Initiative-Genome Neighborhood Tool (EFI−GNT).[1,7,31] Any case where a member of protein family PF04055 occurred within three open-reading frames of any detected RRE was assigned as an RRE-associated rSAM protein. Inversely, all members of the protein family PF05402 (a broad RRE model) were analyzed by EFI−GNT. Any cases where a PF04055 member was found within three open-reading frames of any PF05402 member were retained. After combining sequences from the three strategies and removal of duplicates, we obtained a compendium of ∼15,000 RRE-associated rSAM proteins (UniProt) with a high probability of involvement in RiPP biosynthesis (Supplemental Dataset).

An SSN for the ∼15,000 RRE-associated rSAM proteins was then generated using the Enzyme Function Initiative-Enzyme Similarity Tool (EFI−EST) (Figures S3 and S4).[7] After AS optimization, several groups, such as the PqqE-like enzymes and ranthisynthases, exhibited high convergence ratios (ratio of edges formed to the total number of possible edges in a group), indicating likely isofunctionality. Consistent with previous observations, the sactisynthases partitioned into several smaller groups that do not form an isofunctional group at any AS.

Update to RODEO Sactipeptide/Ranthipeptide Modules

We previously developed a sactipeptide/ranthipeptide scoring module for RODEO, an artificial intelligence tool that analyzes the genomic neighborhood of a query protein and returns a list of putative precursor peptides scored on their likelihood of belonging to a particular RiPP class.[32] To improve module performance, we updated the heuristic scoring and support vector machine classifier to provide better coverage for newly discovered and sequence-diverse sactipeptides (Methods, Figure S5, Table S3). With the new module, we observed fewer false positives among non-thioether-forming rSAM enzymes and fewer false negatives among the population of known thioether-forming rSAM enzymes. At a score threshold of 20, we observed a recall of 100% with 98% precision and an F1 score of ∼0.99 in the test set (Figure S5). We next used the updated module to analyze the RRE-associated rSAM dataset (Methods) after converting the requisite UniProt accession codes into the corresponding NCBI protein codes (see Supplemental Note), which yielded 11,475 rSAM proteins (Figure 2).

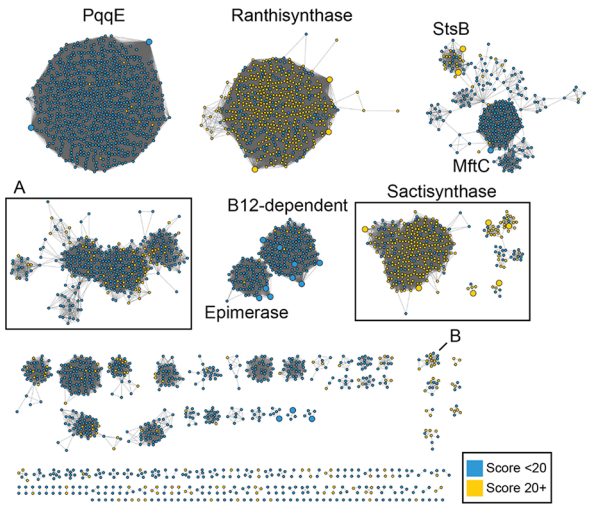


Figure 2. SSN of RRE-associated rSAM proteins. Represented are 11,475 protein sequences (UniProt-identified and successfully mapped to NCBI) visualized at AS = 50 and RepNode = 60 (i.e., edges are drawn between two nodes if they share a BLAST expectation value of $\sim 10^{-50}$ or lower, and any sequences with 60% identity or higher are represented as a single node; full dataset: Figures S3 and S4). Groups are labeled based on the function of characterized rSAMs. No AS isofunctionally groups all known sactisynthases, given their divergence and higher similarity to non-sactisynthases. Larger nodes represent characterized enzymes; yellow, predicted sacti- or ranthisynthase (cognate precursor peptide received a Rapid ORF Detection and Evaluation Online (RODEO) score ≥ 20); blue, not predicted as a sacti- or ranthisynthase. Groups A and B contain previously undetected (predicted) sactisynthases.

The resultant SSN was unable to unify the sactisynthases into a single group, consistent with their high divergence and paraphyletic relationships. We created several hidden Markov models (HMMs) based on the known enzymes to determine if the greater sensitivity in detecting sequence homology would circumvent the shortcomings of BLAST expectation values in identifying additional sactisynthases. Unfortunately, this collection of custom HMMs did not retrieve any additional sequences outside of the already known sactisynthase groups (Figures S3 and S4). Therefore, the HMM-based collection was declared insufficient for sactisynthase detection. During the course of this analysis, we noticed that the group containing mycofactocin synthase (MftC)[33] displayed a

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251

lower convergence ratio and the locally encoded precursor peptides received scores consistent with assignment as either a sactipeptide or a ranthipeptide (conserved $CxCxCx_4GxGxG$ core region motif, Figures 3 and S6). Full-length sequence

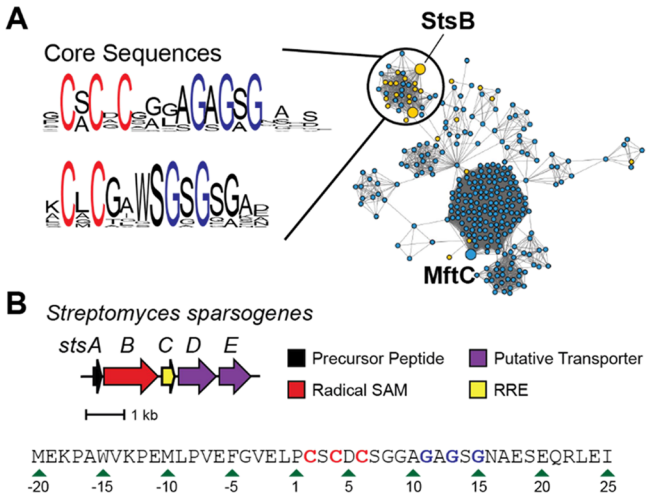


Figure 3. Identification of a putative sactisynthase group related to MftC. (A) Conservation of core sequences in the indicated StsB subgroup (abridged from Figure 2). Logos for full-length precursor peptides are shown in Figure S6. (B) BGC from *S. sparsogenes* associated with the production of modified StsA. The leader peptide cleavage site of StsA is unknown. The numbering system used sets position one as the Pro directly preceding the first conserved Cys.

similarity analysis, the gold standard for protein function prediction, suggests that this group would function similarly to MftC, which performs a two-step reaction on the MftA precursor peptide: (i) oxidative decarboxylation of the C-terminal Tyr and (ii) radical cyclization of the penultimate Val to yield a dimethylated pyrrolidinone.[33] However, this reaction cannot occur thoughout the entire group due to the lack of the essential C-terminal Val-Tyr motif (Figure S6). Included in this group is a rSAM enzyme (hereafter StsB) from *Streptomyces sparsogenes*, which was chosen for further characterization (Figure 3).

Catalytic Site Proximity Profiling of Sactisynthases

We hypothesized that StsB represented another divergent sactisynthase despite it being a poor match to other known sactisynthases based on the full-length similarity. Furthermore, we thought that divergent sactisynthases may share greater conservation when only considering residues within 10 Å of the active site (see Methods). To test this idea, we first determined whether the six largest SSN groups containing a characterized rSAM enzyme had measurably higher active site conservation (ASC) compared to the full-length sequence. Indeed, the median ASC ratios were > 1 in all six tested groups (MftC group = 1.2; PoyB group = 1.6; CteB group = 1.6; TrnC group = 1.8; PqqE group = 1.3; AlbA group = 1.3, Figure S7). The ASC ratios, as expected, vary between groups owing to the membership proteins displaying different rates of divergence. We next examined whether the CteB group displayed the same conservation gradient noted in other enzyme superfamilies.[13] The analysis of secondary shell residues 10−15 Å from the active site gave a median conservation ratio of 1.3, clearly reduced from the ASC ratio of 1.6 for residues <10 Å from the active site.

The above analysis permits the conclusion that across many functionally distinct, RRE-associated rSAM proteins, residues within 10 Å of the active site are more conserved than the full-length sequence.[14] However, we noticed two outlier proteins in the PoyB group with ASC ratios of 1.1 and 1.2 (compared to the group average of 1.6). With ASC ratios >1, the functional annotation as B12-dependent *C*-methyltransferases is nevertheless retained. In contrast, StsB within the MftC group displays a median ASC ratio of 0.96; thus, residues within 10 Å of the active site are less conserved than the full-length protein. Given the MftC group median ASC ratio of 1.2, MftC and StsB are predicted to catalyze different reactions.

We next explored if these findings could be leveraged to unify divergent sactisynthases. To isolate the most functionally determinant portion, we focused on residues proximal to the catalytic site in the characterized members of the RRE-associated rSAM dataset (Figure S8). We selected residues that were (i) within 10 Å of the active site, previously identified in ASC ratio generation, (ii) with side chains facing toward the active site, and (iii) could be identified from a multiple sequence alignment using sequence motifs as landmarks. The latter two steps were added to winnow down the list of residues to those most functionally pertinent and to reduce the analytical complexity that results by considering all residues within 10 Å of the active site. As sactisynthases possess a SPASM or twitch domain, which host auxiliary [Fe−S] centers,[5] we focused on RadicalSAM.org group Mega 1−1 (i.e., SPASM/twitch).[6,34] The crystal structure of the ranthisynthase CteB (PDB: 5WHY) was used as a prototype (Figures 4 and S9).[35] We identified six residues within the 10 Å active site sphere facing into the catalytic site of CteB that lined the substrate-binding cavity near the predicted site of

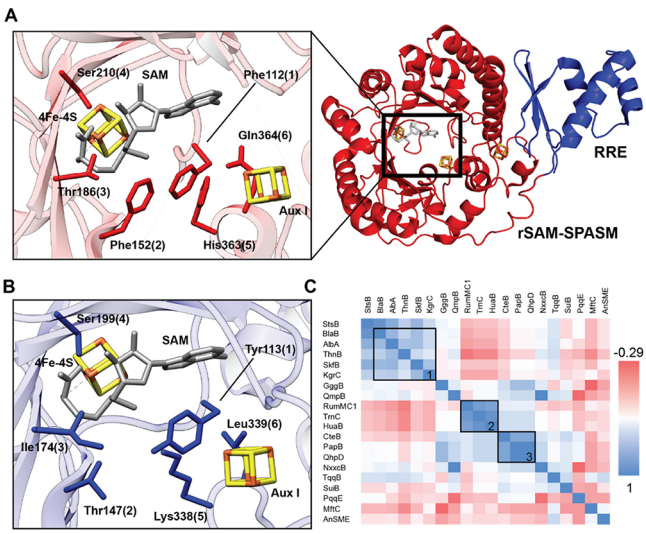


Figure 4. Catalytic site proximity profiling of rSAM proteins. (A) Crystal structure of CteB (PDB 5WHY). Catalytic site proximity residues are red and numbered 1−6. (B) AlphaFold predicted the catalytic site of StsB. Catalytic site proximity residues are blue and numbered 1−6. The [Fe-S] centers and SAM were imported after structural alignment with CteB. (C) Similarity scores for the six catalytic site proximity residues for known sactisynthases (see boxes 1−2) and ranthisynthases (box 3). Blue indicates high catalytic and functional similarity while red suggests the opposite. Non-thioether-forming rSAM proteins TqqB, SuiB, PqqE, MftC, and AnSME are included as outgroups. See Methods for details on the scoring.

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251

radical initiation [CteB: Phe112, Phe152, Thr186, Ser210, His363, Gln364]. CteB-Phe112 directly follows the CxxxCxxC motif that comprises the SAM-binding [4Fe−4S] center. CteB-Phe152, -Thr186, and -Ser210 reside at the ends of three conserved, parallel β strands that run from the surface of the protein toward the SAM-binding site. Finally, CteB-His363 and -Gln364 reside in a loop next directly following a Cys ligand for an auxiliary [Fe−S] center. This loop is present in SPASM and twitch domain-containing rSAM enzymes. We numbered these positions 1−6 based on their N-to-C terminal locations. In accordance with step (iii), all six residues can be readily detected from multiple sequence alignments based on their adjacency to conserved motifs (Figure S10).

We then applied an identical analysis to biochemically characterized sactisynthases and ranthisynthases without experimentally determined structures using AlphaFold. We compared the identities of the six catalytic site proximity residues of CteB to known sacti- and ranthisynthases as well as non-thioether-forming rSAM proteins as functional outgroups. In all cases examined, all six residues were found in the same relative positions as in CteB, suggesting the potential for the comparative analysis of this group of residues (Figures 4 and S9).[35−37] We then profiled the six catalytic site proximity residues and generated a similarity/identity plot (Figures S11 and S12). Catalytic site proximity profiling successfully sorted the hairpin-installing sactisynthases into two groups (groups 1 and 2 in Figure 4, group 3 is QhpD-like ranthisynthases). Group 1 was successfully unified despite the clade being polyfunctional (Figures S1 and S2). These results further show that StsB is a strong match to group 1 sactisynthases and a particularly poor match to MftC. Lastly, we note that the nonhairpin-forming sactisynthases GggB and QmpB are more similar to ranthisynthases upon full-length analysis and catalytic site proximity profiling.[28,29] We speculate that GggB/QmpB may have evolved from ancestral ranthisynthase to become a sactisynthase while retaining ring connectivity processing akin to PapB/NxxcB.

Confirmation of Sactionine Linkages in Modified StsA

Given the StsA sequence (Figure 3) and StsB catalytic site proximity profile (Figure 4), we next tested whether StsB was a bona fide sactisynthase. StsA was expressed in *Escherichia coli* as a fusion to maltose-binding protein (MBP) along with StsB and StsC (RRE). Following the purification and removal of the MBP tag, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) analysis indicated a 6 Da mass loss relative to unmodified StsA, consistent with three thioether linkages (Figure 5). Treatment with excess iodoacetamide showed no evidence of *S*-alkylation, indicating the modification of the three Cys of StsA (Figure S13). Omission of StsC (RRE) from the co-expression resulted in a mixture of unmodified, partial, and fully modified StsA, which was corroborated by the detection of mono-, di-, and tri-alkylated products after iodoacetamide treatment.

Modified StsA was then treated with endoproteinase GluC to afford a smaller peptide amenable for structural characterization. Unexpectedly, GluC digestion preferentially occurred at Glu(-11) and Glu18; proteolysis at Glu(-6) and Glu(-2) were observed as minor products only after extended reaction time (Figure 3). The proteolytic fragment containing the three Cys of StsA was 6 Da lighter than that predicted for unmodified StsA. Hereafter, the term StsA(S-Cα)$_{GluC}$ refers to the Met(-10)−Glu18 fragment of StsA after the StsB/C-modification. HPLC-purified StsA(S-Cα)$_{GluC}$ was analyzed by collision-induced dissociation (CID), which yielded daughter ions with good coverage except for the Cys2-Gly15 region (Figure S14, Table S4). We then prepared a larger quantity of StsA(S-Cα)$_{GluC}$ to pursue NMR-based structure determination. Several attempts to optimize the acquisition conditions were unsuccessful due to solubility challenges.

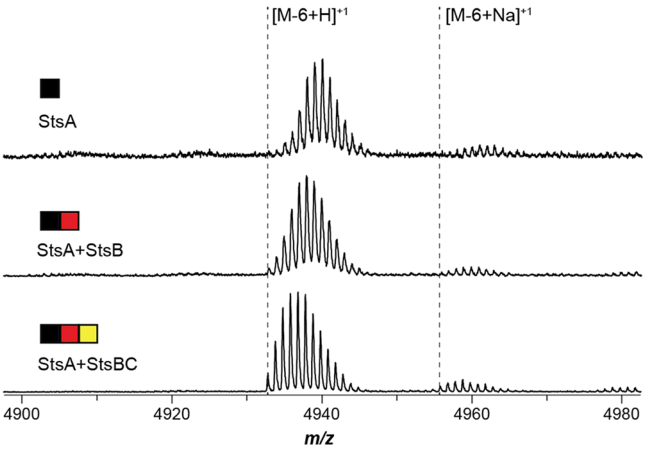


Figure 5. MALDI-TOF-MS of StsA. *Top*, unmodified StsA. *Middle*, StsA co-expressed with rSAM (StsB). *Bottom*, StsA co-expressed with StsB and the cognate RRE, StsC.

In exploring alternatives to establish the linkage chemistry of modified StsA, we were inspired by a previous study that employed reductive desulfurization ($NiCl_2/NaBH_4$).[38] Such treatment of the sactipeptide subtilosin removed the sactionine linkages and permitted traditional amino acid analysis. We reasoned that reductive desulfurization of modified StsA using sodium borodeuteride would selectively introduce deuterium into the sactionine donor and acceptor residues (Figure S15). For instance, a sactionine linkage between Cys and Gly would result in monodeuterated Ala and monodeuterated Gly after $NiCl_2/NaBD_4$ treatment. StsA(S-Cα)$_{GluC}$ was thus subjected to parallel reductive desulfurization reactions, one using $NaBH_4$ and the other with $NaBD_4$. The former reaction yielded a peptide with a mass 136 Da lighter than StsA(S-Cα)$_{GluC}$, consistent with the conversion of Met(-10) into homoalanine and replacement of three sulfur atoms with six hydrogens. The $NaBD_4$ reaction yielded a peptide 7 Da heavier than conditions using $NaBH_4$, consistent with monodeuterated homoalanine and three sulfur atoms replaced with six deuteriums (Figure S15). The desulfurized (linear) peptide was then subjected to CID and the resulting near complete b and y-ion series demonstrated monodeuteration at Cys2, Cys4, Cys6, Gly11, Gly13, and Gly15 (Figures S16−S18, Tables S5−S7). The deuteration pattern was also consistent with residue conservation observed in similar precursor peptides (Figure 3). As Gly-linked thioethers can only occur at the α-carbon, this analytical strategy was sufficient to confirm StsB is a sactisynthase. In the event of a sactipeptide containing linkages to non-Gly residues, this MS-based technique would not distinguish which carbon was modified without additional incorporation of isotopic tracers.

To further evaluate Gly as the sactionine acceptor residues, and to determine the connectivity of the three Cys-Gly linkages, we co-expressed Cys to Ala and Gly to Ser variants of StsA with StsB/C in *E. coli*. Following purification, GluC digestion, and MALDI-TOF-MS analysis, all six variants gave a

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251

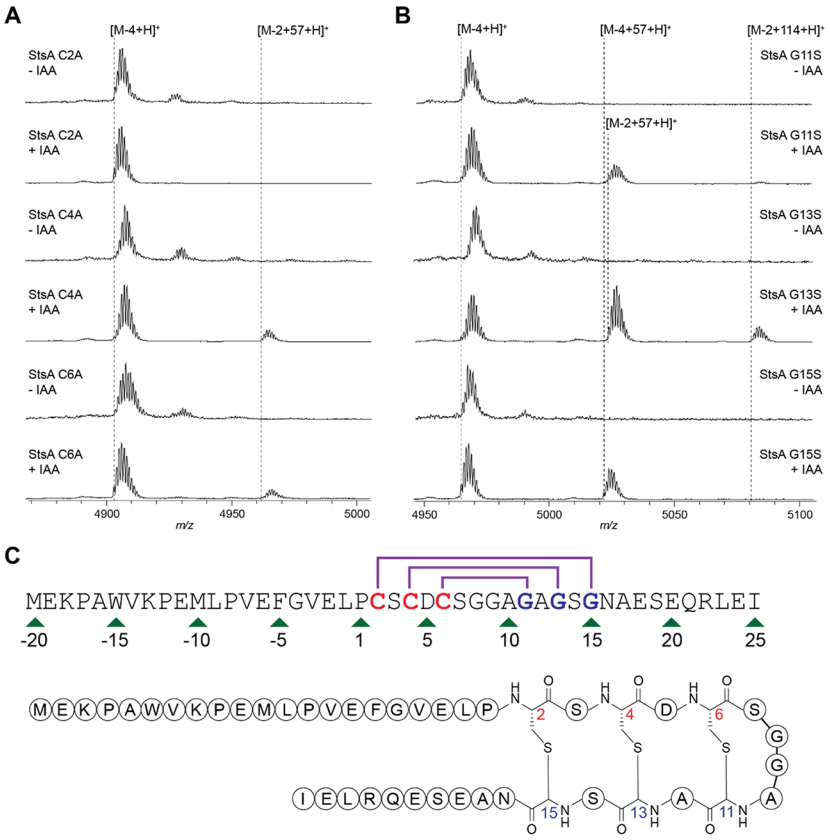


Figure 6. MALDI-TOF-MS of StsA variants. (A) StsA-Cys2, Cys4, and Cys6 were individually replaced with Ala and subjected to iodoacetamide (IAA) labeling (mono- and di-*S*-alkylation yielded +57 and +114 Da, respectively). (B) StsA-Gly11, Gly13, and Gly15 were individually replaced with Ser and similarly treated with IAA. These results were consistent with a hairpin structure where the outermost (Cys2-Gly15) sactionine was not needed to install the other two crosslinks. CID analysis supports these assignments (Figures S19–S23). (C) Proposed structure of modified StsA.

monoisotopic mass four Da lighter than unmodified StsA, supporting the presence of two sactionines (Figure 6). When Cys4 or Cys6 was converted to Ala, we detected mono-*S*-alkylation after iodoacetamide treatment, indicating less efficient processing when one of these two sactionines was absent. Similarly, when Gly11 and Gly13 were replaced with Ser, one sactionine and di-*S*-alkylation were observed. This was not the case with the G15S variant, which only reacted once with iodoacetamide. We reasoned that modified StsA may form a hairpin structure (Cys2-Gly15, Cys4-Gly13, and Cys6-Gly11 linkages) and the compact structure might also explain decreased susceptibility to proteolysis at Glu(-6) and -Glu(-2). To confirm or refute this hypothesis, we subjected five variants to CID with the resulting fragmentation patterns consistent with a hairpin-like connectivity (Figures S19–S23, Tables S8–S12). Importantly, the iodoacetamide-labeled G13S variant produced a fragmentation pattern, suggesting an isobaric mixture of the two parent peptides, each with 4 Da mass loss before alkylation (Figure S22, Table S11). One product displayed two sactionines, Cys4 *S*-alkylation, and a core resistant to fragmentation. The second product contained a Cys6-Gly11 sactionine and a Cys2-Cys4 disulfide (with Met alkylation), indicating that the outermost sactionine (i.e., Cys2-Gly15) forms inefficiently without prior Cys4-Gly13 sactionine formation.

While the current manuscript was in preparation, a publication describing a new sactipeptide became available.[39] The precursor peptide BlaA contains two conserved Cys and Gly residues and is related to StsA (Figures 3 and S6). Similarly, the rSAM enzyme BlaB is related to StsB and MftC (expectation values of 1×10^{-68} and 3×10^{-33}, respectively). Catalytic site proximity profiling of BlaB places it in group 1 with other hairpin-forming sactisynthases (Figure 4). Perhaps, due to increased conformational flexibility (i.e., only two sactionines) and a hydrophilic Ser-Lys-Asn containing loop, modified BlaA was amenable to NMR spectroscopic analysis. Such data unambiguously demonstrated hairpin-like sactionine connectivity between Cys13-Gly22 and Cys15-Gly20 of BlaA.

Enzymatic Activity of Catalytic Site Proximity Variants

We next used site-directed mutagenesis to probe if the catalytic site proximity residues of StsB were critical for StsA modification. StsB variants Y113A, T147A, I174F, S199A, K338A, and L339F were prepared and co-expressed with StsA/C. The reaction products were then purified and analyzed by MALDI-TOF-MS (Figure S24). The activity of StsB-S199A was barely detectable while the other five variants were devoid of activity under the conditions used. To augment these data, we constructed and tested StsB variants F146A, M173A, and V198F under identical expression conditions. These variants are directly adjacent to three of the critical sites but reside on the opposite face of the respective β strands; thus, the side chains face away from the active site (failing step ii of catalytic site proximity residue selection, see above). While StsB-V198F processed StsA with activity equivalent to wild-type, variants F146A and M173A showed masses consistent with the formation of 1–2 sactionine rings. In all three cases, the

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240–251

outward-facing variants performed significantly better than the neighboring catalytic site proximity residue variants.

Analysis of New High-Scoring Sactipeptide Precursors

After confirmation that StsB is a sactisynthase, we revisited the set of RODEO-predicted sactipeptides and ranthipeptides. Following manual dataset curation, we identified 3,421 putative sactipeptide and ranthipeptide BGCs (Figure S25). The Supplemental Dataset was further evaluated by generating a precursor peptide SSN and analyzing every group containing >5 unique sequences (Figures S26−S30). Six of these groups were not associated with any characterized rSAM enzyme, and for such cases, the associated rSAM enzyme was subjected to full-length and catalytic site proximity profiling (Figures S29 and S30). These analyses predict that all six new groups will be sactisynthases. Traditional bioinformatics methods fail to yield a reliable prediction, given their membership in an uncharacterized and polyfunctional group (groups A and B in Figure 2).

Catalytic Site Proximity Profiling in Cyclophane-Forming rSAM Proteins

We next investigated if the utility of catalytic site proximity profiling extends beyond thioether-forming rSAM proteins. Darobactin is an rSAM-modified RiPP featuring a C-O linked macrocycle between two Trp residues and a Csp^2-Csp^3 linked (Trp-Lys) macrocycle (Figure S31).[40] A single rSAM protein (DarE) forms both linkages.[40,41] Outside of the local darobactin group, DarE exhibits higher sequence similarity to anaerobic sulfatase-maturing enzyme (AnSME) than to any other known enzyme. However, AnSME carries out a dissimilar reaction: conversion of Ser/Cys to formylglycine in the active site of anaerobic sulfatase.[42] Catalytic site proximity analysis of DarE using positions equivalent to those previously identified above for the sacti/ranthisynthases yielded a profile unlike AnSME. Instead, the profile was highly similar to other cyclophane-forming rSAMs (e.g., XncB and SjiB). The reactions catalyzed by XncB and SjiB are also Csp^2-Csp^3 couplings between an aromatic side chain and a methylene unit of another residue. Thus, there is a strong chemical similarity between DarE, XncB, and SjiB.[43,44] The recently reported DynB,[45] which installs Csp^2-Csp^3 (Trp-Asn) and N-Csp^3 (His-Tyr) linkages in dynobactin A, also produces a catalytic site proximity profile matching DarE, XncB, and SjiB. Broader scale analyses will be required to establish how often catalytic site proximity preservation will overturn a full-length functional prediction, but with RiPP-modifying rSAM proteins, it occurs for both thioether- and cyclophane-forming pathways.

Comparison with the "RiPP-RaS" Dataset

A set of RiPP-modifying rSAM proteins was recently reported using colocalized ABC transporter genes as a bioinformatic filter.[46] We were interested to examine the extent of overlap between our RRE-associated dataset and the ABC transporter-associated dataset termed "RiPP-RaS". We found only 3,429 overlapping rSAM proteins between the two datasets, despite both coincidentally containing ∼15,000 members (Figure S32). We next mapped the shared rSAM proteins onto SSNs of both datasets and found that the RRE-dependent rSAM proteins present in the RiPP-RAS dataset were well accounted for in our dataset. This comparison indicates that our RRE-associated dataset does well at accounting for RRE-associated rSAMs, but that investigation of rSAMs using a different bioinformatic filter may yield additional interesting RiPP-modifying rSAMs that are not RRE-associated.

CONCLUSIONS

In this work, we have created a dataset of ∼15,000 RRE-associated rSAM proteins and sorted the dataset based on sequence similarity to known rSAM enzymes. Due to the divergent nature of sactisynthases, a thorough sequence conservation analysis of sactisynthases was conducted, which led to the discovery of a new sactipeptide from *Streptomyces sparsogenes*. The requisite rSAM enzyme StsB possesses a catalytic site proximity profile closely resembling known hairpin-forming sactisynthases. The catalytic site proximity profile of StsB was remarkably dissimilar to MftC (mycofactocin biosynthesis), despite MftC sharing the highest full-length sequence similarity of any characterized rSAM enzyme. This finding suggested that StsB is a sactisynthase, which was confirmed by repurposing a reductive desulfurization technique to allow for selective deuterium incorporation at the carbon atoms previously engaged in thioether formation. The use of site-directed mutagenesis and high-resolution/tandem mass spectrometry established a hairpin-like ring connectivity for the three Cys-Gly sactionines of modified StsA. The sactipeptide and ranthipeptide scoring module of RODEO was further updated to take into account new sactipeptide discoveries and to map additional occurrences of sactipeptides that have evaded detection using traditional genome-mining methods. In summary, the disclosed approach shows strong synergy between several sequence analysis tools, including RadicalSAM.org, RODEO, RRE-Finder, and AlphaFold. We note that the utility of catalytic site proximity profiling analysis is likely dependent on (i) the size of the enzyme superfamily, (ii) the functional diversity of the enzyme family, and (iii) whether the reaction is under enzymatic or substrate control. For these reasons, we believe that the approach will have the greatest value when combined with existing bioinformatic techniques. We further anticipate that the detection of catalytic site proximity residues will be most effective in large superfamilies with existing structural characterization to allow for accurate active site identification. We believe that this analysis will be generalizable and applicable to other large and functionally diverse superfamilies.

METHODS

Materials

Materials and reagents were purchased from Gold Biotechnology, Fisher Scientific, or Sigma-Aldrich unless otherwise noted. Yeast extract and tryptone were purchased from Research Products International. Molecular biology reagents for cloning (e.g., restriction enzymes, Q5 polymerase, T4 DNA ligase, and deoxynucleotides) were purchased from New England Biolabs. Oligonucleotide primers and gene blocks were obtained from Integrated DNA Technologies. DNA spin columns were purchased from Epoch Life Sciences. Sanger sequencing was performed by the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign). Polymerase chain reactions were performed using a Bio-Rad S1000 thermal cycler. *Escherichia coli* DH5α and BL21(DE3) strains were used for plasmid maintenance and protein overexpression, respectively. Expressed StsA was purified using an Agilent 1200 series HPLC fitted with a 10 × 250 mm C18 column (Macherey Nagel). Mass spectroscopy was performed using a Bruker Daltonics UltrafleXtreme MALDI-TOF mass spectrometer and a ThermoFisher Scientific Orbitrap Fusion ESI-MS using an Advion TriVersa Nanomate 100.

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251

Collection of rSAM Enzymes Containing Fused RiPP-Recognition Elements

A list of UniProt accession identifiers for all rSAM enzymes collected by the InterPro families and Pfams used in www.radicalSAM.org as of InterPro 87.0, including singletons excluded from the web resource, was provided by Professor John Gerlt.[6] RRE-Finder precision mode with a minimum bitscore of 5 retrieved 6067 sequences from this list of rSAM enzymes.[31]

Collection of rSAM Enzymes Neighboring Discrete RREs

A set of all rSAM enzymes with an annotated PqqD neighbor was generated as follows: the Enzyme Function Initiative-Enzyme Similarity Tool (EFI−EST) was used to generate the sequence similarity network (SSN) of all PqqD homologs identified by PF05402. The resultant network was used to generate a genome neighborhood analysis in the EFI−Genome Neighborhood Tool (GNT) with a gene window of 3. This provided a list of 7221 UniProt accessions from neighboring genes annotated by PF04055 (rSAM superfamily).

A second set of rSAM enzymes with a neighboring RRE was generated as follows: All genes in UniProt as of 2021_3 that were identified as RRE-containing by precision mode RRE-Finder with a bitscore minimum cutoff of 20 were collected. The resultant network was used to generate a genome neighborhood analysis in EFI−GNT with a gene window of 3. This yielded 8217 UniProt identifiers for rSAMs (as defined by PF04055) with a neighboring RRE.

The above two datasets were combined, and 6233 duplicates were removed for a total of 9205 rSAMs with neighboring discrete RRE domains (Supplemental Dataset).

Generation of the RRE-Associated rSAM DataSet

The collection of rSAM sequences with neighboring RREs was combined with the collection of rSAMs with fused RREs and 164 duplicates were removed for a total dataset of 15,108 rSAM enzymes associated with RRE domains. These sequences were supplied to EFI−EST at an alignment score (AS) of 70 to construct an SSN. At AS = 70, fused RRE rSAM proteins were removed when the average RRE-Finder bitscore for the group was below 15 and manual inspection failed to identify an RRE. Singletons in the network with a bitscore value of less than 15, the bitscore at which RREs can be confidently assigned without manual inspection, were also discarded without additional analysis.[31] After manual curation, 14,695 sequences remained and were used as the final set of RRE-associated rSAM enzymes (Figure 2).

Training and Validation of an Improved Sactipeptide/Ranthipeptide RODEO Module

Similar to previous reports where class-specific scoring modules were developed for RODEO,[32] we used a support vector machine (SVM) classifier with a radial basis function kernel in addition to a set of manually curated heuristics (Table S3) to classify potential precursor peptides. However, unlike previous iterations of SVM where only a handful of hard-coded combinations of learning parameters were tested on the training set, this iteration utilized a randomized search pattern supplied by the SciKit-learn package.[47] The use of the randomized search enabled a more thorough approach to SVM optimization than previous RODEO versions. Furthermore, the previous RODEO module analyzed all possible open-reading frames (ORFs) for potential precursor peptides. The current version uses Prodigal to produce a set of candidate substrates.[48] This resulted in a significant reduction of false positives. In addition, we incorporated RRE-Finder to identify local RRE domains, which significantly enhances module performance and reliability.[31]

For the training set, 128 putative sactipeptides were used in addition to 200 nonsactipeptides with similar features (including Cys-richness). We excluded the recently reported KgrA as it contains only a single Cys residue, and there currently is insufficient data to make a training set for single Cys sactipeptides.[27] We trained the SVM with 5-fold cross-validation and obtained a classifier that achieved an F1-score of nearly 0.99 (see eq 1). We then ran a set of 57 withheld putative sactipeptides on the final model and observed nearly complete recall (98%).

$$\text{F1-score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (1)$$

Precursor Peptide Analysis and Curation in the RRE-Associated rSAM Network

UniProt accession identifiers from the RRE-associated rSAM dataset were converted to NCBI protein accession identifiers using a custom script (Supplementary Note). The combined NCBI protein identifiers were used a query for the genome-mining tool Rapid ORF Detection and Evaluation Online (RODEO) with sactipeptide/ranthipeptide scoring selected.[17] Precursor peptides were dereplicated using "Prodigal-shorter" to assign the highest likelihood start codons for each possibility predicted by RODEO. Prodigal-shorter permits gene predictions as short as 15 nucleotides and is a modification to Prodigal and "Prodigal-short."[48,49] The rSAM enzymes were then matched to the highest-scoring local precursor peptide, and the score was used to color the SSN of RRE-associated rSAM enzymes (Figure 2).

Following the analysis of the RRE-associated rSAM dataset with the new sactipeptide/ranthipeptide module, all precursor peptides that scored >19 were collected. Redundancies resulting from rSAMs in the same BGC being separately analyzed (yielding identical precursor peptides) were removed. Regular expressions were then used to collect precursor peptides matching the conserved Cys motifs in characterized sactipeptides and ranthipeptides. Finally, remaining high-scoring peptides were declared false positives and removed if (i) the precursor peptide predicted by RODEO was a large gene (>600 bp), (ii) the gene was distant from the rSAM/BGC (>5 open-reading frames generally, but with consideration for BGC architecture), or (iii) the BGC was consistent with the production of a nonsactipeptide or nonranthipeptide. This procedure removed 163 false positives and generated a final dataset of 3421 ranthipeptides and sactipeptides. Our previously published sactipeptide/ranthipeptide dataset contained 3831 precursor peptides.[17] The difference in the number of retrieved precursor peptides results from the database used to generate the set of rSAM enzymes (previously the NCBI nonredundant protein database was used; here, UniProt).

Profile Hidden Markov Model Generation for Characterized Sactisynthases

The Profile hidden Markov models (pHMM) was generated for the known sactisynthases AlbA, RumMC1, SkfB, ThnB, and TrnC (see Table S13 for identifiers). Each UniProt identifier was queried in RadicalSAM.org and multiple sequence alignment (MSA) at AS = 100 was used as the seed alignment. pHMMs were generated using the HMMER 3.0.[50]

Active Site Conservation Ratio Determination and Heatmap Generation

The RRE-associated rSAM dataset was used to generate an SSN at AS = 50 and RepNode = 40 to maximize sequence diversity. Representative rSAM sequences (n = 50) were selected from the six largest groups. Any literature-reported rSAM enzymes were reintroduced as guideposts for the analysis. These sets of sequences were used to produce MSAs using MAFFT using the G-INS-i method.[51] Next, a reference protein was selected from each set to identify residues within a 10 Å radial sphere of the active site (guided by an experimentally determined structure or high-confidence AlphaFold model, see below).[15] The associated columns in the MSA were extracted and used to produce an "inner sphere" MSA.[14] The percent similarity between all members in the full-length and inner sphere MSAs were generated using SIAS (see below) and the ratio between each pair was generated to create histograms of ASC ratios. Heatmaps were generated using Origin 2022 to display the all-by-all analysis.

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251

Generation of 10 Å Radial Active Site Sphere

The central point of the 10 Å sphere in SPASM domain-containing rSAM enzymes was fixed at the midpoint between the predicted site of radical formation upon reductive cleavage of SAM and the nearest Fe atom of the "AuxI" auxiliary Fe-S center of the SPASM domain. All residues with any atom within 10 Å of this point comprise the 10 Å active site sphere. For rSAM proteins without experimentally determined and publicly available structures, the ligated SAM and auxiliary Fe-S center from CteB (PDB: 5WHY) were superimposed following alignment with the AlphaFold model and used to determine the 10 Å sphere center point. For the PoyB group, the 10 Å sphere was centered on the midpoint between the putative site of radical generation following the SAM cleavage and Co from the B12 group.

Validation of AlphaFold Models

We validated that AlphaFold was generating reliable structures by comparing predictive models to reported crystal structures (PDB codes: CteB, 5WHY; SuiB, 5V1Q; PqqE, 6C8V; AnSME, 4K37). In all cases, AlphaFold generated reliable structures that aligned well with the backbone α-carbons of the crystal structures (RMSDs: CteB, 0.93 Å; SuiB, 1.02 Å; PqqE, 1.72 Å; AnSME, 0.97 Å). While this method cannot validate AlphaFold models for rSAM enzymes not represented in the PDB, the profiling method we later employ only requires knowing the general location of a residue be reliable, not a precise location of each atom.

Similarity Matrix Generation

Similarity and identity matrices for full-length sequences and catalytic proximity residues were generated using SIAS: http://imed.med.ucm.es/Tools/sias.html. Similarity scores were generated using the BLOSUM62 substitution matrix and default settings. SIAS generates a normalized similarity score using eq 2:

$$S = \left(\left(\sum M_{ij}\right) + oP_o + eP_e\right) / \sum M_{ii} \quad (2)$$

where the substitution score (M_{ij}) for each pair of amino acids in the alignment is obtained using the BLOSUM62 substitution matrix (o = number of introduced gaps, P_o = gap introduction penalty, e = number of gap extensions, P_e = gap extension penalty, M_{ii} = score for unchanged amino acid residue).

Precursor Peptide Logo Generation

Precursor peptide alignments were generated by MAFFT using the G-INS-i method. For new sactipeptide groups (see Figures S26–S30), MSAs were constructed and used to create sequence conservation logos using Skylign with the "remove mostly-empty columns" and "Information Content–Above Background" options.[52] All other sequence logos were generated using WebLogo.[53]

Molecular Biology Techniques

A modified pETDuet-1 vector was used for StsABC expression. The StsA precursor peptide was fused to the C-terminus of MBP and the construct includes a tobacco etch virus (TEV) protease site. After TEV protease treatment, a Ser-Gly-Ser sequence remains at the N-terminus of StsA. *StsBC* (rSAM and RRE proteins, respectively) were cloned into the second multiple cloning site and were untagged. This vector was used for initial heterologous co-expression experiments. A construct lacking *stsC* was prepared, in which only *stsB* was present in the second multiple cloning site. The generation of StsA point variants was accomplished by restriction cloning of new oligonucleotide sequences into the first multiple cloning site. The oligonucleotide sequences for StsA variants and primers used for cloning are listed in Table S1. *StsA* was similarly cloned into pETDuet-1 with an N-terminal MBP tag and an empty second multiple cloning site for the purification of unmodified precursor peptides. To increase expression yields and improve DNA manipulability, we used a pET28a+ vector containing *E. coli* optimized *stsA* (His-tagged), *B*, and *C* genes synthesized by Twist Bioscience. This construct was used for coexpressions of *stsB* point mutants.

Preparation of Modified StsA and StsA Variants

E. coli BL21 (DE3) cells were transformed with pETDuet_MBP-StsA_StsBC or the point mutant being analyzed (e.g., pET-Duet_MBP-StsA C2A_StsBC) and cultured on lysogeny broth (LB) agar plates supplemented with 50 μg/mL kanamycin. A single colony was used to inoculate a 5 mL culture of LB with 50 μg/mL kanamycin. After overnight growth at 37 °C, 1 L of LB in a 4 L flat bottom flask was inoculated and grown to ∼0.8 OD_{600} with shaking at 220 rpm. Cultures were then cooled at 4 °C for 30 min and then expression was induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). The culture was shaken at 90 rpm for 16 h at room temperature. The cells were harvested by centrifugation at 4000 × g for 15 min. Cell pellets were frozen at −20 °C until purification.

Harvested cells were resuspended in ice-cold lysis buffer [50 mM Tris pH 7.5, 500 mM NaCl, 2.5% (v/v) glycerol, 0.1% (v/v) Triton X-100] supplemented with 3 mg/mL lysozyme and 0.5 mL of a protease inhibitor cocktail [16 mg/mL benzamidine HCl, 6 mM phenylmethylsulfonyl fluoride, 0.1 mM leupeptin, 0.1 mM E64] The samples were subjected to three rounds of sonication for 45 s with 10 min of equilibration at 4 °C between sonication steps. The resultant lysate was clarified by centrifugation at 17,000 × g for 90 min at 4 °C. The supernatant was loaded onto a gravity flow column with amylose resin (pre-equilibrated in cold lysis buffer). The column was then washed with five column volumes of lysis buffer followed by five column volumes of wash buffer [50 mM Tris pH 7.5, 500 mM NaCl, 2.5% (v/v) glycerol]. The MBP-fused peptide was eluted with elution buffer [50 mM Tris pH 7.5, 300 mM NaCl, 2.5% (v/v) glycerol, 10 mM D-maltose]. Eluted protein was collected in Amicon Ultra 15 mL centrifugal filters [30 kDa NMWL (Nominal Molecular Weight Limit)] and concentrated by centrifugation at 3800 × g until the volume had decreased to ∼1 mL. The sample was then subjected to a 10-fold buffer exchange by the addition of storage buffer [50 mM, 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 300 mM NaCl, 2.5% (v/v) glycerol, pH 7.4] and then further concentrated by centrifugation. The resultant aliquots were used for all subsequent analyses.

Purification of Modified StsA

MBP-tagged and StsBC-modified StsA was obtained by the methods described above. The resultant isolates (concentrated to ∼50 mg/mL) were combined with TEV protease in a 1:100 [TEV protease: isolate] ratio by mass and were allowed to equilibrate at room temperature for 1 h to remove the MBP tag. Free MBP was removed by heating the sample at 90 °C for 5 min. The sample was then subjected to centrifugation at 13,000 × g for 20 min to remove the precipitate. The supernatant was transferred to a new tube and acetonitrile (MeCN) was added to 5%. The addition of 5% MeCN resulted in the precipitation of residual proteinaceous impurities. The sample was again subjected to centrifugation at 13,000 × g for 20 min. The supernatant, containing the TEV-cleaved StsA peptide, was purified from any remaining contaminants using an Agilent 1200 series HPLC fitted with a 10 × 250 mm C18 column (Macherey Nagel). A gradient elution was used with solvent A (10 mM ammonium bicarbonate in Milli-Q water) and solvent B (MeCN) according to the following linear gradient combinations: at t = 0 min, 10% B; t = 5 min, 15% B; t = 35, 65% B; t = 37 min, 95% B; t = 39 min, 95% B; t = 42 min, 10% B. The fraction containing StsA was collected and lyophilized to dryness for later experimentation.

Modified StsA for high-resolution mass spectrometry (HR-MS) analysis was combined with GluC in a 1:100 [GluC protease: sample] ratio. The digest was allowed to proceed at 37 °C for 24 h and yielded a product digested preferentially after Glu(-11) and Glu18. The resultant peptide was purified using the same linear gradient and solvents described above.

Iodoacetamide Labeling of Isolates

Alkylation of StsA and variants was performed using 375 mM iodoacetamide in 200 mM ammonium bicarbonate. Alkylation was allowed to proceed in the dark at room temperature for 1 h. Following labeling, the sample was desalted by ZipTip and used for MS analysis.

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240–251

MALDI-TOF-MS Analysis

MALDI-TOF-MS was used to identify peptide products based on their characteristic mass changes upon post-translational modification, protease digestion, or chemical modification. The StsA precursor peptide and fragments were mixed with a matrix consisting of 20 mg/mL of sinapinic acid in 60% MeCN with 0.1% formic acid. The samples were ionized using a Bruker Daltonics UltrafleXtreme MALDI-TOF mass spectrometer in reflector/positive mode and linear/positive mode. Data processing was performed using Bruker FlexAnalysis software.

HR-ESI-MS/MS Analysis

Samples for high-resolution electrospray ionization tandem mass spectrometry (HR-ESI-MS/MS) were either desalted by ZipTip or purified by HPLC as described above. Next, samples were diluted 1:1 into an ESI mix (80% methanol, 19% H_2O, 1% acetic acid). Samples were directly infused into a ThermoFisher Scientific Orbitrap Fusion ESI-MS using an Advion TriVersa Nanomate 100. The MS was calibrated and tuned with Pierce LTQ Velos ESI Positive Ion Calibration Solution (ThermoFisher). The MS was operated using the following parameters: resolution, 100,000; isolation width (MS/MS), 1 m/z; normalized collision energy (MS/MS), 35; activation q value (MS/MS), 0.4; activation time (MS/MS), 30 ms. Fragmentation was performed using CID at 30%. Data analysis was conducted using the Qualbrowser application of Xcalibur software (ThermoFisher Scientific).

Desulfurization and Deuterium Labeling of StsA

Desulfurization of StsA was adapted from previous work on subtilosin A.[38] $NiCl_2$ hexahydrate (500 μg, Acros Organics) was added to a 2 mL screw-capped vial. GluC-digested StsA (125 μg) was dissolved in 150 μL of 60% methanol: Milli-Q water and transferred to the vial followed by 500 μg of sodium borohydride (Fluka). The addition of sodium borohydride causes rapid formation of the nickel boride catalyst as a fine black particulate and gas evolution. The vial was quickly sealed and heated for 5 min at 50 °C. Two further portions of 500 μg sodium borohydride were added followed by 5 min of heating at 50 °C after each portion. The reaction was quenched with 30 μL of trifluoroacetic acid (TCI) and the nickel boride particulates were removed by centrifugation. The supernatant was transferred to a 1.5 mL Eppendorf tube and dried using a Speedvac concentrator (Savant ISS110). The white residue was reconstituted in 100 μL of 1% methanol in Milli-Q water, desalted using a C_{18} ZipTip, and eluted into 80% methanol Milli-Q water with 1% acetic acid. For desulfurization under deuterated conditions, the same method was used with the following modifications: sodium borodeuteride was used in place of sodium borohydride, anhydrous nickel chloride was used in place of the hexahydrate salt, and 60% D_4 methanol−D_2O was used as the reaction solvent (Cambridge Isotope Laboratories).

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsbiomedchemau.2c00085.

Experimental methods, supporting figures/tables, and supporting references (PDF)

Supplemental Dataset. RRE-associated rSAM accession identifiers and listing of sactipeptide ranthipeptide precursor peptides (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Douglas A. Mitchell − *Department of Chemistry and Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States;* orcid.org/0000-0002-9564-0953; Phone: 1-217-333-1345; Email: douglasm@illinois.edu; Fax: 1-217-333-0508

Authors

Timothy W. Precord − *Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

Sangeetha Ramesh − *Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

Shravan R. Dommaraju − *Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States; Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States;* orcid.org/0000-0002-0565-1748

Lonnie A. Harris − *Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

Bryce L. Kille − *Department of Computer Science, Rice University, Houston, Texas 77005, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsbiomedchemau.2c00085

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by grants from NIH/NIGMS (R01 GM123998 and R24 GM141196 to D.A.M.) and the Chemistry-Biology Interface Research Training Program (T32 GM070421 to T.W.P.). B.K. is supported by a fellowship from the National Library of Medicine Training Program in Biomedical Informatics and Data Science (T15 LM007093). We further thank Dr. John Gerlt for providing the sequences from RadicalSAM.org used in our dataset generation.

■ REFERENCES

(1) UniProt: The Universal Protein Knowledgebase in 2021. *Nucleic Acids Res.* **2020**, *49* (), D480−D489.

(2) Booker, S. J. Radical SAM Enzymes and Radical Enzymology. *Biochim. Biophys. Acta* **2012**, *1824*, 1151−1153.

(3) Broderick, W. E.; Hoffman, B. M.; Broderick, J. B. Mechanism of Radical Initiation in the Radical S-Adenosyl-L-Methionine Superfamily. *Acc. Chem. Res.* **2018**, *51*, 2611−2619.

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

(5) Mendauletova, A.; Kostenko, A.; Lien, Y.; Latham, J. How a Subfamily of Radical S-Adenosylmethionine Enzymes Became a Mainstay of Ribosomally Synthesized and Post-Translationally Modified Peptide Discovery. *ACS Bio. Med. Chem. Au* **2022**, *2*, 53−59.

(6) 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.

(7) 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.
(8) Zallot, R.; Oberg, N. O.; Gerlt, J. A. 'Democratized' Genomic Enzymology Web Tools for Functional Assignment. *Curr. Opin. Chem. Biol.* **2018**, *47*, 77−85.
(9) Bershtein, S.; Tawfik, D. S. Ohno's Model Revisited: Measuring the Frequency of Potentially Adaptive Mutations under Various Mutational Drifts. *Mol. Biol. Evol.* **2008**, *25*, 2311−2318.
(10) Tararina, M. A.; Allen, K. N. Bioinformatic Analysis of the Flavin-Dependent Amine Oxidase Superfamily: Adaptations for Substrate Specificity and Catalytic Diversity. *J. Mol. Biol.* **2020**, *432*, 3269−3288.
(11) Neugebauer, M. E.; Kissman, E. N.; Marchand, J. A.; Pelton, J. G.; Sambold, N. A.; Millar, D. C.; Chang, M. C. Y. Reaction Pathway Engineering Converts a Radical Hydroxylase into a Halogenase. *Nat. Chem. Biol.* **2022**, *18*, 171−179.
(12) Zallot, R.; Harrison, K. J.; Kolaczkowski, B.; De Crécy-Lagard, V. Functional Annotations of Paralogs: A Blessing and a Curse. *Life* **2016**, *6*, 39.
(13) Jack, B. R.; Meyer, A. G.; Echave, J.; Wilke, C. O. Functional Sites Induce Long-Range Evolutionary Constraints in Enzymes. *PLoS Biol.* **2016**, *14*, No. e1002452.
(14) Das, R.; Gerstein, M. A Method Using Active-Site Sequence Conservation to Find Functional Shifts in Protein Families: Application to the Enzymes of Central Metabolism, Leading to the Identification of an Anomalous Isocitrate Dehydrogenase in Pathogens. *Proteins* **2004**, *55*, 455−463.
(15) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596*, 583−589.
(16) Baek, M.; DiMaio, F.; Anishchenko, I.; Dauparas, J.; Ovchinnikov, S.; Lee, G. R.; Wang, J.; Cong, Q.; Kinch, L. N.; Schaeffer, R. D.; Millán, C.; Park, H.; Adams, C.; Glassman, C. R.; DeGiovanni, A.; Pereira, J. H.; Rodrigues, A. V.; van Dijk, A. A.; Ebrecht, A. C.; Opperman, D. J.; Sagmeister, T.; Buhlheller, C.; Pavkov-Keller, T.; Rathinaswamy, M. K.; Dalwadi, U.; Yip, C. K.; Burke, J. E.; Garcia, K. C.; Grishin, N. V.; Adams, P. D.; Read, R. J.; Baker, D. Accurate Prediction of Protein Structures and Interactions Using a Three-Track Neural Network. *Science* **2021**, *373*, 871−876.
(17) 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.
(18) 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.
(19) 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.
(20) Zheng, G.; Yan, L. Z.; Vederas, J. C.; Zuber, P. Genes of the Sbo-Alb Locus of Bacillus Subtilis Are Required for Production of the Antilisterial Bacteriocin Subtilosin. *J. Bacteriol.* **1999**, *181*, 7346−7355.
(21) 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*, 6249−6252.
(22) Sit, C. S.; van Belkum, M. J.; McKay, R. T.; Worobo, R. W.; Vederas, J. C. The 3D Solution Structure of Thurincin H, a Bacteriocin with Four Sulfur to α-Carbon Crosslinks. *Angew. Chem. Int. Ed.* **2011**, *50*, 8718−8721.
(23) Wieckowski, B. M.; Hegemann, J. D.; Mielcarek, A.; Boss, L.; Burghaus, O.; Marahiel, M. A. The PqqD Homologous Domain of the Radical SAM Enzyme ThnB Is Required for Thioether Bond Formation during Thurincin H Maturation. *FEBS Lett.* **2015**, *589*, 1802−1806.
(24) Rea, M. C.; Sit, C. S.; Clayton, E.; O'Connor, P. M.; Whittal, R. M.; Zheng, J.; Vederas, J. C.; Ross, R. P.; Hill, C. Thuricin CD, a Posttranslationally Modified Bacteriocin with a Narrow Spectrum of Activity against Clostridium Difficile. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 9352−9357.
(25) Balty, C.; Guillot, A.; Fradale, L.; Brewee, C.; Boulay, M.; Kubiak, X.; Benjdia, A.; Berteau, O. Ruminococcin C, an Anti-Clostridial Sactipeptide Produced by a Prominent Member of the Human Microbiota Ruminococcus Gnavus. *J. Biol. Chem.* **2019**, *294*, 14512−14525.
(26) Flühe, L.; Burghaus, O.; Wieckowski, B. M.; Giessen, T. W.; Linne, U.; Marahiel, M. A. 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*, 959−962.
(27) Clark, K. A.; Covington, B. C.; Seyedsayamdost, M. R. Biosynthesis-Guided Discovery Reveals Enteropeptins as Alternative Sactipeptides Containing N-Methylornithine. *Nat. Chem.* **2022**, *14*, 1390−1398.
(28) 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.
(29) 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.
(30) Burkhart, B. J.; Hudson, G. A.; Dunbar, K. L.; Mitchell, D. A. A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis. *Nat. Chem. Biol.* **2015**, *11*, 564−570.
(31) 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.
(32) Tietz, J. I.; Schwalen, C. J.; Patel, P. S.; Maxson, T.; Blair, P. M.; Tai, H.-C.; Zakai, U. I.; Mitchell, D. A. A New Genome-Mining Tool Redefines the Lasso Peptide Biosynthetic Landscape. *Nat. Chem. Biol.* **2017**, *13*, 470−478.
(33) Khaliullin, B.; Aggarwal, P.; Bubas, M.; Eaton, G. R.; Eaton, S. S.; Latham, J. A. Mycofactocin Biosynthesis: Modification of the Peptide MftA by the Radical S-Adenosylmethionine Protein MftC. *FEBS Lett.* **2016**, *590*, 2538−2548.
(34) 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.
(35) Grove, T. L.; Himes, P. M.; Hwang, S.; Yumerefendi, H.; Bonanno, J. B.; Kuhlman, B.; Almo, S. C.; Bowers, A. A. Structural Insights into Thioether Bond Formation in the Biosynthesis of Sactipeptides. *J. Am. Chem. Soc.* **2017**, *139*, 11734−11744.
(36) 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.
(37) Barr, I.; Stich, T. A.; Gizzi, A. S.; Grove, T. L.; Bonanno, J. B.; Latham, J. A.; Chung, T.; Wilmot, C. M.; Britt, R. D.; Almo, S. C.; Klinman, J. P. X-Ray and EPR Characterization of the Auxiliary Fe−S Clusters in the Radical SAM Enzyme PqqE. *Biochemistry* **2018**, *57*, 1306−1315.
(38) Kawulka, K. E.; Sprules, T.; Diaper, C. M.; Whittal, R. M.; McKay, R. T.; Mercier, P.; Zuber, P.; Vederas, J. C. Structure of Subtilosin A, a Cyclic Antimicrobial Peptide from Bacillus Subtilis with Unusual Sulfur to α-Carbon Cross-Links: Formation and

Reduction of α-Thio-α-Amino Acid Derivatives. *Biochemistry* **2004**, *43*, 3385−3395.
(39) He, B.-B.; Cheng, Z.; Zhong, Z.; Gao, Y.; Liu, H.; Li, Y.-X. Expanded Sequence Space of Radical S-Adenosylmethionine-Dependent Enzymes Involved in Post-Translational Macrocyclization. *Angew. Chem. Int. Ed.* **2022**, *134*, No. e202212447.
(40) Imai, Y.; Meyer, K. J.; Iinishi, A.; Favre-Godal, Q.; Green, R.; Manuse, S.; Caboni, M.; Mori, M.; Niles, S.; Ghiglieri, M.; Honrao, C.; Ma, X.; Guo, J. J.; Makriyannis, A.; Linares-Otoya, L.; Böhringer, N.; Wuisan, Z. G.; Kaur, H.; Wu, R.; Mateus, A.; Typas, A.; Savitski, M. M.; Espinoza, J. L.; O'Rourke, A.; Nelson, K. E.; Hiller, S.; Noinaj, N.; Schäberle, T. F.; D'Onofrio, A.; Lewis, K. A New Antibiotic Selectively Kills Gram-Negative Pathogens. *Nature* **2019**, *576*, 459−464.
(41) Guo, S.; Wang, S.; Ma, S.; Deng, Z.; Ding, W.; Zhang, Q. Radical SAM-Dependent Ether Crosslink in Daropeptide Biosynthesis. *Nat. Commun.* **2022**, *13*, 2361.
(42) Benjdia, A.; Leprince, J.; Guillot, A.; Vaudry, H.; Rabot, S.; Berteau, O. Anaerobic Sulfatase-Maturating Enzymes: Radical SAM Enzymes Able To Catalyze in Vitro Sulfatase Post-Translational Modification. *J. Am. Chem. Soc.* **2007**, *129*, 3462−3463.
(43) Nguyen, T. Q. N.; Tooh, Y. W.; Sugiyama, R.; Nguyen, T. P. D.; Purushothaman, M.; Leow, L. C.; Hanif, K.; Yong, R. H. S.; Agatha, I.; Winnerdy, F. R.; Gugger, M.; Phan, A. T.; Morinaka, B. I. Post-Translational Formation of Strained Cyclophanes in Bacteria. *Nat. Chem.* **2020**, *12*, 1042−1053.
(44) Ma, S.; Chen, H.; Li, H.; Ji, X.; Deng, Z.; Ding, W.; Zhang, Q. Post-Translational Formation of Aminomalonate by a Promiscuous Peptide-Modifying Radical SAM Enzyme. *Angew. Chem. Int. Ed.* **2021**, *60*, 19957−19964.
(45) Miller, R. D.; Iinishi, A.; Modaresi, S. M.; Yoo, B.-K.; Curtis, T. D.; Lariviere, P. J.; Liang, L.; Son, S.; Nicolau, S.; Bargabos, R.; Morrissette, M.; Gates, M. F.; Pitt, N.; Jakob, R. P.; Rath, P.; Maier, T.; Malyutin, A. G.; Kaiser, J. T.; Niles, S.; Karavas, B.; Ghiglieri, M.; Bowman, S. E. J.; Rees, D. C.; Hiller, S.; Lewis, K. Computational Identification of a Systemic Antibiotic for Gram-Negative Bacteria. *Nat. Microbiol.* **2022**, *7*, 1661−1672.
(46) Clark, K. A.; Seyedsayamdost, M. R. Bioinformatic Atlas of Radical SAM Enzyme-Modified RiPP Natural Products Reveals an Isoleucine−Tryptophan Crosslink. *J. Am. Chem. Soc.* **2022**, *144*, 17876−17888.
(47) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* , *12*, 2825.
(48) Hyatt, D.; Chen, G.-L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinform.* **2010**, *11*, 119.
(49) 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.
(50) Eddy, S. R. Profile Hidden Markov Models. *Bioinformatics* **1998**, *14*, 755−763.
(51) Katoh, K.; Rozewicki, J.; Yamada, K. D. MAFFT Online Service: Multiple Sequence Alignment, Interactive Sequence Choice and Visualization. *Brief. Bioinform.* **2019**, *20*, 1160−1166.
(52) Wheeler, T. J.; Clements, J.; Finn, R. D. Skylign: A Tool for Creating Informative, Interactive Logos Representing Sequence Alignments and Profile Hidden Markov Models. *BMC Bioinform.* **2014**, *15*, 7.
(53) Crooks, G. E.; Hon, G.; Chandonia, J.-M.; Brenner, S. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **2004**, *14*, 1188−1190.

https://doi.org/10.1021/acsbiomedchemau.2c00085
ACS Bio Med Chem Au 2023, 3, 240−251