A cell-free workflow for detecting and characterizing RiPP recognition element-precursor peptide interactions.

Derek A. Wong[a,b,c], Maria D. Cabezas[a,b,c], Martin Daniel-Ivad[d,e], Deepali V. Prasanna[a,b,c], Regina Fernandez[a,b,c], Robert Nicol[d], Emily P. Balskus[d,e,f,*], Ashty S. Karim[a,b,c,*], and Michael C. Jewett[a,b,c,g,*]

Affiliations

[a]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA
[b]Chemistry of Life Processes Institute, Northwestern University, Evanston, IL 60208, USA
[c]Center for Synthetic Biology, Northwestern University, Evanston, IL 60208, USA
[d]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[e]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA
[f]Howard Hughes Medical Institute, Harvard University, Cambridge, MA 02138
[g]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA

[*]To whom correspondence should be addressed:

Michael Jewett, Stanford University, 443 Via Ortega, Stanford, CA 94305, mjewett@stanford.edu; Tel (+1) 847 467 5007

Ashty Karim, Northwestern University, 2145 Sheridan Road, Tech E-135, Evanston IL 60208, ashty.karim@northwestern.edu; Tel; (+1) 847 467 0526

Emily Balskus, Harvard University, 12 Oxford Street, Conant 208, Cambridge, MA 02138, balskus@chemistry.harvard.edu; Tel (+1) 617 496 9921

Abstract

Ribosomally synthesized and post-translationally modified peptides (RiPPs) represent a promising class of new therapeutics and antimicrobials. Unfortunately, RiPP discovery efforts are hampered by low-throughput methods for characterizing RiPP recognition elements (RRE), which direct tailoring enzymes to their peptide substrates for RiPP maturation. To address this bottleneck, we report a high-throughput, cell-free workflow for parallelized expression and assaying of RREs with their associated precursor peptide substrates in a process that takes hours instead of weeks. We show the utility of our platform by rapidly scanning precursor peptide sequences for residues important for RRE binding, comprehensively mapping essential residues for RRE binding, and engineering peptides with synthetic RRE recognition sites. We also test 72 computationally predicted lasso peptide RRE and precursor peptide pairs for binding activity, enabling the discovery of a class II lasso peptide. We anticipate that our cell-free workflow will provide a tool for discovering, understanding, and engineering RiPPs.

Introduction

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a class of natural products with high research interest due to their wide variety of potent bioactivities, including antimicrobial[1-3] (e.g., nisin[4], Ruminococcin C[5]) and anti-cancer (e.g., Thioholgamide A[6], prethioviridamide[7]) activity, and large degree of chemical and structural diversity. Composed of an amino acid backbone, RiPPs biosynthetically originate as a precursor peptide composed of an N-terminal leader sequence and C-terminal core sequence[8]. Tailoring enzymes encoded within the same biosynthetic gene cluster (BGC) as the precursor peptide recognize a portion of the leader sequence and install post-translational modifications on the core sequence, producing the mature RiPP[8]. Examples of post-translational modifications include methylation[9,10], glycosylation[11-13], heterocyclization[14,15], and dehydration[16]. Due to the relatively conserved nature of RiPP BGCs, popular search algorithms (e.g., AntiSMASH[17], PRISM[18], RODEO[19], RiPPMiner[20]) can identify on the order of thousands of putative RiPP BGCs using only genome sequences. With the increasing number of sequenced genomes, there is an opportunity to discover and engineer new classes of RiPPs.

In over 50% of RiPP BGCs, a standalone protein or fusion protein containing a RiPP precursor peptide recognition element (RRE) is essential for biosynthesis[21]. RREs are conserved protein

domains with homology to the enzyme PqqD, which is involved in pyrroloquinoline quinone (PQQ) biosynthesis. Acting as a chaperone, PqqD guides PqqE, a radical *S*-adenosylmethionine (rSAM) enzyme, to its peptide substrate, PqqA[22]. Proteins or protein domains found in several RiPP classes share homology to PqqD and function in the same manner, guiding tailoring enzymes to their peptide substrate. In the absence of the RRE, individual reactions catalyzed by the tailoring enzymes often suffer from slow kinetics and low conversion rates[23]. Understanding how these domains function and their individual peptide sequence specificities is important for interrogating the natural diversity of RiPPs as well as discovering new-to-nature RiPPs by engineering RiPP biosynthesis.

Current methods for validating and studying RRE-peptide interactions are low-throughput, typically studying single digits or tens of interactions at a time. For example, fluorescence polarization assays can measure interactions of a fluorescently conjugated precursor peptide with an RRE-containing protein[21,24,25]. However, these assays require a fluorescently conjugated peptide for each individual sequence of interest, which is time consuming to make or expensive to obtain from commercial sources. In some studies, co-crystallization of the known peptide substrate and RRE have been used to elucidate the RRE residues directly interacting with the peptide of interest[26]. While useful for studying interactions between an enzyme and a limited number of peptide substrates, for screening larger RRE-peptide combinations these methods are generally time prohibitive, not scalable, low-throughput, and not able to provide information on the relative affinity of binding interactions without additional experiments using fluorescence polarization.

In this work, we develop a cell-free (or *in vitro*) plate-based workflow that combines cell-free protein synthesis (CFPS)[27-31] with AlphaLISA[32], an in-solution, bead-based ELISA, to evaluate RRE recognition. We begin by validating our workflow with a panel of model RREs and peptides from different classes of RiPPs, demonstrating that we can detect binding interactions between these RREs and their native peptide substrate. Next, we demonstrate the workflow's utility for in depth characterization of a specific RRE-peptide pair by identifying specific residues within a peptide substrate that are important for binding recognition by a computationally identified F protein-dependent cyclodehydratase. We then demonstrate how our workflow can experimentally screen computationally predicted RiPP BGCs, leading to the discovery of a lasso peptide we refer to as Alphalassin. Our workflow addresses some major limitations associated with standard

assays used for characterizing RiPPs and RREs, which we anticipate will accelerate the discovery and engineering of RiPPs.

Results

A cell-free, AlphaLISA based workflow can detect RRE-peptide interactions

Our cell-free, plate-based workflow combines cell-free protein synthesis (CFPS) to express RiPP precursor peptides and RRE-containing proteins with AlphaLISA to semi-quantitatively measure RRE recognition of the precursor peptide. The key goal was to express and quantify protein-peptide interactions in less than a day, rather than weeks that would be needed with typical chemical peptide synthesis approaches and fluorescence polarization. We chose to work with the PURE*frex* CFPS system, a commercially available cell-free expression system which utilizes purified transcription and translation machinery with a user supplied DNA template[33], because of its completely defined reaction environment.

To begin, we selected RREs from three RiPPs – pyrroloquinoline quinone (PqqA – PqqD)[34], Pantocin A (PaaA – PaaP)[35,36] and Streptide (SuiA – SuiB)[37] - and several RiPP classes including a lanthipeptide (NisA – NisB)[38], four linear azole-containing peptides (LAPs) (BalhA1 – BalhC[21,39], CurA – CurC[21,40], HcaA – HcaF[21,41], and McbA-McbB[42]), a computationally predicted thiopeptide (TbtA - TbtF)[21], a mycofactin (MftB – MftA)[43], a sactipeptide (SkfA – SkfB)[44], a ranthipeptide (PapA – PapB)[45], and a lasso peptide (TbiAβ - TbiB1)[26] to assess with our workflow. As a first step, we produced linear expression templates (LETs) via PCR encoding each of the 13 listed RREs and tested their expression in PURE*frex* via incorporation of FluoroTect^{TM} Green$_{Lys}$, fluorescently labeled lysine (**Supplementary Fig. 1**). For 9 of these proteins, we tested the native sequence as well as fusion proteins in which the predicted PqqD domain of the protein was fused either N-terminally or C-terminally to maltose-binding protein (MBP) due to their size and/or origin from an rSAM enzyme potentially making PURE*frex* expression difficult. While many of the full-length constructs did not produce soluble protein, we did observe soluble expression of full-length PqqD, MftB, and TbiB1 fused to MBP, and 8/9 of the fusion proteins composed of the predicted RRE domain fused to MBP.

Encouraged by these results, we then tested the functionality of these RRE-containing proteins in an AlphaLISA based reaction with each of the RREs' respective peptide substrates. AlphaLISA is an in-solution, bead-based assay version of ELISA that is amenable to acoustic liquid handling

robots and small (2 μL) reaction sizes in a 384-well plate format[32] and has previously been used with cell-free systems to assess protein-protein[46,47] and protein-peptide interactions[46,48].

We first expressed MBP-tagged RRE fusion proteins and N-terminally tagged sFLAG peptide substrates in individual PURE*frex* reactions (**Fig. 1a**). We then assayed RRE-peptide recognition by mixing an RRE protein-expressing PURE*frex* reaction and the corresponding peptide substrate-expressing reaction with anti-FLAG donor beads and anti-MBP acceptor beads. Only in instances in which the RRE binds the peptide will the acceptor and donor bead be brought within close enough proximity to produce a chemiluminescent signal. A cross-titration of four different RRE-peptide pairs (PqqD, HcaF, TbtF, and TbiB1) across multiple dilutions revealed a clear binding pattern consistent with RRE-peptide engagement (**Fig. 1b-1e**). Two of the pairs utilize fusion proteins containing only the predicted PqqD-like domain (TbtF and HcaF), suggesting that predicted RRE domains, rather than full-length proteins, can be assessed for binding in this workflow.

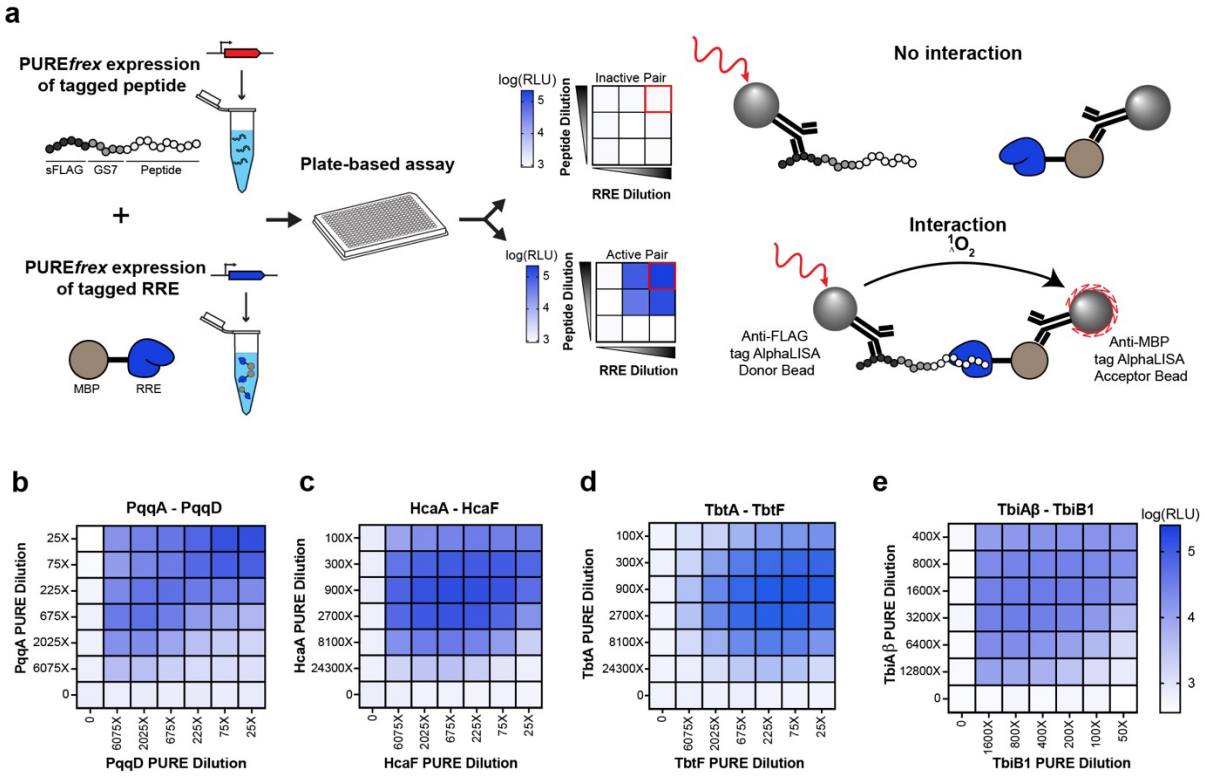


Figure 1. A cell-free, plate-based assay for detecting RRE-peptide interactions. (a) Schematic of the cell-free workflow. sFLAG-tagged peptides and MBP-tagged RREs are expressed in individual PURE*frex* reactions, mixed in a 384 well plate, and incubated to enable binding interactions. Addition of anti-FLAG AlphaLISA donor beads and anti-MBP AlphaLISA acceptor beads enables detection of interactions between the RRE and peptide of interest. PURE*frex* reactions of precursor peptide and RRE for (b) pyrroloquinoline quinone (PQQ), (c) a heterocycloanthracin from *Bacillus* sp. Al Hakam, (d) a GE2270

derivative thiopeptide from *Thermobispora bispora*, and (e) a lasso peptide from *Thermobacculum terrenum* ATCC BAA-798 were cross-titrated across different dilutions and assessed for binding interactions via AlphaLISA. Data are representative of at least three independent experiments.

Cell-free workflow enables mapping of residues important for RRE-peptide binding

We next sought to benchmark our workflow's ability to identify peptide residues important for RRE binding. As a test case, we assayed binding interactions between the HcaF protein and the leader sequence of HcaA (**Supplementary Fig. 2**). We specifically tested the ability of MBP-HcaF to bind to a mutant library composed of HcaA variants with each amino acid individually mutated to alanine. Previous work has found that mutating M1 and F4 of HcaA results in a 4-fold and 10-fold decrease in binding affinity respectively with HcaF[41], a finding that our workflow confirms as evidenced by a decrease in AlphaLISA signal compared to reactions involving HcaF and the wild-type peptide sequence, with a larger decrease observed for the F4A mutation. The results from our workflow also corroborate the finding that mutating any of the three leucine residues in HcaA results in a large decrease in binding affinity with HcaF and that of the six tyrosine residues in HcaA, mutating Y16 results in the largest decrease in binding affinity with HcaF[41]. Importantly, by utilizing simple DNA manipulation techniques, we were able to construct the library of DNA templates, express peptide and protein constructs, and assay for binding activity within hours. This is faster than conventional cloning, transformation, expression, purification, and characterization workflows, which can take weeks to months.

With the ability to identify important peptide residues for binding, we next assessed if our platform could accurately determine residues important for binding within the leader sequence of a previously discovered, but uncharacterized precursor peptide. The N-terminal leader sequence is important because it is the portion of the peptide that is recognized by the RRE and therefore enables the recognition, and subsequent modification, of the peptide by additional tailoring enzymes[8]. We characterized binding of TbtF, a computationally predicted F-protein cyclodehydratase involved in thiopeptide biosynthesis, to the leader sequence of TbtA (**Fig. 2a**)[21]. We specifically tested a mutant library of the TbtA leader sequence containing individual alanine mutations scanning the peptide. We found that mutation of six residues within the TbtA leader sequence, L9, D11, L12, M14, D15, and F17, to an alanine abolish binding by TbtF, as evidenced by a greater than 100-fold decrease in AlphaLISA signal for all alanine variants. A second set of alanine variants, composed of L7A, P13A, M19A, also appear to decrease binding by TbtF; relative AlphaLISA signal for reactions with each of these variants was reduced by approximately 28-fold, 25-fold, and 9-fold respectively compared to signal obtained using the wild-type leader

sequence. Lastly, reactions with a third group of alanine variants, composed of N8A, E18A, and D21A also had slightly lower relative AlphaLISA signal relative to reactions with the wild-type TbtA leader sequence.

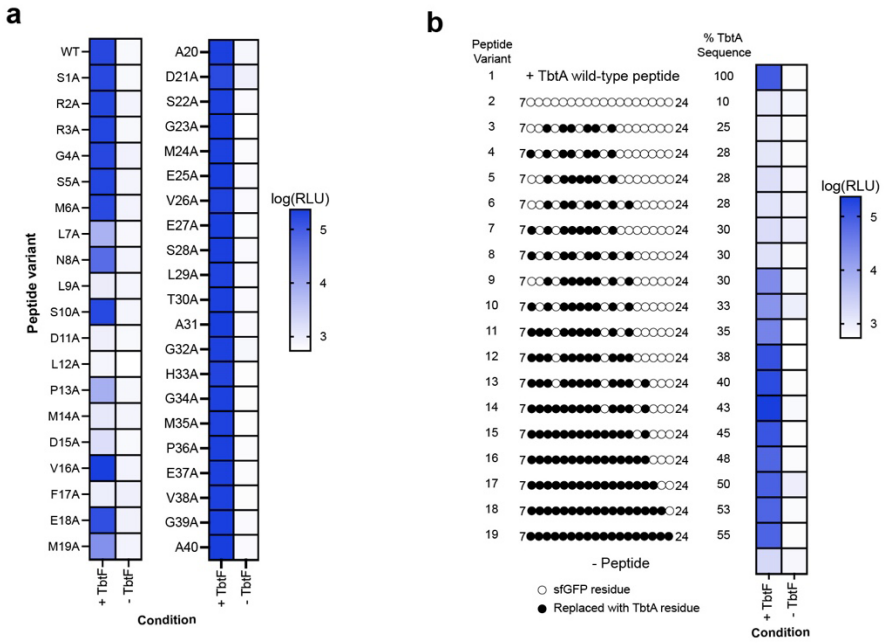


Figure 2. Cell-free workflow identifies peptide residues important for binding by TbtF. (a) An alanine scan library of the leader sequence of TbtA was expressed in individual PURE*frex* reactions and assessed for binding interactions in the presence and absence of TbtF using AlphaLISA. (b) A synthetic peptide library was constructed using the first 40 amino acids of sfGFP. Variants of the sfGFP peptide were then constructed by replacing residues in the peptide identified by the alanine scan as important for binding by TbtF with the corresponding residue in the wild-type TbtA leader sequence. Each peptide variant was expressed in an individual PURE*frex* reaction, and then assessed for binding interactions in the presence and absence of TbtF using AlphaLISA. Peptide variant 2 contains 10% identity to TbtA wild-type peptide due to sharing residues S1, G32, G34, and G39. Sequences for each of the peptide variants assayed in panel b are provided in **Supplementary Table 1**. All data are presented as the mean of n = 3 technical replicates.

We next wanted to use the characterized peptide-binding landscape to inform the design of a synthetic peptide capable of binding to TbtF. As a starting point, we constructed a synthetic peptide sequence the same length as the leader sequence of TbtA that does not bind to TbtF (**Fig. 2b**; peptide variant 2), using the first 40 amino acids of sfGFP with a G23T mutation to ensure all residues in the region of interest differed from the wild-type TbtA leader sequence. We then created peptide variants by replacing residues in the synthetic peptide with residues identified from the alanine scan as important for binding by TbtF, starting with the six residues (L9, D11, L12, M14, D15, and F17) that when mutated to an alanine resulted in the greatest decrease in AlphaLISA signal. We were unable to detect binding interactions between this engineered peptide variant (peptide variant 3) and TbtF. Next, we created peptide variants 4-10

by adding residues (L7, P13, and M19) individually or in combination to peptide variant 3. Adding both P13 and M19 (peptide variant 9) to peptide variant 3 enabled weak binding by TbtF, with ~25% AlphaLISA signal of the wild-type TbtA leader sequence. Further addition of residues resulted in a synthetic peptide (peptide variant 12) that is 38% identical to the leader sequence of TbtA (L7, N8, L9, D11, L12, P13, M14, D15, F17, E18, and M19) and exhibits binding to TbtF (AlphaLISA signal) that is approximately equal to that observed with the wild-type TbtA leader sequence peptide. Interestingly, adding residues D21 and S10 (peptide variant 14) increased the signal further to ~2-fold higher than that observed with the wild-type TbtA leader sequence. These results confirm our assay's ability to identify key residues involved in RRE-peptide binding interactions and highlight the applicability of the workflow as a tool towards the development of new-to-nature RiPPs.

Cell-free workflow as a screening tool for computationally identified RRE-peptide pairs

A challenge in RiPP discovery from computationally predicted BGCs is that RREs may be incorrectly identified[25]. We next showed that our workflow could be used to characterize RRE binding for uncharacterized lasso peptide BGCs computationally predicted via AntiSMASH[17]. Lasso peptides are a class of high interest RiPPs due to their unique lariat structure which imparts the molecule with a wide range of beneficial characteristics, such as heat and protease stability[49,50]. Additionally, previously discovered lasso peptides have displayed a variety of bioactivities, including antimicrobial activity[51], with several lasso peptides having distinct cellular targets, such as against RNA polymerase[52-54], components of the cell wall[55,56], or the ClpC1 unit of the ClpC1P1P2 protease complex[57]. Biosynthetically, lasso peptide BGCs typically encode (i) a precursor peptide, (ii) an RRE and (iii) a protease, or a fusion protein encoding both the RRE and protease, as well as (iv) a cyclase[49]. In all reported lasso peptide BGCs, RREs are important for guiding the protease to the precursor peptide substrate and in some cases is also required for cyclization by the cyclase[58-63]. Previous works have also successfully utilized cell-free systems to study lasso peptide biosynthesis[64].

To begin, we used AntiSMASH[17] to identify a total of 2,574 lasso peptide BGCs from a collection of 39,311 diverse genomes (**Fig. 3a, Supplementary Table 2**). Of these, 1,882 BGCs were predicted to contain a complete collection of essential lasso peptide biosynthetic enzymes (**Supplementary Table 3**). With an eye towards discovering lasso peptides, we compared the identified BGCs to known lasso peptides by constructing a sequence similarity network of the predicted pro-peptide sequences and annotating known sequences within the resulting network.

From the remaining predicted BGCs, 47 were selected for study from 32 unique genera based on their predicted pro-peptide length and whether they would be expected to carry a positive charge at a neutral pH (**Extended Data Set**). Rationale for the latter criteria was that cationic peptides could enrich for peptides with antibiotic activity against gram-negative bacteria, a desirable trait in antibiotic discovery as they could accumulate towards the negatively charged lipopolysaccharide sugars[65,66].

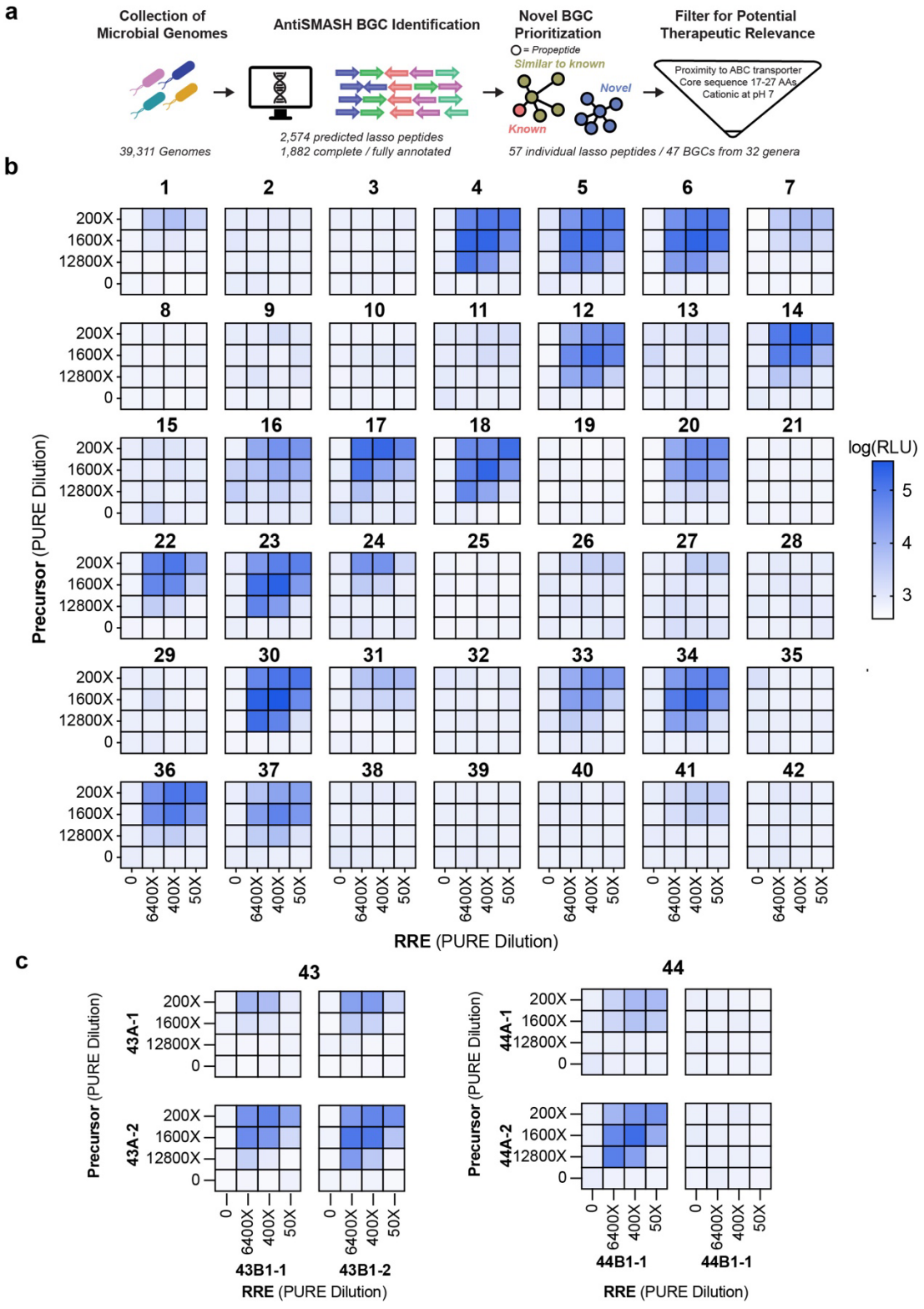


Figure 3. Computational guided screen of lasso peptide RREs. (a) Computational prediction workflow for identifying lasso peptide BGCs. (b) For predicted BGCs with a single RRE and precursor peptide, individual PURE*frex* reactions expressing RREs and respective peptide were cross-titrated and assessed for binding activity via AlphaLISA. (c) For predicted biosynthetic gene clusters with multiple RREs and

precursor peptides, all possible combinations of RRE and precursor peptide were assessed for binding activity via AlphaLISA (select combinations shown, see **Supplementary Figure 4** for additional pairwise combinations). BGCs 4 and 23 are representative of n = 5 independent experiments and the combination of 43A-1 with 43B1-1 is representative of n = 2 independent experiments, with all others having a single replicate.

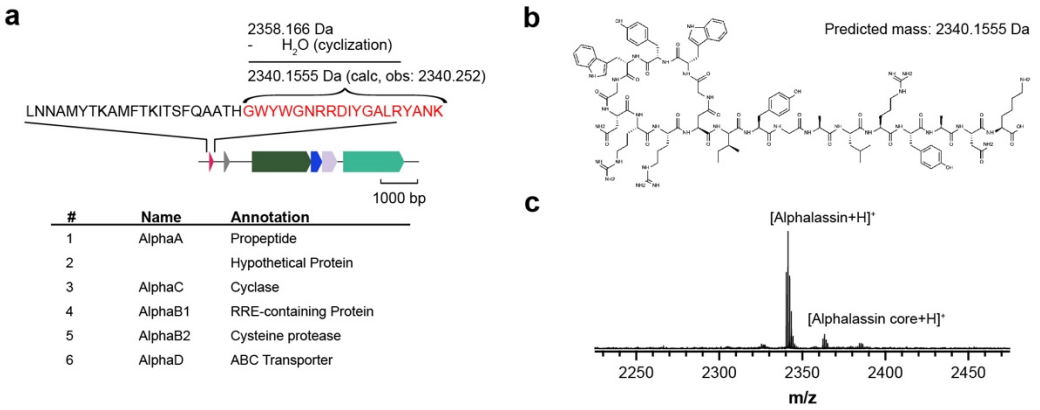
Of the 47 predicted lasso peptide BGCs, 5 were predicted to contain more than one precursor peptide and/or RRE, bringing the total number of predictions to 57 unique precursor peptides and 52 unique RREs. We applied our cell-free workflow to screen all 57 predicted precursor peptides with their associated predicted RREs (**Fig. 3b** and **3c**; **Supplementary Fig. 4**). To account for potential differences in expression levels in the PURE*frex* reactions as well as the fact that RREs reported in literature have a range of binding affinities, we tested each peptide-RRE pair at multiple concentrations. In instances where multiple RREs or precursor peptides were predicted in the same BGC, we screened all pairwise combinations. In total, we screened 72 different RRE-peptide pairs, 42 RRE-peptide pairs from clusters with a single predicted RRE and peptide pair (**Fig. 3b**) and 30 different combinations of RRE and peptides from clusters with multiple predicted genes for each (**Fig. 3c** and **Supplementary Fig. 4**). When assessing for binding interactions, we considered any RRE-peptide pair a "hit" if any of the 9 different RRE-peptide dilution conditions produced AlphaLISA signal at least 2.5-fold higher than the signal obtained from either of the corresponding no RRE or no peptide control reactions.

Our initial screen yielded hits for 23 of the 42 individual RRE-peptide pairs and 22 of the 30 RRE-peptide combinations from larger clusters (**Fig. 3**). A subsequent validation experiment assaying all RRE-peptide pairs at the dilution condition that yielded the highest AlphaLISA signal agreed with 69/70 of the results from the initial screen, with the lone exception being cluster 1 (**Supplementary Fig. 5**).

In vitro production of a lasso peptide from *Rothia*

Using the results from our large-scale RRE screen, we set out to test any clusters identified as "hits" for complete biosynthesis of a mature lasso peptide *in vitro*. To do so, we expressed each precursor peptide in a PURE*frex* reaction and purified each related tailoring enzyme heterologously expressed in *E. coli.* To increase the soluble yield of each of the cyclases, we incorporated an N-terminal MBP-tag and co-expressed folding chaperones during protein production *in vivo*. Small scale (10 μL) reactions were assembled by combining precursor peptides and purified tailoring enzymes and analyzed via matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) MS after overnight incubation at 37 ºC.

We successfully produced a lasso peptide, from BGC 24, which we refer to as Alphalassin (**Fig. 4**). Alphalassin is a 19-residue, class II lasso peptide identified from the genome of an unclassified species of *Rothia,* originally isolated from kefir. Commonly found in the upper respiratory tract or in the flora of the human oropharynx, *Rothia* are gram-positive, anaerobic or facultatively anaerobic bacteria[67]. Biosynthetically, the Alphalassin BGC encodes a five-gene split protease system consisting of the precursor peptide, an RRE, a protease, a cyclase, a predicted ABC transporter, and a fifth protein of unknown function (**Fig. 4a**). The Alphalassin BGC lies directly adjacent on the genome to a predicted novel lanthipeptide BGC.



#	Name	Annotation
1	AlphaA	Propeptide
2		Hypothetical Protein
3	AlphaC	Cyclase
4	AlphaB1	RRE-containing Protein
5	AlphaB2	Cysteine protease
6	AlphaD	ABC Transporter

Figure 4. *In vitro* production of Alphalassin, a class II lasso peptide. (a) Schematic of Alphalassin biosynthetic gene cluster and precursor peptide sequence. (b) Predicted structure of Alphalassin. (c) MALDI-TOF-MS spectra demonstrating *in vitro* production of Alphalassin. Data is representative of three independent experiments.

After overnight incubation of the Alphalassin precursor peptide with purified samples of the predicted RRE, protease, and cyclase, we observe the emergence of an m/z peak at 2341.260 which agrees with the predicted structure of Alphalassin (predicted $[M+H]^+$ m/z: 2341.156) and does not appear in reactions in which individual components are removed from the reaction (**Fig. 4b** and **4c**; **Supplementary Fig. 7**). In instances in which both the RRE and protease, but not cyclase, are incubated with the precursor peptide, we observe formation of a mass with an m/z of 2620.229, which corresponds to the uncyclized core peptide (predicted $[M+H]^+$ m/z: 2359.166) (**Supplementary Fig. 7**). A subsequent experiment confirmed that the production of Alphalassin is time dependent (**Supplementary Fig. 8**).

Treatment with carboxypeptidase Y is often used to distinguish threaded lasso peptides from branched, cyclic peptides, due to the macrolactam ring's ability to shield the C-terminus of threaded lasso peptides from protease degradation[68]. When we treat reactions producing only the

uncyclized core peptide with carboxypeptidase Y, we observe complete degradation after overnight incubation (**Supplementary Fig. 9**). In contrast, we do not observe degradation of the cyclized product after overnight treatment with carboxypeptidase Y, suggesting the threaded structure of Alphalassin (**Supplementary Fig. 9**).

Discussion

Here we developed an integrated workflow for RRE screening to enhance RiPP discovery pipelines by combining methods for cell-free DNA assembly and amplification, cell-free protein synthesis, and binding characterization via AlphaLISA. We show that the platform is generalizable, fast (steps are carried out in hours), and readily scalable to 96- or 384-well plates without the need for time intensive protein purification or cell-based cloning techniques. Moreover, the platform is designed with automation in mind, with each step consisting of simple liquid handling and temperature incubation steps. We showed the utility of the platform by identifying residues important for binding by an RRE and coupled our cell-free workflow with genome mining and computational analysis to identify and produce a previously undiscovered lasso peptide, which we name Alphalassin.

As a screening tool, there are multiple advantages to incorporating our cell-free workflow prior to testing computationally predicted RiPP BGCs. First, our workflow serves as a filtering tool to down select clusters with active RiPP peptide recognition. We note that many predicted RREs in our study did not exhibit binding interactions to the respective predicted precursor peptide (50% of predicted RREs did not bind to an associated precursor peptide). We hypothesize that this could be due to incorrect annotation of RREs and precursor peptides in our computational workflow and prioritizing the investigation of BGCs with definable and quantifiable biochemical interactions increases the success rate of complete *in vitro* biosynthesis and characterization of the mature RiPP. Second, the throughput of our workflow enables testing each pairwise combination of RRE and precursor peptide for BGCs with multiple predictions for each component. Not only do we observe potential differences in binding affinity within a given BGC (e.g., 43-1A compared to 43-2A with both RREs), but we also observe several non-functional RREs in the more complex clusters. For example, while 44B1-1 is capable of binding to both predicted precursor peptides, 44B1-2 is unable to bind to either peptide. Understanding which RRE-peptide pairs are functional can narrow the number of different enzyme combinations that need to be tested when assembling complete BGCs.

Biosynthesis of lasso peptides requires a concerted sequence of events, including both proteolytic and cyclization activity, of which recognition of the precursor peptide by the RRE is only the first step[49]. In some instances where the computationally predicted RREs are functional but we observe no maturation to the lasso peptide structure, the associated predicted proteases or cyclases may be incorrectly annotated. Alternatively, for some lasso peptides, additional chaperones not encoded within the BGC may be required for maturation into the final product, as has been demonstrated for other natural products[69,70]. Future studies should consider methods for identifying potential auxiliary genes required for lasso peptide biosynthesis and incorporating those additional genes into the *in vitro* reconstitution of the BGCs. By acting as an initial screen, our RRE focused workflow reduced the number of enzymes required for *in vivo* production and purification by approximately 40% by prioritizing BGCs with demonstrated functional RREs.

RiPP tailoring enzymes represent a potentially modular approach to creating new-to-nature molecules. Recent work has, for example, created novel RiPP products by engineering peptide substrates to contain leader sequences recognized by tailoring enzymes from multiple classes of RiPPs[71]. In doing so, a peptide substrate was modified with RRE-dependent tailoring enzymes from two different BGCs. Creating more complex systems with even greater numbers of RRE-dependent modifications will require an understanding of appropriate design rules for enabling recognition of the precursor peptide by the desired tailoring enzymes. Using the information gained by mutational scanning, we were able to systematically produce a synthetic peptide with less than 40% identity to the wild-type peptide that exhibits AlphaLISA binding signal on par with the wild-type peptide. Understanding the minimal set of amino acid residues required for recognition will be important for engineering increasingly complex molecules. Our workflow provides a method for understanding and prototyping these requirements.

From initial biodiscovery to in-depth characterization, our work provides an important tool for studying RiPP biosynthesis. Looking forward, we anticipate that the cell-free AlphaLISA workflow will accelerate our understanding of RiPP enzymology and enable engineering of complex new-to-nature molecules.

Materials and Methods

DNA design and preparation

For the initial screen of known RRE's, gene constructs were ordered from Twist Biosciences (synthesized into pJL1 backbone between NdeI and SalI restriction sites). Briefly, sequences were retrieved from literature or Uniprot and codon optimized using the IDT Codon Optimization Tool. For full length RRE constructs, a codon optimized sequence for a Twin-Strep tag and PAS11 linker were added to the N-terminus of the nucleotide sequence. MBP-fusion RRE constructs were constructed by replacing either the C-terminus (for proteins in which the RRE domain was predicted to occur in the N-terminus) or N-terminus (for proteins in which the RRE domain was predicted to occur in the C-terminus) portion of the sequence with codon optimized sequences for MBP and a GS7 linker. For precursor peptide sequences, sequences encoding either the full-length precursor or leader sequence were fused to an N-terminal sFLAG tag and GS7 linker.

For all peptide sequences used in AlphaLISA based assays, an N-terminal sFLAG tag and GS7 linker were incorporated into the design. For sequences utilized in the AlphaLISA alanine scan workflow, each alanine variant peptide was constructed by replacing the corresponding wild-type codon with "GCC". To construct synthetic sfGFP peptides, the first 40 amino acids of sfGFP (with a G23T mutation) was first codon optimized. Each variant was then constructed by replacing the appropriate wild-type codon with the codon corresponding to the desired residue change. All peptide sequences were ordered as eBlocks with overhang to a linearized pJL1 backbone for use in Gibson Assembly reactions.

For all computationally predicted lasso peptide proteases and cyclases, the predicted gene sequences were codon optimized using the IDT Codon Optimization Tool. At the N-terminus of each sequence, maltose binding protein (MBP) and a short linker were incorporated to enable soluble expression and detection via AlphaLISA based assays. All genes were synthesized by Twist Biosciences either in pJL1 (for expression in PURE*frex*) or in a modified pET vector (for *in vivo* expression). The corresponding (untagged) precursor sequences were also synthesized by Twist Biosciences in pJL1 for use in assembling complete lasso peptide BGCs.

DNA templates for expression in PURE*frex* were prepared either in plasmid form using ZymoPURE II Plasmid Midiprep Kit (Zymo Research) or as linear expression templates (LETs). For LETs, eBlocks were inserted into pJL1 using Gibson Assembly with a linearized pJL1

backbone. Following Gibson Assembly (GA), each reaction was then diluted 10x in nuclease free water. 1 µL of diluted GA reaction was then used in a 50 µL PCR reaction using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs).

Linearized pJL1 backbone:

gagcatcaaatgaaactgcaatttattcatatcaggattatcaataccatatttttgaaaaagccgtttctgtaatgaaggagaaaactca
ccgaggcagttccataggatggcaagatcctggtatcggtctgcgattccgactcgtccaacatcaatacaacctattaatttcccctcgt
caaaaataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggtgagaatggcaaaagcttatgcatttctttccagact
tgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttattcattcgtgattgcgcctgagcgagacg
aaatacgcgatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgccagcgcatcaacaat
attttcacctgaatcaggatattcttctaatacctggaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatcaggagtac
ggataaaatgcttgatggtcggaagaggcataaattccgtcagccagtttagtctgaccatctcatctgtaacatcattggcaacgctac
ctttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaatcgatagattgtcgcacctgattgcccgacattatcgcgagc
ccatttatacccatataaatcagcatccatgttggaatttaatcgcggcttcgagcaagacgtttcccgttgaatatggctcataacacccc
ttgtattactgtttatgtaagcagacagttttattgttcatgatgatatatttttatcttgtgcaatgtaacatcagagattttgagacacaacgtg
agatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgc
cggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagtt
aggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtc
gtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtgcacacagccca
gcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaagg
cggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatctttata
gtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaac
gcgatcccgcgaaattaatacgactcactatagggagaccacaacggtttc

Forward primer for LET PCR: ctgagatacctacagcgtgagc
Reverse primer for LET PCR: cgtcactcatggtgatttctcacttg

FluoroTect™ gel

PURE*frex* 2.1 (Gene Frontier) reactions were assembled according to manufacturer instructions, using 1 µL of unpurified template LET and 0.5 µL of FluoroTect™ (Promega) per 10 µL reaction. Following incubation at 37 °C for 6 hours, samples were centrifuged at 12,000 xg for 10 min at 4 °C. 3 µL of supernatant was then mixed with 1 µL of 40 µg/mL RNase A and incubated at 37 °C for 10 minutes. Following incubation, 1 µL of 1M DTT, 2.5 µL of 4X Protein Sample Loading Buffer for Western Blots (Li-COR Biosciences), and 2.5 µL of water were added to each sample and the

samples were then incubated at 70 °C for 10 minutes. Samples were then loaded on a NuPAGE 4-12% Bis-Tris Protein Gel and run for 40 minutes at 200 V in MES Running Buffer. For comparison, a lane was loaded with BenchMark fluorescent protein standard (Thermo Fisher Scientific). The resulting gel was then imaged using both the 600 and 700 fluorescent channel on a LICOR Odyssey Fc (Li-COR Biosciences).

AlphaLISA reactions

PURE*frex* 2.1 (Gene Frontier) reactions were assembled according to manufacturer instructions. Briefly, 1 μL of the unpurified LET reaction – encoding for the precursor peptide or RRE - was added as a template per 10 μL PURE*frex* reaction. Reactions were then incubated at 37 °C for 5 hours. After incubation, these samples were then diluted in a buffer consisting of 50 mM HEPES pH 7.4, 150 mM NaCl, 1 mg/mL BSA, and 0.015% v/v Triton X-100. Following dilution, an Echo 525 acoustic liquid handler was used to dispense 0.5 μL of diluted RRE, 0.5 μL of diluted peptide, and 0.5 μL of blank buffer from a 384-well polypropylene 2.0 Plus Source microplate (Labcyte) using the 384PP_Plus_GPSA fluid type into a ProxiPlate-384 Plus, White 384-shallow well destination microplate. The plate was then sealed and equilibrated at room temperature for one hour. Next, anti-FLAG Alpha Donor beads (Perkin Elmer) were used to immobilize the sFLAG tagged peptides and anti-Maltose-Binding (MBP) AlphaLISA acceptor beads were used to immobilize the MBP-tagged RREs. 0.5 μL of acceptor and donor beads diluted in buffer were added to each reaction to a final concentration of 0.08 mg/mL and 0.02 mg/mL donor and acceptor beads respectively. Reactions were then equilibrated an additional hour at room temperature in the dark. For analysis, reactions were incubated for 10 minutes in a Tecan Infinite M1000 Pro plate reader at room temperature and then chemiluminescence signal was read using the AlphaLISA filter with an excitation time of 100 ms, an integration time of 300 ms, and a settle time of 20 ms. Results were visualized using Prism version 9.5.1 (GraphPad).

Computational prediction of lasso peptide BGCs

A diverse collection of 39,311 publicly genomes available (2020 April) spanning soil bacteria, metagenomes and extremophiles were analyzed using AntiSMASH 5.1.2 identifying 315,876 biosynthetic gene clusters (**Supplementary Table 2**). A total of 2,574 lasso peptide clusters were identified, and from this set, 1,882 BGCs contained a complete collection of essential biosynthetic enzymes (**Supplementary Table 3**). Further prioritizing these clusters, a sequence similarity network[72,73] of the identified propeptide genes with a collection of known lasso peptide sequences was created to assess the novelty of each cluster. Subsequent filtering of the remaining novel

BGCs included selecting BGCs based on a propeptide length of 17-27 amino acids and whether the mature lasso peptide is predicted to carry a positive charge at a neutral pH. Calculation of the predicted isoelectric point of the predicted core peptides used Thermo Fisher Scientific's peptide analysis tool (https://www.thermofisher.com/us/en/home/life-science/protein-biology/peptides-proteins/custom-peptide-synthesis-services/peptide-analyzing-tool.html). This narrowed the selection to 202 BGCs, of which 47 were chosen. A total of 210 genes were synthesized by Twist Bioscience. All amino acid sequences and metadata for the 47 selected BGCs are provided in the **Extended Data Set**.

In vivo expression and purification of lasso peptide tailoring enzymes

For computationally predicted MBP-RREs and MBP-proteases, constructs of the target protein in pET.BCS.RBSU.NS backbone were transformed into BL21 Star (DE3) cells, plated on LB agar plates containing 100 μg/mL carbenicillin, and incubated at 37 °C. Single colonies were cultured in 50 mL of LB containing 100 μg/mL carbenicillin at 37 °C and 250 RPM. After overnight incubation, 20 mL of the overnight culture were used to inoculate 1L of LB supplemented with 2 g/L glucose and 100 μg/mL carbenicillin. Cells were grown at 37 °C and 250 RPM and induced for protein production at OD_{600} 0.6-0.8 with 500 μL of 1M IPTG. Four hours post induction, cells were harvested via centrifugation at 5,000 xg for 10 minutes at 4 °C and flash frozen in liquid nitrogen.

After thawing on ice, cell pellets were resuspended in lysis buffer composed of 50 mM Tris-HCl pH 7.4, 500 mM NaCl, 2.5 % (v/v) glycerol, and 0.1% Triton X-100. For cell pellets used to overexpress RREs and cyclases, the lysis buffer also contained 6 mM PMSF, 100 μM Leupeptin, and 100 μM E64. Cell suspensions were then supplemented with 1 mg/mL lysozyme and lysed via sonication using a Qsonica sonicator at 50% amplitude for 2 minutes with 10 seconds on 10 second off cycles. Following sonication, insoluble debris were removed via centrifugation at 14,000 xg for 30 minutes at 4 °C. Per 1L of cell culture, 5 mL of amylose resin was equilibrated with 5 to 10 column volumes of wash buffer (50 mM Tris HCl, 500 mM NaCl, 2.5 % (v/v) glycerol, pH 7.4) in a 50 mL conical tube and mixed via inversion. Resin was separated from wash buffer by spinning at 2,000 xg for 2 min at 4 °C and the supernatant was then poured off. Equilibration was repeated for a total of 4 times with fresh equilibration buffer. Following the last equilibration, the cleared cell lysis supernatant was added to the resin and incubated for 2 hours at 4 °C with constant agitation on a shake table. Following incubation on the resin, the resin was washed once with 5 column volumes of lysis buffer followed by 5 column volumes of wash buffer four times.

For the last wash, the resuspended resin was loaded in a 25 mL gravity flow column and drained via gravity flow. For elution, 15 mL of elution buffer (50 mM Tris HCl, 300 mM NaCl, 10 mM maltose, 2.5% (v/v) glycerol, pH 7.4) was added to the gravity flow column and collected. Samples were then buffer exchanged into storage buffer (50 mM HEPES, 300 mM NaCl, 0.5 mM TCEP, 2.5% (v/v) glycerol, pH 7.5) using amicon spin filters (50 kDa MWCO) by spinning at 4,500 xg for 10-15 minutes. Samples were then aliquoted, flash frozen, and stored at -80 °C until use. Total protein concentration of each purified sample was determined using a Bradford assay (Biorad). Percent purity of each sample was determined by running diluted aliquots of each purified protein on a 4-12% Bis-Tris gel and staining with Optiblot Blue (Abcam). After destaining, each gel was imaged using the 700 fluorescent channel on a LICOR Odyssey Fc (Li-COR Biosciences, USA) and percent purity was determined via densitometry using Licor Image Studio Lite. Final concentrations of each protein were then calculated by multiplying the total protein content by the percent purity.

Computationally identified cyclases were expressed and purified according to the process outlined above for computationally identified RREs except for transforming into BL21 Star (DE3) cells already transformed with pG-KJE8. LB agar and media for cell growth were supplemented with 20 μg/mL chloramphenicol in addition to 100 μg/mL carbenicillin. At inoculation, LB was supplemented with 2 g/L glucose, 100 μg/mL carbenicillin, 20 μg/mL chloramphenicol, and 2 ng/mL anhydrotetracycline per 1 L of media for induction of folding chaperones.

In vitro enzymatic assembly of lasso peptide BGCs

PURE*frex* 2.1 (Gene Frontier) reactions to express the precursor peptide were assembled according to manufacturer instructions using 1 μL of 200 ng/μL plasmid (pJL1 backbone encoding precursor peptide of interest) per 10 μL reaction and incubated at 37 °C for at least five hours. Purified proteins were buffer exchanged using Zeba Micro Spin Desalting Columns (7K MWCO) into synthetase buffer (50 mM Tris-HCl pH 7.5, 125 mM NaCl, 20 mM $MgCl_2$). 10 μL reactions were then assembled using 5 μL of PURE*frex* reaction, and the appropriate volume of each individual purified enzyme or buffer such that both the RRE and protease were at a final concentration of 10 μM and the cyclase was at a final concentration of 1 μM. Reactions were supplemented to a final concentration of 10 mM DTT and 5 mM ATP and incubated at 37 °C for varying lengths of time. For analysis, samples were desalted using Pierce C18 spin tips (10 μL bed), spotted on a MALDI target plate using 50% saturated CHCA matrix in 80% ACN with 0.1%

TFA, and analyzed using a Bruker RapiFlex MALDI-TOF mass spectrometer in reflector positive mode at Northwestern University's Integrated Molecular Structure Education and Research Center (IMSERC).

Carboxypeptidase treatment of lasso peptides

Assembled reactions (20 μL scale) were desalted using Pierce C18 spin column and eluted into 20 μL of acetonitrile. After solvent removal under vacuum, reactions were resuspended in a solution containing carboxypeptidase Y at 50 ng/μL in 1X PBS (10 μL) and incubated at room temperature overnight. The mixtures were evaporated to dryness and resuspended in 3 μL saturated α-Cyano-4-hydroxycinnamic acid (CHCA) matrix solution in TFA (trifluoroacetic acid). Samples were then spotted on a matrix assisted laser desorption/ionization (MALDI) plate and analyzed using a Bruker RapiFlex MALDI-TOF mass spectrometer in reflector positive mode at Northwestern University's Integrated Molecular Structure Education and Research Center (IMSERC).

Acknowledgements

The authors would like to thank Rui Gan, Jonathan Bogart, and Thuy Aziz for helpful discussions. This work was supported by the National Institutes of Health (NIH) 1U19AI142780-01. D.A.W. acknowledges support from the National Science Foundation Graduate Research Fellowship under grant no. DGE-1842165. M.D. acknowledges support from the Canadian Institutes of Health Research Postdoctoral Fellowship under grant no. MFE-176575. This work made use of the IMSERC MS facility at Northwestern University, which has received support from the Soft and Hybrid Nanotechnology Experimental (SHyNE) Resource (NSF ECCS-2025633), the State of Illinois, and the International Institute for Nanotechnology (IIN).

Author Contributions

D.A.W. designed research, performed experiments, performed MALDI-MS on reactions, analyzed data, and wrote the paper. M.D.C. designed research, performed experiments, performed MALDI-MS on reactions, analyzed data, and edited the paper. M.D. computationally identified all lasso peptide BGCs and wrote the paper. D.V.P. performed experiments. R.F. performed experiments. R.N. supervised research and edited the paper. E.P.B. supervised research and edited the paper. A.S.K. supervised research, analyzed data, and edited the paper. M.C.J. designed and directed research, analyzed data, and wrote the paper.

Conflict of Interest

M.C.J. has a financial interest in Resilience, Gauntlet Bio, Pearl Bio, Stemloop Inc., and Synolo Therapeutics. M.C.J.'s interests are reviewed and managed by Northwestern University and Stanford University in accordance with their competing interest policies. All other authors declare no competing interests.

References

1 Ayikpoe, R. S. *et al.* A scalable platform to discover antimicrobials of ribosomal origin. *Nature Communications* **13**, 6135 (2022). https://doi.org/10.1038/s41467-022-33890-w

2 Hudson, G. A. & Mitchell, D. A. RiPP antibiotics: biosynthesis and engineering potential. *Current Opinion in Microbiology* **45**, 61-69 (2018). https://doi.org/https://doi.org/10.1016/j.mib.2018.02.010

3 Fu, Y., Jaarsma, A. H. & Kuipers, O. P. Antiviral activities and applications of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Cellular and Molecular Life Sciences* **78**, 3921-3940 (2021). https://doi.org/10.1007/s00018-021-03759-0

4 Shin, J. M. *et al.* Biomedical applications of nisin. *J Appl Microbiol* **120**, 1449-1465 (2016). https://doi.org/10.1111/jam.13033

5 Chiumento, S. *et al.* Ruminococcin C, a promising antibiotic produced by a human gut symbiont. *Science Advances* **5**, eaaw9969 (2019). https://doi.org/doi:10.1126/sciadv.aaw9969

6 Dahlem, C. *et al.* Thioholgamide A, a New Anti-Proliferative Anti-Tumor Agent, Modulates Macrophage Polarization and Metabolism. *Cancers* **12** (2020).

7 Takase, S. *et al.* Mechanism of Action of Prethioviridamide, an Anticancer Ribosomally Synthesized and Post-Translationally Modified Peptide with a Polythioamide Structure. *ACS Chemical Biology* **14**, 1819-1828 (2019). https://doi.org/10.1021/acschembio.9b00410

8 Arnison, P. G. *et al.* Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* **30**, 108-160 (2013). https://doi.org/10.1039/c2np20085f

9 Ongpipattanakul, C. & Nair, S. K. Molecular Basis for Autocatalytic Backbone N-Methylation in RiPP Natural Product Biosynthesis. *ACS Chemical Biology* **13**, 2989-2999 (2018). https://doi.org/10.1021/acschembio.8b00668

10 Miller, F. S. *et al.* Conformational rearrangements enable iterative backbone N-methylation in RiPP biosynthesis. *Nature Communications* **12**, 5355 (2021). https://doi.org/10.1038/s41467-021-25575-7

11 Ortiz-López, F. J. *et al.* Cacaoidin, First Member of the New Lanthidin RiPP Family. *Angewandte Chemie International Edition* **59**, 12654-12658 (2020). https://doi.org/https://doi.org/10.1002/anie.202005187

12 Iorio, M. *et al.* A Glycosylated, Labionin-Containing Lanthipeptide with Marked Antinociceptive Activity. *ACS Chemical Biology* **9**, 398-404 (2014). https://doi.org/10.1021/cb400692w

13 Wang, H. *et al.* The glycosyltransferase involved in thurandacin biosynthesis catalyzes both O- and S-glycosylation. *J Am Chem Soc* **136**, 84-87 (2014). https://doi.org/10.1021/ja411159k

14 Schmidt, E. W. *et al.* Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. *Proceedings of the National Academy of Sciences* **102**, 7315-7320 (2005). https://doi.org/doi:10.1073/pnas.0501424102

15 Philmus, B., Christiansen, G., Yoshida, W. Y. & Hemscheidt, T. K. Post-translational Modification in Microviridin Biosynthesis. *ChemBioChem* **9**, 3066-3073 (2008). https://doi.org/https://doi.org/10.1002/cbic.200800560

16 Claesen, J. & Bibb, M. Genome mining and genetic analysis of cypemycin biosynthesis reveal an unusual class of posttranslationally modified peptides. *Proceedings of the National Academy of Sciences* **107**, 16297-16302 (2010). https://doi.org/doi:10.1073/pnas.1008608107

17 Blin, K. *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**, W81-W87 (2019). https://doi.org/10.1093/nar/gkz310

18 Skinnider, M. A., Merwin, N. J., Johnston, C. W. & Magarvey, N. A. PRISM 3: expanded prediction of natural product chemical structures from microbial genomes. *Nucleic Acids Res* **45**, W49-w54 (2017). https://doi.org/10.1093/nar/gkx320

19 Tietz, J. I. *et al.* A new genome-mining tool redefines the lasso peptide biosynthetic landscape. *Nature chemical biology* **13**, 470-478 (2017). https://doi.org/10.1038/nchembio.2319

20 Agrawal, P., Khater, S., Gupta, M., Sain, N. & Mohanty, D. RiPPMiner: a bioinformatics resource for deciphering chemical structures of RiPPs based on prediction of cleavage and cross-links. *Nucleic Acids Res* **45**, W80-w88 (2017). https://doi.org/10.1093/nar/gkx408

21 Burkhart, B. J., Hudson, G. A., Dunbar, K. L. & Mitchell, D. A. A prevalent peptide-binding domain guides ribosomal natural product biosynthesis. *Nature Chemical Biology* **11**, 564-570 (2015). https://doi.org/10.1038/nchembio.1856

22 Latham, J. A., Iavarone, A. T., Barr, I., Juthani, P. V. & Klinman, J. P. PqqD is a novel peptide chaperone that forms a ternary complex with the radical S-adenosylmethionine protein PqqE in the pyrroloquinoline quinone biosynthetic pathway. *J Biol Chem* **290**, 12908-12918 (2015). https://doi.org/10.1074/jbc.M115.646521

23 Kretsch, A. M. *et al.* Peptidase Activation by a Leader Peptide-Bound RiPP Recognition Element. *Biochemistry* **62**, 956-967 (2023). https://doi.org/10.1021/acs.biochem.2c00700

24 Dunbar, K. L., Tietz, J. I., Cox, C. L., Burkhart, B. J. & Mitchell, D. A. Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *Journal of the American Chemical Society* **137**, 7672-7677 (2015). https://doi.org/10.1021/jacs.5b04682

25 Shelton, K. E. & Mitchell, D. A. Bioinformatic prediction and experimental validation of RiPP recognition elements. *Methods Enzymol* **679**, 191-233 (2023). https://doi.org/10.1016/bs.mie.2022.08.050

26 Chekan, J. R., Ongpipattanakul, C. & Nair, S. K. Steric complementarity directs sequence promiscuous leader binding in RiPP biosynthesis. *Proceedings of the National Academy of Sciences* **116**, 24049-24055 (2019). https://doi.org/doi:10.1073/pnas.1908364116

27 Silverman, A. D., Karim, A. S. & Jewett, M. C. Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics* **21**, 151-170 (2020). https://doi.org/10.1038/s41576-019-0186-3

28 Carlson, E. D., Gan, R., Hodgman, C. E. & Jewett, M. C. Cell-free protein synthesis: applications come of age. *Biotechnol Adv* **30**, 1185-1194 (2012). https://doi.org/10.1016/j.biotechadv.2011.09.016

29 Garenne, D. *et al.* Cell-free gene expression. *Nature Reviews Methods Primers* **1**, 49 (2021). https://doi.org/10.1038/s43586-021-00046-x

30 Gregorio, N. E., Levine, M. Z. & Oza, J. P. A User's Guide to Cell-Free Protein Synthesis. *Methods Protoc* **2** (2019). https://doi.org/10.3390/mps2010024

31 Kightlinger, W. *et al.* A cell-free biosynthesis platform for modular construction of protein glycosylation pathways. *Nature Communications* **10**, 5404 (2019). https://doi.org/10.1038/s41467-019-12024-9

32 Beaudet, L. *et al.* AlphaLISA immunoassays: the no-wash alternative to ELISAs for research and drug discovery. *Nature Methods* **5**, an8-an9 (2008). https://doi.org/10.1038/nmeth.f.230

33 Shimizu, Y. *et al.* Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751-755 (2001). https://doi.org/10.1038/90802

34 Klinman, J. P. & Bonnot, F. Intrigues and intricacies of the biosynthetic pathways for the enzymatic quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ. *Chem Rev* **114**, 4343-4365 (2014). https://doi.org/10.1021/cr400475g

35 Jin, M., Liu, L., Wright, S. A., Beer, S. V. & Clardy, J. Structural and functional analysis of pantocin A: an antibiotic from Pantoea agglomerans discovered by heterologous expression of cloned genes. *Angew Chem Int Ed Engl* **42**, 2898-2901 (2003). https://doi.org/10.1002/anie.200351053

36 Fleming, S. R. *et al.* Exploring the Post-translational Enzymology of PaaA by mRNA Display. *Journal of the American Chemical Society* **142**, 5024-5028 (2020). https://doi.org/10.1021/jacs.0c01576

37 Davis, K. M. *et al.* Structures of the peptide-modifying radical SAM enzyme SuiB elucidate the basis of substrate recognition. *Proceedings of the National Academy of Sciences* **114**, 10420-10425 (2017). https://doi.org/doi:10.1073/pnas.1703663114

38 Mavaro, A. *et al.* Substrate Recognition and Specificity of the NisB Protein, the Lantibiotic Dehydratase Involved in Nisin Biosynthesis*. *Journal of Biological Chemistry* **286**, 30552-30560 (2011). https://doi.org/https://doi.org/10.1074/jbc.M111.263210

39 Melby, J. O., Dunbar, K. L., Trinh, N. Q. & Mitchell, D. A. Selectivity, Directionality, and Promiscuity in Peptide Processing from a Bacillus sp. Al Hakam Cyclodehydratase. *Journal of the American Chemical Society* **134**, 5309-5316 (2012). https://doi.org/10.1021/ja211675n

40 Molohon, K. J. *et al.* Structure Determination and Interception of Biosynthetic Intermediates for the Plantazolicin Class of Highly Discriminating Antibiotics. *ACS Chemical Biology* **6**, 1307-1313 (2011). https://doi.org/10.1021/cb200339d

41 Dunbar, K. L., Tietz, J. I., Cox, C. L., Burkhart, B. J. & Mitchell, D. A. Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. *J Am Chem Soc* **137**, 7672-7677 (2015). https://doi.org/10.1021/jacs.5b04682

42 Li, Y.-M., Milne, J. C., Madison, L. L., Kolter, R. & Walsh, C. T. From Peptide Precursors to Oxazole and Thiazole-Containing Peptide Antibiotics: Microcin B17 Synthase. *Science* **274**, 1188-1193 (1996). https://doi.org/doi:10.1126/science.274.5290.1188

43 Khaliullin, B. *et al.* Mycofactocin biosynthesis: modification of the peptide MftA by the radical S-adenosylmethionine protein MftC. *FEBS Letters* **590**, 2538-2548 (2016). https://doi.org/https://doi.org/10.1002/1873-3468.12249

44 Morinaka, B. I. *et al.* Natural noncanonical protein splicing yields products with diverse β-amino acid residues. *Science* **359**, 779-782 (2018). https://doi.org/doi:10.1126/science.aao0157

45 Precord, T. W., Mahanta, N. & Mitchell, D. A. Reconstitution and Substrate Specificity of the Thioether-Forming Radical S-Adenosylmethionine Enzyme in Freyrasin Biosynthesis. *ACS Chemical Biology* **14**, 1981-1989 (2019). https://doi.org/10.1021/acschembio.9b00457

46 Hunt, A. C. *et al.* Multivalent designed proteins neutralize SARS-CoV-2 variants of concern and confer protection against infection in mice. *Science Translational Medicine* **14**, eabn1252 (2022). https://doi.org/doi:10.1126/scitranslmed.abn1252

47 Hunt, A. C. *et al.* A rapid cell-free expression and screening platform for antibody discovery. *Nature Communications* **14**, 3897 (2023). https://doi.org/10.1038/s41467-023-38965-w

48 DeWinter, M. A. *et al.* Point-of-Care Peptide Hormone Production Enabled by Cell-Free Protein Synthesis. *ACS Synthetic Biology* **12**, 1216-1226 (2023). https://doi.org/10.1021/acssynbio.2c00680

49 Hegemann, J. D., Zimmermann, M., Xie, X. & Marahiel, M. A. Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *Accounts of Chemical Research* **48**, 1909-1919 (2015). https://doi.org/10.1021/acs.accounts.5b00156

50 Cheng, C. & Hua, Z.-C. Lasso Peptides: Heterologous Production and Potential Medical Application. *Frontiers in Bioengineering and Biotechnology* **8** (2020). https://doi.org/10.3389/fbioe.2020.571165

51 Tan, S., Moore, G. & Nodwell, J. Put a Bow on It: Knotted Antibiotics Take Center Stage. *Antibiotics* **8**, 117 (2019).

52 Kuznedelov, K. *et al.* The antibacterial threaded-lasso peptide capistruin inhibits bacterial RNA polymerase. *J Mol Biol* **412**, 842-848 (2011). https://doi.org/10.1016/j.jmb.2011.02.060

53 Metelev, M. *et al.* Acinetodin and Klebsidin, RNA Polymerase Targeting Lasso Peptides Produced by Human Isolates of Acinetobacter gyllenbergii and Klebsiella pneumoniae. *ACS Chemical Biology* **12**, 814-824 (2017). https://doi.org/10.1021/acschembio.6b01154

54 Cheung-Lee, W. L., Parry, M. E., Jaramillo Cartagena, A., Darst, S. A. & Link, A. J. Discovery and structure of the antimicrobial lasso peptide citrocin. *J Biol Chem* **294**, 6822-6830 (2019). https://doi.org/10.1074/jbc.RA118.006494

55 Tan, S., Ludwig, K. C., Müller, A., Schneider, T. & Nodwell, J. R. The Lasso Peptide Siamycin-I Targets Lipid II at the Gram-Positive Cell Surface. *ACS Chemical Biology* **14**, 966-974 (2019). https://doi.org/10.1021/acschembio.9b00157

56 Metelev, M. *et al.* Structure, bioactivity, and resistance mechanism of streptomonomicin, an unusual lasso Peptide from an understudied halophilic actinomycete. *Chem Biol* **22**, 241-250 (2015). https://doi.org/10.1016/j.chembiol.2014.11.017

57 Gavrish, E. *et al.* Lassomycin, a ribosomally synthesized cyclic peptide, kills mycobacterium tuberculosis by targeting the ATP-dependent protease ClpC1P1P2. *Chem Biol* **21**, 509-518 (2014). https://doi.org/10.1016/j.chembiol.2014.01.014

58 Zhu, S. *et al.* The B1 Protein Guides the Biosynthesis of a Lasso Peptide. *Scientific Reports* **6**, 35604 (2016). https://doi.org/10.1038/srep35604

59 Koos, J. D. & Link, A. J. Heterologous and in Vitro Reconstitution of Fuscanodin, a Lasso Peptide from Thermobifida fusca. *Journal of the American Chemical Society* **141**, 928-935 (2019). https://doi.org/10.1021/jacs.8b10724

60 DiCaprio, A. J., Firouzbakht, A., Hudson, G. A. & Mitchell, D. A. Enzymatic Reconstitution and Biosynthetic Investigation of the Lasso Peptide Fusilassin. *J Am Chem Soc* **141**, 290-297 (2019). https://doi.org/10.1021/jacs.8b09928

61 Sánchez-Hidalgo, M., Martín, J. & Genilloud, O. Identification and Heterologous Expression of the Biosynthetic Gene Cluster Encoding the Lasso Peptide Humidimycin, a Caspofungin Activity Potentiator. *Antibiotics* **9**, 67 (2020).

62 Cheung, W. L., Chen, M. Y., Maksimov, M. O. & Link, A. J. Lasso Peptide Biosynthetic Protein LarB1 Binds Both Leader and Core Peptide Regions of the Precursor Protein LarA. *ACS Central Science* **2**, 702-709 (2016). https://doi.org/10.1021/acscentsci.6b00184

63 Yan, K.-P. *et al.* Dissecting the Maturation Steps of the Lasso Peptide Microcin J25 in vitro. *ChemBioChem* **13**, 1046-1052 (2012). https://doi.org/https://doi.org/10.1002/cbic.201200016

64 Si, Y., Kretsch, A. M., Daigh, L. M., Burk, M. J. & Mitchell, D. A. Cell-Free Biosynthesis to Evaluate Lasso Peptide Formation and Enzyme–Substrate Tolerance. *Journal of the American Chemical Society* **143**, 5917-5927 (2021). https://doi.org/10.1021/jacs.1c01452

65 He, S. & Deber, C. M. Interaction of designed cationic antimicrobial peptides with the outer membrane of gram-negative bacteria. *Scientific Reports* **14**, 1894 (2024). https://doi.org/10.1038/s41598-024-51716-1

66 Lei, J. *et al.* The antimicrobial peptides and their potential clinical applications. *Am J Transl Res* **11**, 3919-3931 (2019).

67 Ramanan, P., Barreto Jason, N., Osmon Douglas, R. & Tosh Pritish, K. Rothia Bacteremia: a 10-Year Experience at Mayo Clinic, Rochester, Minnesota. *Journal of Clinical Microbiology* **52**, 3184-3189 (2020). https://doi.org/10.1128/jcm.01270-14

68 Hegemann, J. D. in *Methods in Enzymology* Vol. 663 (ed Leslie M. Hicks) 177-204 (Academic Press, 2022).

69 Garcie, C. *et al.* The Bacterial Stress-Responsive Hsp90 Chaperone (HtpG) Is Required for the Production of the Genotoxin Colibactin and the Siderophore Yersiniabactin in Escherichia coli. *The Journal of Infectious Diseases* **214**, 916-924 (2016). https://doi.org/10.1093/infdis/jiw294

70 Washio, K., Lim, S. P., Roongsawang, N. & Morikawa, M. Identification and Characterization of the Genes Responsible for the Production of the Cyclic Lipopeptide Arthrofactin by <I>Pseudomonas</I> sp. MIS38. *Bioscience, Biotechnology, and Biochemistry* **74**, 992-999 (2010). https://doi.org/10.1271/bbb.90860

71 Burkhart, B. J., Kakkar, N., Hudson, G. A., van der Donk, W. A. & Mitchell, D. A. Chimeric Leader Peptides for the Generation of Non-Natural Hybrid RiPP Products. *ACS Central Science* **3**, 629-638 (2017). https://doi.org/10.1021/acscentsci.7b00141

72 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* **58**, 4169-4182 (2019). https://doi.org/10.1021/acs.biochem.9b00735

73 Oberg, N., Zallot, R. & Gerlt, J. A. EFI-EST, EFI-GNT, and EFI-CGFP: Enzyme Function Initiative (EFI) Web Resource for Genomic Enzymology Tools. *Journal of Molecular Biology* **435**, 168018 (2023). https://doi.org/https://doi.org/10.1016/j.jmb.2023.168018