

Reactivity-based screening for citrulline-containing natural products reveals a family of bacterial peptidyl arginine deiminases

Lonnie A. Harris¹, Patricia M. B. Saint-Vincent^{1,2†}, Xiaorui Guo¹, Graham A. Hudson¹, Douglas A. Mitchell^{1,2,3*}

¹Department of Chemistry, ²Carl R. Woese Institute for Genomic Biology, ³Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, United States

†Current address: Leidos, National Energy Technology Laboratory, Pittsburgh, PA

*Correspondence: douglas.mitchell@illinois.edu, phone: 1-217-333-1345, fax: 1-217-333-0508
600 S. Mathews Avenue, Urbana, IL 61802

Author ORCID accounts:

LAH: 0000-0002-0312-8806

PMBS: 0000-0001-6360-8206

XRG: 0000-0001-8668-2839

GAH: 0000-0002-3715-4279

DAM: 0000-0002-9564-0953

ABSTRACT

Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a family of natural products defined by a genetically encoded precursor peptide that is tailored by associated biosynthetic enzymes to form the mature product. Lasso peptides are a class of RiPP defined by an isopeptide linkage between the N-terminal amine and an internal Asp/Glu residue with the C-terminus threaded through the macrocycle. This unique lariat topology, which provides considerable stability towards heat and proteases, has stimulated interest in lasso peptides as potential therapeutics. Post-translational modifications beyond the class-defining, threaded macrolactam have been reported, including one example of arginine deimination to yield citrulline. Although a citrulline-containing lasso peptide (i.e., citrulassassin) was serendipitously discovered during a genome-guided campaign, the gene(s) responsible for arginine deimination has remained unknown. Herein we describe the use of reactivity-based screening to discriminate bacteria that produce arginine- versus citrulline-bearing citrulassins, culminating in the discovery and characterization of 11 new lasso peptide variants. Phylogenetic profiling identified a distally encoded peptidyl arginine deiminase (PAD) gene ubiquitous to the citrulline-containing variants. Absence of this gene correlated strongly with citrulassassin variants only containing arginine (*des*-citrulassassin). Heterologous expression of the PAD in a non-citrulassassin producer resulted in the production of the deiminated analog, confirming PAD involvement in arginine deimination. The family of PADs were then bioinformatically surveyed for a deeper understanding of its genomic context and potential role in post-translational modification of RiPPs.

Introduction

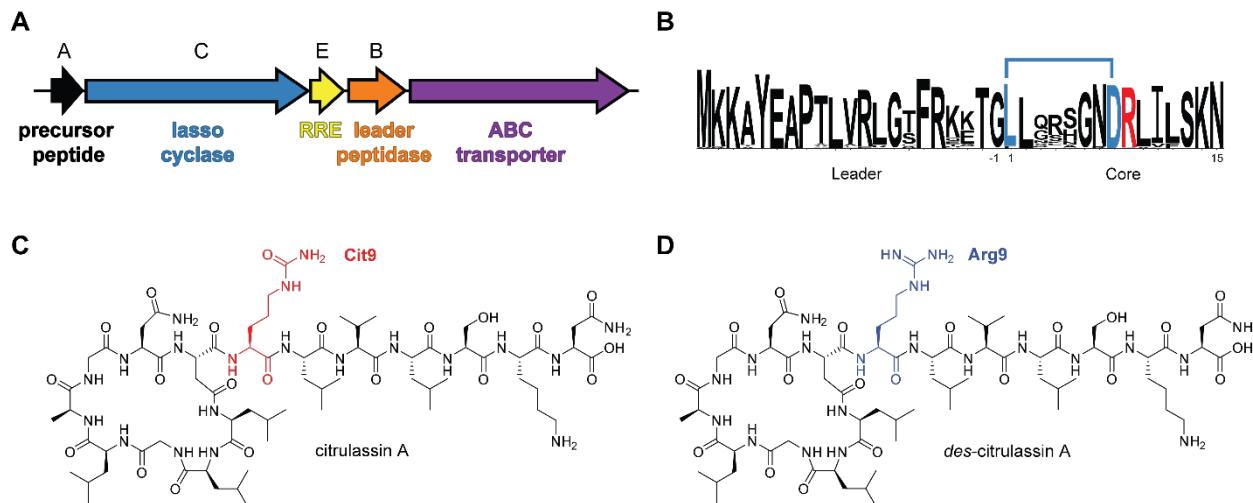
Ribosomally synthesized and post-translationally modified peptides (RiPPs) are a family of natural products (NPs) defined by a common biosynthetic logic, wherein a genetically encoded precursor peptide is modified by associated biosynthetic enzymes to form the mature NP.^{1, 2} Precursor peptides are typically bipartite, composed of an N-terminal leader region responsible for enzyme recognition and a C-terminal core region which is modified and cleaved from the leader region to form the mature RiPP.^{1, 2} Because substrate recognition is guided by the leader sequence, RiPP biosynthetic enzymes are typically permissive of sequence changes in the core region, facilitating access to analogs.³ One RiPP class that has attracted interest are the lasso peptides, which are defined by a unique lariat topology that grants them thermal stability and protease resistance.⁴ These properties, in addition to the substrate permissiveness of several characterized lasso biosynthetic enzymes, have engendered interest in the design of lasso peptides therapeutics.^{5, 6, 7, 8, 9}

After ribosomal synthesis of the lasso precursor peptide, the next biosynthetic step involves RiPP recognition element (RRE) engagement of the precursor peptide, which occurs through binding to a recognition sequence within the N-terminal leader region. The RRE domain mediates recruitment of the subsequent enzymes^{10, 11} and is found as either a discretely encoded protein or fused to the leader peptidase. Upon RRE binding, the leader peptidase removes the leader region from the precursor peptide.^{12, 13} The characteristic lariat topology is then formed by an ATP-dependent lasso cyclase, which installs a macrolactam linkage between the newly formed N-terminus of the core with the carboxylate of a downstream Asp or Glu acceptor residue, trapping the C-terminal tail within the ring.^{12, 13}

As with other RiPP classes, post-translational modifications (PTMs) beyond the class-defining, threaded macrolactam are known, among them are disulfide formation (e.g. BI-32169 and others),¹⁴ C-terminal *O*-methylation (e.g. lassomycin),¹⁵ Asp β -hydroxylation (e.g. canucin A),¹⁶ Lys ϵ -acylation (e.g. albusnordin),¹⁷ Trp 7-hydroxylation (e.g. RES-701-2),¹⁸ epimerization to D-Trp (e.g. MS-271),¹⁹ Ser *O*-phosphorylation (e.g. paeninodin),²⁰ and Arg deimination (e.g. citrulassassin A).²¹ The latter modification was discovered during a genomics-guided discovery of novel lasso peptides, enabled by the bioinformatics tool Rapid ORF Description and Evaluation Online (RODEO). Briefly, RODEO uses profile hidden Markov models (pHMMs), heuristic scoring, and supervised machine learning to identify

putative RiPP biosynthetic gene clusters (BGCs) and score potential precursors based on known family members.²¹ Since its initial application to lasso peptides, additional RODEO modules have been developed and used to mine and classify other RiPP classes, such as thiopeptides,²² lanthipeptides,²³ sactipeptides, and ranthipeptides.²⁴

During our previous RODEO-guided lasso peptide discovery effort, an uncharacterized family of 55 lasso peptides was bioinformatically identified (Figure 1).²¹ The first member of this family, citrulassin A, was isolated and characterized from *Streptomyces albulus* NRRL B-3066. Citrulassin A was so named owing to an unprecedented citrulline at position nine of the core region (Cit9) rather than the genetically encoded Arg. Because citrulline is a non-proteinogenic amino acid, it was hypothesized that a peptidyl arginine deiminase (PAD) might be responsible for forming the citrulline moiety of citrulassin A, as this enzyme family carries out post-translational deimination of Arg in other biological contexts. However, genome sequencing of *S. albulus* NRRL B-3066 revealed no enzymes with predicted PAD activity within 15 open-reading frames of the citrulassin A cyclase (WP_079136914.1). Moreover, heterologous expression of the citrulassin A BGC with large flanking regions (~20 kbp on each side of the BGC) in *Streptomyces lividans* resulted in the production of *des*-citrulassin A, which contains Arg at position nine of the core (Arg9) instead of citrulline (Figure 1). Given this result, the gene(s) responsible for Arg deimination was presumed to be distally encoded to the citrulassin A BGC.²¹



(A) General lasso peptide biosynthetic gene cluster architecture. (B) Multiple sequence alignment (<http://weblogo.berkeley.edu/logo.cgi>)⁸ of the 181 bioinformatically identified citrulassin precursor peptides. Isopeptide bond formation between the Asp side chain and N-terminus of the core region is indicated with a blue bar. The Arg residue modified to citrulline in citrulassin A is red. (C) Two-dimensional structure of citrulassin A from *S. albulus* NRRL B-3066, with Cit9

highlighted in red. (D) Two-dimensional structure of *des*-citrulassassin A heterologously expressed in *S. lividans*, with Arg9 highlighted in blue. Non-threaded versions of the structures are shown for improved clarity.

To the best of our knowledge, only one bacterial PAD has been characterized. This PAD is from *Porphyromonas gingivalis*, conserved amongst periodontal pathogens,²⁵ and contributes to *P. gingivalis* virulence. The *P. gingivalis* PAD enzyme deimines human fibrinogen, leading to gingivitis and an increased risk of rheumatoid arthritis.²⁶ This PAD virulence factor (protein family²⁷ PF08527, PAD_porph) is distinct from human PADs (protein family PF03068), of which the PAD4 isoform is the best studied. PAD4 deimines histones, which can affect transcription as well as cellular differentiation, and has been implicated in the onset of rheumatoid arthritis and in certain cancers.²⁸

To identify the putative PAD involved in the deimination of citrulassassin A, further screening of citrulassassin producers for the presence of citrulline was required prior to pursuing a comparative genomic approach. Arg deimination, however, is difficult to discern by mass spectrometry, as the PTM results in a small mass deviation (replacement of NH with O, +0.98 Da) that is easily overlooked, especially when both Arg- and citrulline-containing congeners are produced. Moreover, deimination of Asn/Gln to the corresponding carboxylic acid is isobaric with deamidation, a commonly observed, spontaneous side chain modification, further complicating identification of deimination.²⁹ Building off of precedent in natural product discovery and proteomics, we have previously described strategies for the rapid discovery of NPs bearing specific functional groups in the extracts of cultured bacteria, termed reactivity-based screening (RBS) (**Figure S1**).³⁰ RBS involves the chemoselective labeling of an organic functional group, whereby NPs bearing the targeted functional group are rapidly detected via comparative mass spectrometry (MS) of reacted and unreacted samples. RBS has enabled the discovery and subsequent characterization of RiPPs, non-ribosomal peptides, polyketides, and hybrids thereof (e.g. cyclothiazomycin C,³¹ *deimino*-antipain,³² hygrobafilomycin JBIR-100,³³ and tirandalydigin,³⁴ *etc.*). RBS is therefore a valuable tool for bridging bioinformatics-guided NP identification with NP production. Additionally, an RBS-based NP discovery strategy permits rapid dereplication of already-characterized NPs,³¹ detection of low-abundance species using sensitive MS-based techniques such as matrix-assisted laser/desorption ionization time-of-flight MS (MALDI-TOF-MS), and can employ variable probe chemistries that

enhance NP detection due to unique isotopic patterns of labeled compounds or selective enrichment through affinity purification.^{32, 35, 36}

Herein, we describe the expansion of RBS as a tool to aid in the discovery of the PAD responsible for deimination activity on the citrulassin family of lasso peptides. Repurposing the work of Thompson,^{37, 29} a phenylglyoxal probe was employed to specifically label primary ureido groups over primary guanidino groups. This probe was used to survey for the presence of citrulline in a subset of lasso peptides and enable phylogenetic profiling, thus revealing a putative PAD in each citrulassin producer. A broader bioinformatic survey of bacterial PADs was also performed, revealing their taxonomic breadth and diversity of genomic contexts. Lastly, we heterologously express a PAD from *Streptomyces glaucescens* and demonstrate the deimination of a citrulassin precursor peptide *in vivo*, providing the first gene-to-function link for RiPP citrulline formation.

Results and Discussion

Validation of 3-bromophenylglyoxal as a primary ureido-selective probe

Previous work has demonstrated selective labeling of citrulline in eukaryotic cellular extracts at low pH with phenylglyoxal probes (**Figure 2**).^{29, 37} Under sufficiently acidic conditions, guanidino groups are protonated and thus unreactive towards the glyoxal moiety; however, ureido groups remain nucleophilic and thus react rapidly.³⁷ The reaction of citrulline with phenylglyoxal has been further elaborated to readily identify labeled species in complex samples by employing brominated phenylglyoxal derivatives. The bromine provides a distinct isotopic pattern (~1:1 ⁷⁹Br/⁸¹Br) while the UV-absorbing arene increases signal intensity in MALDI-TOF-MS.³⁸ Moreover, the probe can easily be modified to enrich for labeled compounds using a biotin-conjugated glyoxal probe.²⁷ Thus, we imagined that 3-bromophenylglyoxal (**1**) could be deployed to discover primary ureido containing NPs from bacterial extracts. As predicted from the known reactivity of glyoxal, the free amino acid L-citrulline was robustly modified by **1** while L-Arg was unmodified under identical conditions (**Figure S2**). To evaluate the suitability of reaction within the context of a cellular extract, *S. albulus* NRRL B-3066 (the native producer of citrulassin A) and *Streptomyces lividans* 3H4 (a

heterologous producer of *des*-citrulassassin A)²⁰ were grown for 7 d and metabolites were extracted from the cell-surface with methanol. When treated with **1** under identical reaction conditions, citrulassassin A was nearly fully labeled while negligible product was observed for *des*-citrulassassin A (**Figure 2**). Moreover, *deimino*-antipain, another citrulline-containing compound produced by *S. albulus* NRRL B-3066, was labeled under these conditions (**Figure S3**). Extract from *S. lividans* 3H4²¹ heterologously producing antipain confirmed selectivity for citrulline over Arg, with antipain displaying negligible labeling by **1**. Taken together, these experiments validate **1** as useful in the rapid identification of NPs containing a primary ureido group.

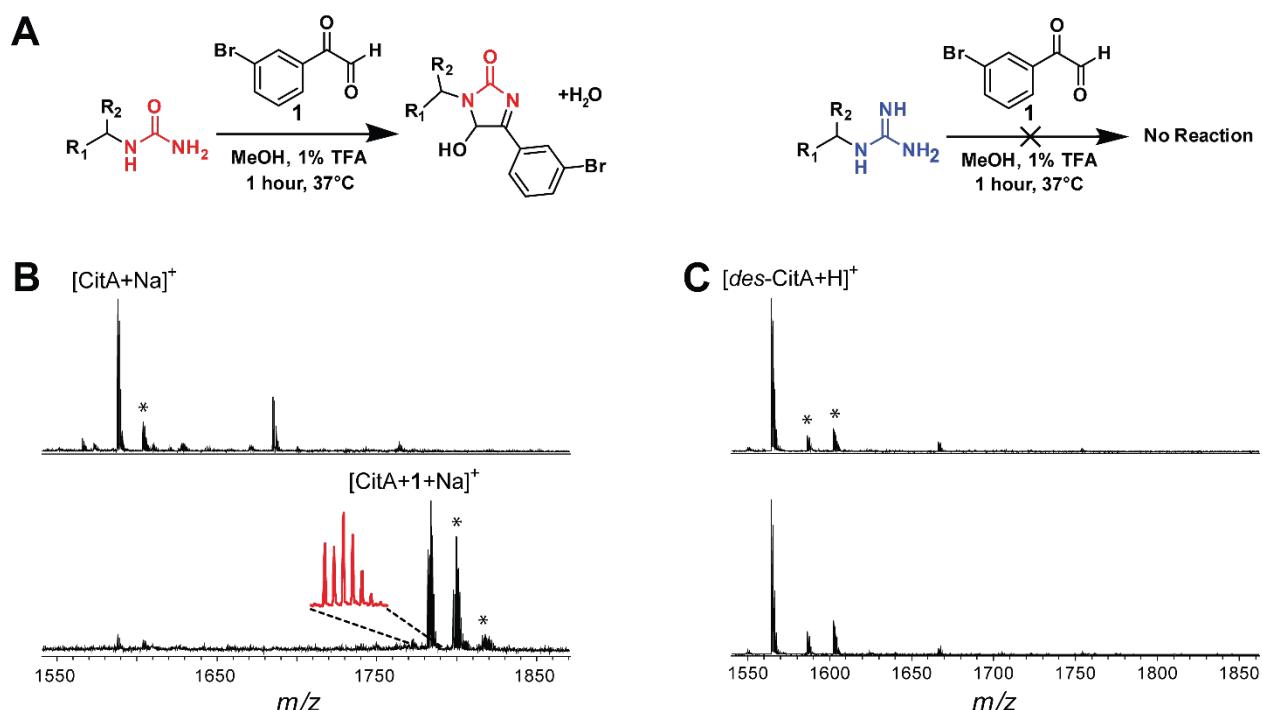


Figure 2. Validation of 3-bromophenylglyoxal (**1**) selectivity. (A) Reaction conditions that selectively label citrulassin (left) over *des*-citrulassin (right). TFA = trifluoroacetic acid. (B) MALDI-TOF mass spectra of a methanolic extract of *Streptomyces albulus* (citrulassin A producer) prior to addition of **1** (*top*) and after reaction with **1** (*bottom*). Magnified inset shows the $^{79}\text{Br}:\mathbf{^{81}\text{Br}}$ isotopic pattern. Peaks indicated with an asterisk are (from left to right): $[\text{CitA}+\text{K}]^+$, $[\text{CitA}+\mathbf{1}+\text{K}]^+$, and $[\text{CitA}+\mathbf{1}+\text{K}+\text{H}_2\text{O}]^+$. (C) MALDI-TOF mass spectra of a methanolic extract of *des*-citrulassin A heterologously expressed in *Streptomyces lividans* (ref. 21) prior to addition of **1** (*top*) and after reaction with **1** (*bottom*). Peaks indicated with an asterisk are (from left to right): $[\text{des-CitA}+\text{Na}]^+$ and $[\text{des-CitA}+\text{K}]^+$.

3-bromophenylglyoxal (1) screening of citrulassin producers

With **1** validated, we next sought to identify which members of the citrulassin family of lasso peptides contain a post-translationally derived citrulline. Previous screening of bacterial extracts suggested that not all citrulassins contain citrulline (the Arg-containing counterpart is referred to as *des*-citrulassin),²¹ but owing to a small mass difference

between guanido- and ureido-containing products, we tested all putative citrulassins through reactivity with **1**. To identify probable citrulassins, a BLAST-p search was performed using the citrulassin A cyclase as a query (WP_079136914.1). Cyclases with an e-value $<10^{-200}$ were then subjected to RODEO analysis. The identified citrulassin precursor peptides were remarkably well conserved, with only a few positions in the ring and loop regions showing sequence variation (**Figure 1** and **Supplemental Dataset 1**). In addition, co-occurrence analysis indicated that no putative PADs are encoded in the genomic neighborhood of the citrulassin BGCs, in accord with the earlier heterologous expression study.²¹ With this list in hand, a selection of 109 strains predicted to contain citrulassin BGCs were obtained from the Agriculture Research Service (ARS) Culture Collection and grown for 10 d on a variety of media at 30 °C. Afterwards, the bacteria were harvested and metabolites extracted from the cell surface using MeOH for 2 h (see methods). Reactions with **1** were performed directly on the methanolic extracts and subsequently dried, reconstituted in methanol, and analyzed by MALDI-TOF-MS for the presence of primary ureido-containing metabolites. Extracts from which a predicted citrulassin were detected ($n = 12$) were screened for Arg deimination using reactivity towards **1** (**Table 1**). Extracts from six of these strains underwent labeling and were subjected to high-resolution and tandem mass spectrometry (HR-MS/MS), which confirmed the molecular formula and location of citrulline(s) (**Figure S5**). Several citrulassins identified contain Arg within the ring, typically at position four, in addition to the conserved Arg9 encoded within the loop region. Citrulline formation was confined to position 9 for all citrulassins with the exception of one variant from *Streptomyces* sp. NRRL S-920, which contained a second citrulline moiety within the ring. The extent of deimination was accurately reflected by reactivity towards **1**, with all singly deiminated citrulassins showing exactly one labeling event, while the doubly deiminated citrulassin was labeled twice (**Figure S5**). Thus, this approach is useful for determining the stoichiometry of targetable functional groups on a compound of interest.

Table 1: Characterized citrulassins

Bacterial strain (NRRL)	Core Sequence	Citrulassin	Deimination	PAD
<i>Streptomyces albulus</i> B-3066	LLGLAGNDRLVLSKN	A	Y	Y
<i>Streptomyces tricolor</i> B-16925	LLQRSGNDRLILSKN	C	N	N
<i>Streptomyces katrae</i> B-16271	LLGRSGNDRLVLSKN	D	N	N
<i>Streptomyces glaucescens</i> B-11408	LLQRHGNDRLILSKN	E	Y	Y
<i>Streptomyces avermitilis</i> B-16169	LLGRSGNDRLILSKN	F	N	N
<i>Streptomyces torulosus</i> S-189	LLGRSGNDRLILSKN	F	Y	Y
<i>Streptomyces auratus</i> 8097	LLNSSGNDRLILSKN	G	N	N
<i>Streptomyces</i> sp. S-118	LLAFHGNDRLILSKN	H	Y	Y
<i>Streptomyces natalensis</i> B-5314	LLEFRGNDRLILSKN	J	N	N
<i>Streptomyces</i> sp. S-920	LLRRSGNDRLILSKN	K	Y	Y
<i>Streptomyces</i> sp. S-481	LLGDAGNDRLILSKN	L	N	Y
<i>Streptomyces pharetrae</i> B-24333	LLQRNGNDRLILSKN	M	Y	Y

Phylogenetic profiling identifies a putative PAD for citrulline formation

Knowing which sequenced *Streptomyces* strains produce citrulassin versus *des*-citrulassin allowed a phylogenetic profiling strategy to identify the putative gene(s) responsible for Arg deimination. Using PhyloProfile,^{39,40} a search was conducted for genes present in the *S. albulus* genome that were also present in four other citrulassin producers but not in four *des*-citrulassin producers. This analysis returned 32 *S. albulus* genes that correlated with the citrulline phenotype, one of which encoded a hypothetical protein containing a PAD domain (**Supplemental Dataset 1**). This gene encodes a member of protein family PF03068 and thus is homologous to human PADs. No significant sequence homology is shared between the *S. albulus* PAD (WP_064069847.1) and the only other characterized bacterial PAD from *P. gingivalis* (protein family PF04371, WP_005873463.1).^{27,25} (**Table S3**) Position-specific iterative BLAST (PSI-BLAST)⁴¹ of citrulassin producers revealed a correlation between the genome of a bacterial strain encoding a PAD and the presence of citrulline in the mature citrulassin, with only one strain producing *des*-citrulassin while having a PAD in the genome. Analogously, all organisms that lacked a PAD were associated with *des*-citrulassin productions. However, the small sample size prevented a statistically relevant correlation between genotype and chemotype from being established.⁴² Nonetheless, having identified a candidate PAD, we set out to further examine bacterial members of protein family PF03068.

Bioinformatic survey of PF03068 bacterial PADs

While the *S. albulus* PAD is classified within the same protein family as eukaryotic PADs (PF03068), the sequence similarity is only 42-47% between each of the five human PAD isoforms and the PAD from *S. albulus* (**Table S4**). Despite this modest sequence similarity, an alignment of the *S. albulus* PAD with human PAD4, the best-characterized isoform, shows conservation of functionally important residues. The human PADs are calcium-dependent enzymes with a conserved catalytic Cys645 and stabilizing triad of Asp350-His471-Asp473;⁴³ these catalytic residues, as well as numerous Ca²⁺-binding residues are retained between the human and *S. albulus* PADs (**Figure S6**). In contrast, the *P. gingivalis* (PAD_porph) is calcium-independent and cannot be aligned with the *S. albulus* PAD outside of short motifs near the active site (**Figure S7**). Moreover, few functional residues are conserved beyond those common to all members of the guanidino-modifying enzyme superfamily.^{44,45}

The *S. albulus* PAD was used to generate a dataset of 837 bacterial PF03068 PADs using a PSI-BLAST query against all bacterial genomes in the NCBI non-redundant database (**Supplemental Dataset 1**).⁴⁶ Protein accession identifiers were compiled, analyzed by RODEO,²¹ and then used to generate a sequence similarity network (SSN) and a maximum-likelihood phylogenetic tree to visualize the distribution of bacterial PADs belonging to PF03068 (**Figures 4 and S8**).⁴⁷ Homologs of the *S. albulus* PAD are present in diverse bacterial phyla, but are predominantly in Actinobacteria, Cyanobacteria, and Proteobacteria, with very few homologs in Bacteroidetes and Firmicutes. In contrast, members of PF04371 (PAD_porph), are more evenly distributed across prokaryotic phyla and are present in Actinobacteria, Proteobacteria, Firmicutes and Bacteroidetes, the latter including *P. gingivalis* (<https://pfam.xfam.org/family/PF04371#tabview=tab7>).²⁷ Genomes that encode a PAD within these genera is fairly limited, as illustrated by only 1.6%, 0.1%, and 3.9% of sequenced actinobacteria, proteobacteria, and cyanobacteria, respectively, featuring an annotated homolog. Interestingly, the highest occurrence of annotated PADs is in Riflebacteria, which despite having only 11 genomes in the NCBI database, includes 15 annotated PADs. With a few notable examples, there is little evidence of horizontal gene transfer with PF03068, as proteins tend to cluster with other PADs from species within the same phylum (**Figures 4 and S8**). Additionally, the %GC content of the genes encoding the PAD correlate strongly with the overall %GC content of the genome (**Figure S9**).⁴⁸ These data indicate

bacterial members of PF03068 are not disseminated across bacterial genomes due to recent horizontal gene transfer, although we cannot rule out the possibility that the genes were acquired by organisms having similar GC content. Notably, several proteobacterial PADs are more closely related to the actinobacterial PAD clade that includes those implicated in citrulassassin Arg deimination. Intriguingly, most of these PAD-containing Proteobacteria are uncultured bacteria detected in aquatic and sediment metagenomes, raising questions concerning the genomic origins and ecology of PADs.

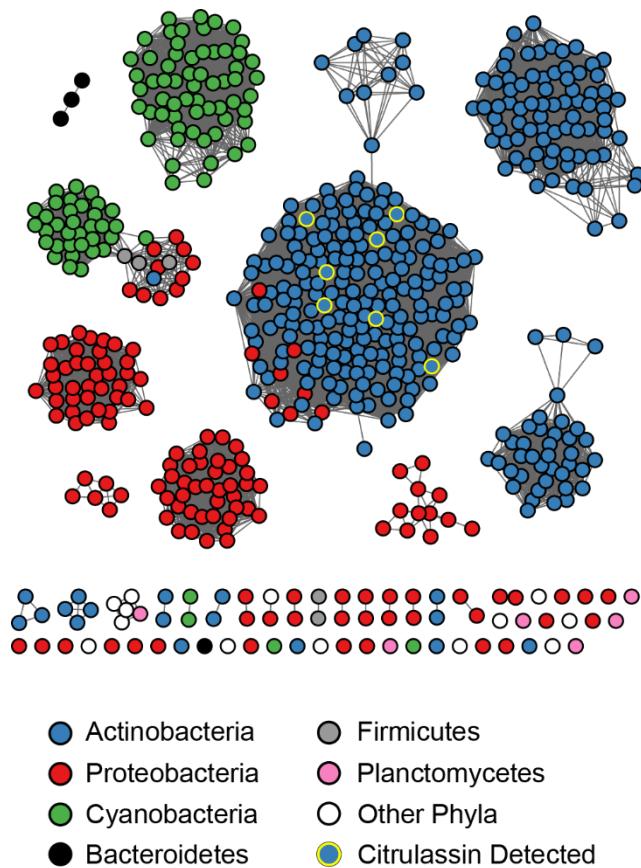


Figure 3. Protein sequence similarity network (SSN) of bacterial PADs. The SSN was generating using EFI-EST⁴⁷, Sequences sharing 100% identity are conflated as a single node, and connected nodes indicate an alignment score of 125, see methods) and visualized by Cytoscape.⁴⁹ Nodes are colored by phylum and detection of citrulassassin (deiminated product) within the same strain.

Co-occurrence analysis of citrulassassin BGCs (± 8 coding sequences) shows little correlation beyond known lasso peptide biosynthetic proteins and furthermore demonstrates that PADs do not appear within lasso peptide biosynthetic operons (Table S3). Even PADs with a high degree of similarity appear in diverse genome neighborhoods, which do

not bear resemblance to RiPP BGCs. The contexts within which the bacterial members of PF03068 appear are highly diverse (**Figure S10**). Thus, these PADs likely exert their activity on distally encoded substrate(s).

PAD complementation in a *des*-citrulassassin producer results in citrulassassin production

The co-occurrence of PAD with the production of citrulassassin provided strong, but ultimately circumstantial, support for the proposed role in Arg deimination. To further evaluate this putative biochemical role, a functional expression experiment designed to introduce the PAD-encoding gene from a verified citrulassassin producer to the genome of a verified *des*-citrulassassin producer that lacks a PAD-encoding gene. Using the *Streptomyces* phage φ C31 integrase, plasmids containing an *attP* locus can be inserted into the chromosome of a recipient organism at the corresponding *attB* locus. Nearly all identified *des*-citrulassassin producer genomes contain a φ C31 *attB*, enabling *Escherichia coli*-*Streptomyces* conjugative transfer.^{50,51} The integrative plasmid places the PAD under strong constitutive promotion (*ermE*^{*}*p*) to facilitate a higher level of expression within the heterologous host.^{51, 52} The *S. glaucescens* PAD was integrated into *Streptomyces katrae* NRRL B-16271, a native producer of *des*-citrulassassin D (**Figure 4**, donor and recipient were arbitrarily chosen). Exconjugants were isolated by iterative cultivation on a selective medium and chromosomal insertion of the PAD gene was confirmed via DNA sequencing (**Figure S11**). Successfully integrated strains were then grown on a medium that elicited *des*-citrulassassin D production in wild-type *S. katrae*. Metabolites were extracted with MeOH and subsequently analyzed via MALDI-TOF-MS. PAD integration in *S. katrae* resulted in the production of a ion consistent with citrulline-containing citrulassassin D (**Figure 4**). Reaction with **1** resulted in complete labeling, while wild-type *des*-citrulassassin D was not labeled (**Figure 4**), suggesting that PAD integration had resulted in the conversion of the Arg within *des*-citrulassassin D to citrulline. HRMS/MS confirmed that the produced lasso peptide was 0.98 Da heavier than expected for *des*-citrulassassin D and elicited a fragmentation pattern consistent with the sequence of citrulassassin D, featuring Arg9 converted to citrulline. Notably, the site-specific deimination of Arg9 of *des*-citrulassassin D over Arg4 confirms our previously assigned site of deimination on citrulassassin and demonstrates that expression of PAD is necessary and sufficient for conversion of Arg to citrulline in the context of citrulassassin biosynthesis. Future work will be necessary to uncover the origins of the selectivity and the timing of deimination during citrulassassin biosynthesis.

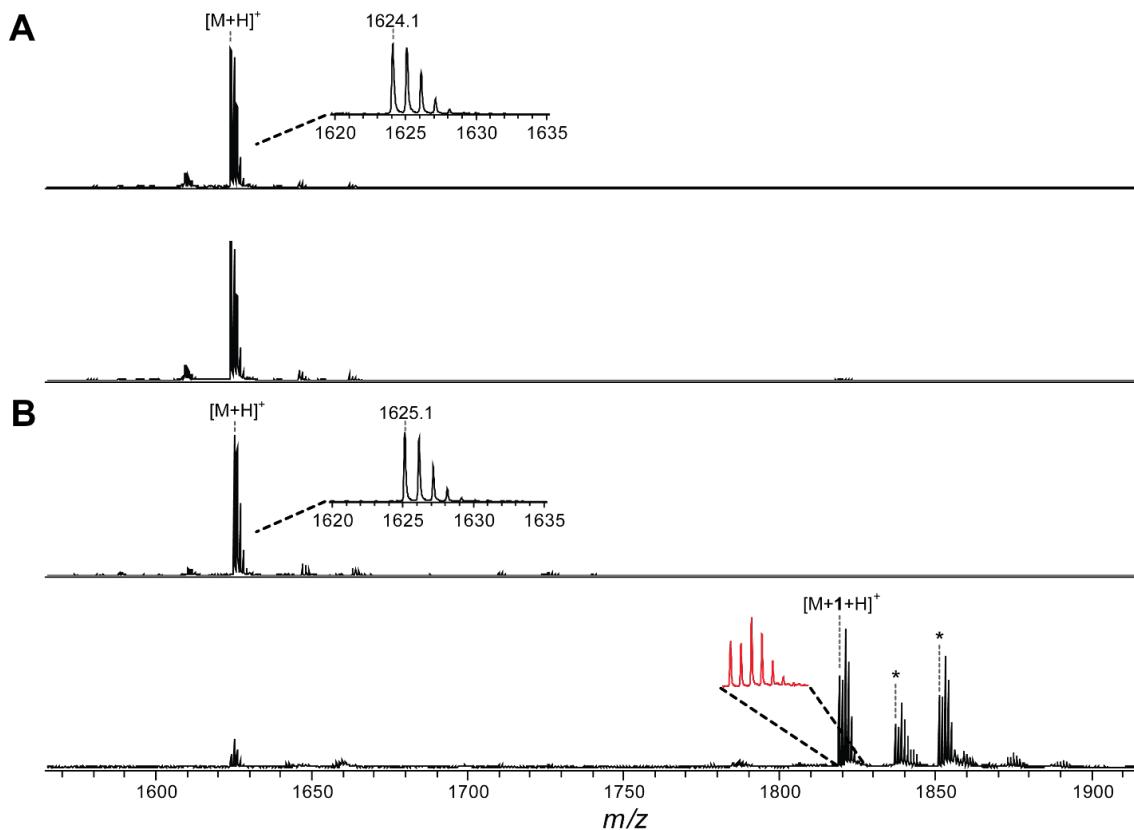


Figure 4 Citrulassassin production in a *des*-citrulassassin producer by PAD heterologous expression (A) MALDI-TOF mass spectrum of *Streptomyces katrae* extract containing *des*-citrulassassin D unreacted (top) and reacted (bottom) with 1. (B) MALDI-TOF mass spectrum of *S. katrae* extract expressing the PAD from *Streptomyces glaucescens*. Shown are samples unreacted (top) and reacted (bottom) with 1. Insets show zoomed regions to visualize $^{79}\text{Br}:\text{Br}^{81}$ isotopic ratios. Asterisks denote solvated adducts corresponding to $[M+1+H_2O+H]^+$ (left) and $[M+1+MeOH+H]^+$ (right).

Conclusion

Advances in genomic sequencing and bioinformatics have provided a foundation for a renaissance in NP discovery using genome-guided techniques, previously illustrated by the discovery of numerous new lasso peptides, including the citrulline-containing citrulassins. The versatility of genome-guided natural product discovery is strongly complemented by reactivity-based screening which exploits chemoselective targeting of organic functional groups to facilitate metabolite identification and isolation. In this work, we expand the toolkit of RBS to include primary ureido groups (i.e. citrulline) through their reactivity towards glyoxal-based probes, which, under mildly acidic conditions, are largely inert towards guanidino groups (i.e. Arg). Harnessing this reactivity led to the discovery of 11 new citrulassins and provided a means to perform comparative genomic analysis to discover the origins of the previously enigmatic Arg

deimination in citrulassin biosynthesis. These analyses suggested that a family of bacterial PADs, divergent from the better-known bacterial PAD from *P. gingivalis*, were responsible for citrulassin deimination despite a lack of co-genomic context conservation. We confirmed the Arg deimination activity of this PAD converting a *des*-citrulassin producer into a citrulassin producer through heterologous expression of a PAD. This work highlights the synergy between genome- and reactivity-guided discovery approaches and provides a platform for identifying additional bacteria that produce NPs featuring the rare, non-proteinogenic primary ureido functional group. More broadly, we expect this workflow to be amenable towards uncovering both the prevalence and enzymatic origins of other reactive moieties present in RiPPs and other NPs.

ASSOCIATED CONTENT

The Supporting Information is available free of charge via the web. This document contains a description of experimental methods, mass spectral characterization of newly discovered citrulassins, and bioinformatic analysis of the citrulassins and PAD enzymes.

AUTHOR INFORMATION

Corresponding Author

* douglasm@illinois.edu

Funding sources

This work was supported by a grant from the National Institutes of Health (GM123998 to D.A.M.) and the Seemon Pines Fellowship from the Department of Chemistry at the University of Illinois at Urbana-Champaign (to G.A.H.). Funds to purchase the Bruker UltrafleXtreme MALDI TOF/TOF mass spectrometer were from the National Institutes of Health (S10 RR027109 A).

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENT

We are grateful to W. W. Metcalf for access to many of the actinobacterial strains screened in this study, as well as the *E. coli* strains used for bacterial conjugation. P. Patel provided the script for automated % GC analysis.

References

- (1) Arnison, P.; Bibb, M.; Bierbaum, G.; Bowers, A.; Bugni, T.; Bulaj, G.; Camarero, J.; Campopiano, D.; Challis, G.; Clardy, J.; Cotter, P.; Craik, D.; Dawson, M.; Dittmann, E.; Donadio, S.; Dorrestein, P.; Entian, K.-D.; Fischbach, M.; Garavelli, J.; Göransson, U.; Gruber, C.; Haft, D.; Hemscheidt, T.; Hertweck, C.; Hill, C.; Horswill, A.; Jaspars, M.; Kelly, W.; Klinman, J.; Kuipers, O.; James Link, A.; Liu, W.; Marahiel, M.; Mitchell, D.; Moll, G.; Moore, B.; Müller, R.; Nair, S.; Nes, I.; Norris, G.; Olivera, B.; Onaka, H.; Patchett, M.; Piel, J.; Reaney, M. J.; Rebuffat, S.; Paul Ross, R.; Sahl, H.-G.; Schmidt, E.; Selsted, M.; Severinov, K.; Shen, B.; Sivonen, K.; Smith, L.; Stein, T.; Süssmuth, R.; Tagg, J.; Tang, G.-L.; Truman, A.; Vedera, J.; Walsh, C.; Walton, J.; Wenzel, S.; Willey, J.; Donk, W. A. van der. Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products: Overview and Recommendations for a Universal Nomenclature. *Nat. Prod. Rep.* **2013**, *30* (1), 108–160. <https://doi.org/10.1039/C2NP20085F>.
- (2) Ortega, M. A.; van der Donk, W. A. New Insights into the Biosynthetic Logic of Ribosomally Synthesized and Post-Translationally Modified Peptide Natural Products. *Cell Chem. Biol.* **2016**, *23* (1), 31–44. <https://doi.org/10.1016/j.chembiol.2015.11.012>.
- (3) Hudson, G. A.; Mitchell, D. A. RiPP Antibiotics: Biosynthesis and Engineering Potential. *Curr. Opin. Microbiol.* **2018**, *45*, 61–69. <https://doi.org/10.1016/j.mib.2018.02.010>.
- (4) Hegemann, J. D.; Zimmermann, M.; Xie, X.; Marahiel, M. A. Lasso Peptides: An Intriguing Class of Bacterial Natural Products. *Acc. Chem. Res.* **2015**, *48* (7), 1909–1919. <https://doi.org/10.1021/acs.accounts.5b00156>.
- (5) Pavlova, O.; Mukhopadhyay, J.; Sineva, E.; Ebright, R. H.; Severinov, K. Systematic Structure-Activity Analysis of Microcin J25. *J. Biol. Chem.* **2008**, *283* (37), 25589–25595. <https://doi.org/10.1074/jbc.M803995200>.
- (6) Knappe, T. A.; Manzenrieder, F.; Mas-Moruno, C.; Linne, U.; Sasse, F.; Kessler, H.; Xie, X.; Marahiel, M. A. Introducing Lasso Peptides as Molecular Scaffolds for Drug Design: Engineering of an Integrin Antagonist. *Angew. Chem. Int. Ed.* **2011**, *50* (37), 8714–8717. <https://doi.org/10.1002/anie.201102190>.
- (7) Al Toma, R. S.; Kuthning, A.; Exner, M. P.; Denisiuk, A.; Ziegler, J.; Budisa, N.; Süssmuth, R. D. Site-Directed and Global Incorporation of Orthogonal and Isostructural Noncanonical Amino Acids into the Ribosomal Lasso Peptide Capistruin. *ChemBioChem* **2015**, *16* (3), 503–509. <https://doi.org/10.1002/cbic.201402558>.
- (8) Piscotta, F.; Tharp, J.; Liu, W.; James Link, A. Expanding the Chemical Diversity of Lasso Peptide MccJ25 with Genetically Encoded Noncanonical Amino Acids. *Chem. Commun.* **2015**, *51* (2), 409–412. <https://doi.org/10.1039/C4CC07778D>.
- (9) Zong, C.; Maksimov, M. O.; Link, A. J. Construction of Lasso Peptide Fusion Proteins. *ACS Chem. Biol.* **2016**, *11* (1), 61–68. <https://doi.org/10.1021/acschembio.5b00745>.
- (10) Burkhardt, 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* (8), 564–570. <https://doi.org/10.1038/ncchembio.1856>.
- (11) Zhu, S.; Fage, C. D.; Hegemann, J. D.; Mielcarek, A.; Yan, D.; Linne, U.; Marahiel, M. A. The B1 Protein Guides the Biosynthesis of a Lasso Peptide. *Sci. Rep.* **2016**, *6* (1), 1–12. <https://doi.org/10.1038/srep35604>.
- (12) Yan, K.-P.; Li, Y.; Zirah, S.; Goulard, C.; Knappe, T. A.; Marahiel, M. A.; Rebuffat, S. Dissecting the Maturation Steps of the Lasso Peptide Microcin J25 in Vitro. *ChemBioChem* **2012**, *13* (7), 1046–1052. <https://doi.org/10.1002/cbic.201200016>.
- (13) 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.* **2019**, *141* (1), 290–297. <https://doi.org/10.1021/jacs.8b09928>.

(14) Knappe, T. A.; Linne, U.; Xie, X.; Marahiel, M. A. The Glucagon Receptor Antagonist BI-32169 Constitutes a New Class of Lasso Peptides. *FEBS Lett.* **2010**, *584* (4), 785–789. <https://doi.org/10.1016/j.febslet.2009.12.046>.

(15) Gavrish, E.; Sit, C. S.; Cao, S.; Kandror, O.; Spoering, A.; Peoples, A.; Ling, L.; Fetterman, A.; Hughes, D.; Bissell, A.; Torrey, H.; Akopian, T.; Mueller, A.; Epstein, S.; Goldberg, A.; Clardy, J.; Lewis, K. Lassomycin, a Ribosomally Synthesized Cyclic Peptide, Kills Mycobacterium Tuberculosis by Targeting the ATP-Dependent Protease ClpC1P1P2. *Chem. Biol.* **2014**, *21* (4), 509–518. <https://doi.org/10.1016/j.chembiol.2014.01.014>.

(16) Xu, F.; Wu, Y.; Zhang, C.; Davis, K. M.; Moon, K.; Bushin, L. B.; Seyedsayamdst, M. R. A Genetics-Free Method for High-Throughput Discovery of Cryptic Microbial Metabolites. *Nat. Chem. Biol.* **2019**, *15* (2), 161–168. <https://doi.org/10.1038/s41589-018-0193-2>.

(17) Zong, C.; Ling Cheung-Lee, W.; E. Elashal, H.; Raj, M.; James Link, A. Albusnadin: An Acetylated Lasso Peptide from Streptomyces Albus. *Chem. Commun.* **2018**, *54* (11), 1339–1342. <https://doi.org/10.1039/C7CC08620B>.

(18) Ogawa, T.; Ochiai, K.; Tanaka, T.; Tsukuda, E.; Chiba, S.; Yano, K.; Yamasaki, M.; Yoshida, M.; Matsuda, Y. RES-701-2, -3 and -4, Novel and Selective Endothelin Type B Receptor Antagonists Produced by Streptomyces Sp. I. Taxonomy of Producing Strains, Fermentation, Isolation, and Biochemical Properties. *J. Antibiot. (Tokyo)* **1995**, *48* (11), 1213–1220. <https://doi.org/10.7164/antibiotics.48.1213>.

(19) Feng, Z.; Ogasawara, Y.; Nomura, S.; Dairi, T. Biosynthetic Gene Cluster of a D-Tryptophan-Containing Lasso Peptide, MS-271. *ChemBioChem* **2018**, *19* (19), 2045–2048. <https://doi.org/10.1002/cbic.201800315>.

(20) Zhu, S.; Hegemann, J. D.; Fage, C. D.; Zimmermann, M.; Xie, X.; Linne, U.; Marahiel, M. A. Insights into the Unique Phosphorylation of the Lasso Peptide Paeninodin. *J. Biol. Chem.* **2016**, *291* (26), 13662–13678. <https://doi.org/10.1074/jbc.M116.722108>.

(21) 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* (5), 470–478. <https://doi.org/10.1038/nchembio.2319>.

(22) Schwalen, C. J.; Hudson, G. A.; Kille, B.; Mitchell, D. A. Bioinformatic Expansion and Discovery of Thiopeptide Antibiotics. *J. Am. Chem. Soc.* **2018**, *140* (30), 9494–9501. <https://doi.org/10.1021/jacs.8b03896>.

(23) Walker, M. C.; Mitchell, D. A.; van der Donk, W. A. Precursor Peptide-Targeted Mining of More than One Hundred Thousand Genomes Expands the Lanthipeptide Natural Product Family. *bioRxiv* **2020**, 2020.03.13.990614. <https://doi.org/10.1101/2020.03.13.990614>.

(24) 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 C α , C β , and C γ -Linked Thioether-Containing Peptides. *J. Am. Chem. Soc.* **2019**, *141* (20), 8228–8238. <https://doi.org/10.1021/jacs.9b01519>.

(25) Gabarrini, G.; Smit, M. de; Westra, J.; Brouwer, E.; Vissink, A.; Zhou, K.; Rossen, J. W. A.; Stobernack, T.; Dijl, J. M. van; Winkelhoff, A. J. van. The Peptidylarginine Deiminase Gene Is a Conserved Feature of *Porphyromonas Gingivalis*. *Sci. Rep.* **2015**, *5* (1), 1–8. <https://doi.org/10.1038/srep13936>.

(26) Nielen, M. M. J.; Schaardenburg, D. van; Reesink, H. W.; Stadt, R. J. van de; Horst-Bruinsma, I. E. van der; Koning, M. H. M. T. de; Habibuw, M. R.; Vandenbroucke, J. P.; Dijkmans, B. A. C. Specific Autoantibodies Precede the Symptoms of Rheumatoid Arthritis: A Study of Serial Measurements in Blood Donors. *Arthritis Rheum.* **2004**, *50* (2), 380–386. <https://doi.org/10.1002/art.20018>.

(27) El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S. R.; Luciani, A.; Potter, S. C.; Qureshi, M.; Richardson, L. J.; Salazar, G. A.; Smart, A.; Sonnhammer, E. L. L.; Hirsh, L.; Paladin, L.; Piovesan,

D.; Tosatto, S. C. E.; Finn, R. D. The Pfam Protein Families Database in 2019. *Nucleic Acids Res.* **2019**, *47* (D1), D427–D432. <https://doi.org/10.1093/nar/gky995>.

(28) Fuhrmann, J.; Clancy, K. W.; Thompson, P. R. Chemical Biology of Protein Arginine Modifications in Epigenetic Regulation. *Chem. Rev.* **2015**, *115* (11), 5413–5461. <https://doi.org/10.1021/acs.chemrev.5b00003>.

(29) Clancy, K. W.; Weerapana, E.; Thompson, P. R. Detection and Identification of Protein Citrullination in Complex Biological Systems. *Curr. Opin. Chem. Biol.* **2016**, *30*, 1–6. <https://doi.org/10.1016/j.cbpa.2015.10.014>.

(30) Cox, C. L.; Tietz, J. I.; Sokolowski, K.; Melby, J. O.; Doroghazi, J. R.; Mitchell, D. A. Nucleophilic 1,4-Additions for Natural Product Discovery. *ACS Chem. Biol.* **2014**, *9* (9), 2014–2022. <https://doi.org/10.1021/cb500324n>.

(31) The genomic landscape of ribosomal peptides containing thiazole and oxazole heterocycles | BMC Genomics | Full Text <https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-2008-0> (accessed Apr 10, 2020).

(32) Maxson, T.; Tietz, J. I.; Hudson, G. A.; Guo, X. R.; Tai, H.-C.; Mitchell, D. A. Targeting Reactive Carbonyls for Identifying Natural Products and Their Biosynthetic Origins. *J. Am. Chem. Soc.* **2016**, *138* (46), 15157–15166. <https://doi.org/10.1021/jacs.6b06848>.

(33) Molloy, E. M.; Tietz, J. I.; Blair, P. M.; Mitchell, D. A. Biological Characterization of the Hygrobafilomycin Antibiotic JBIR-100 and Bioinformatic Insights into the Hygrolide Family of Natural Products. *Bioorg. Med. Chem.* **2016**, *24* (24), 6276–6290. <https://doi.org/10.1016/j.bmc.2016.05.021>.

(34) Castro-Falcón, G.; Hahn, D.; Reimer, D.; Hughes, C. C. Thiol Probes To Detect Electrophilic Natural Products Based on Their Mechanism of Action. *ACS Chem. Biol.* **2016**, *11* (8), 2328–2336. <https://doi.org/10.1021/acscchembio.5b00924>.

(35) Palaniappan, K. K.; Pitcher, A. A.; Smart, B. P.; Spiciarich, D. R.; Iavarone, A. T.; Bertozzi, C. R. Isotopic Signature Transfer and Mass Pattern Prediction (IsoStamp): An Enabling Technique for Chemically-Directed Proteomics. *ACS Chem. Biol.* **2011**, *6* (8), 829–836. <https://doi.org/10.1021/cb100338x>.

(36) Capehart, S. L.; Carlson, E. E. Mass Spectrometry-Based Assay for the Rapid Detection of Thiol-Containing Natural Products. *Chem. Commun.* **2016**, *52* (90), 13229–13232. <https://doi.org/10.1039/C6CC07111B>.

(37) Bicker, K. L.; Subramanian, V.; Chumanovich, A. A.; Hofseth, L. J.; Thompson, P. R. Seeing Citrulline: Development of a Phenylglyoxal-Based Probe To Visualize Protein Citrullination. *J. Am. Chem. Soc.* **2012**, *134* (41), 17015–17018. <https://doi.org/10.1021/ja308871v>.

(38) Choi, M.; Song, J.-S.; Kim, H.-J.; Cha, S.; Lee, E. Y. Matrix-Assisted Laser Desorption Ionization–Time of Flight Mass Spectrometry Identification of Peptide Citrullination Site Using Br Signature. *Anal. Biochem.* **2013**, *437* (1), 62–67. <https://doi.org/10.1016/j.ab.2013.03.003>.

(39) Vallenet, D.; Calteau, A.; Dubois, M.; Amours, P.; Bazin, A.; Beuvin, M.; Burlot, L.; Bussell, X.; Fouteau, S.; Gautreau, G.; Lajus, A.; Langlois, J.; Planel, R.; Roche, D.; Rollin, J.; Rouy, Z.; Sabatet, V.; Médigue, C. MicroScope: An Integrated Platform for the Annotation and Exploration of Microbial Gene Functions through Genomic, Pangenomic and Metabolic Comparative Analysis. *Nucleic Acids Res.* **2020**, *48* (D1), D579–D589. <https://doi.org/10.1093/nar/gkz926>.

(40) Tran, N.-V.; Greshake Tzovaras, B.; Ebersberger, I. PhyloProfile: Dynamic Visualization and Exploration of Multi-Layered Phylogenetic Profiles. *Bioinformatics* **2018**, *34* (17), 3041–3043. <https://doi.org/10.1093/bioinformatics/bty225>.

(41) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–3402. <https://doi.org/10.1093/nar/25.17.3389>.

(42) McDonald, J. H. *Handbook of Biological Statistics*, 3rd ed.; Sparky House Publishing: Baltimore, MD, 2014.

- (43) Bicker, K. L.; Thompson, P. R. The Protein Arginine Deiminases: Structure, Function, Inhibition, and Disease. *Biopolymers* **2013**, *99* (2), 155–163. <https://doi.org/10.1002/bip.22127>.
- (44) Shirai, H.; Mokrab, Y.; Mizuguchi, K. The Guanidino-Group Modifying Enzymes: Structural Basis for Their Diversity and Commonality. *Proteins Struct. Funct. Bioinforma.* **2006**, *64* (4), 1010–1023. <https://doi.org/10.1002/prot.20863>.
- (45) Linsky, T.; Fast, W. Mechanistic Similarity and Diversity among the Guanidine-Modifying Members of the Pentein Superfamily. *Biochim. Biophys. Acta BBA - Proteins Proteomics* **2010**, *1804* (10), 1943–1953. <https://doi.org/10.1016/j.bbapap.2010.07.016>.
- (46) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- (47) The EFI Web Resource for Genomic Enzymology Tools: Leveraging Protein, Genome, and Metagenome Databases to Discover Novel Enzymes and Metabolic Pathways. - Abstract - Europe PMC <https://europepmc.org/article/med/31553576> (accessed Apr 10, 2020).
- (48) Lawrence, J. G.; Hartl, D. L. Inference of Horizontal Genetic Transfer from Molecular Data: An Approach Using the Bootstrap. *Genetics* **1992**, *131* (3), 753–760.
- (49) Kohl, M.; Wiese, S.; Warscheid, B. Cytoscape: Software for Visualization and Analysis of Biological Networks. In *Data Mining in Proteomics: From Standards to Applications*; Hamacher, M., Eisenacher, M., Stephan, C., Eds.; Methods in Molecular Biology; Humana Press: Totowa, NJ, 2011; pp 291–303. https://doi.org/10.1007/978-1-60761-987-1_18.
- (50) Bierman, M.; Logan, R.; O'Brien, K.; Seno, E. T.; Nagaraja Rao, R.; Schoner, B. E. Plasmid Cloning Vectors for the Conjugal Transfer of DNA from Escherichia Coli to Streptomyces Spp. *Gene* **1992**, *116* (1), 43–49. [https://doi.org/10.1016/0378-1119\(92\)90627-2](https://doi.org/10.1016/0378-1119(92)90627-2).
- (51) Herrmann, S.; Siegl, T.; Luzhetska, M.; Petzke, L.; Jilg, C.; Welle, E.; Erb, A.; Leadlay, P. F.; Bechthold, A.; Luzhetsky, A. Site-Specific Recombination Strategies for Engineering Actinomycete Genomes. *Appl. Environ. Microbiol.* **2012**, *78* (6), 1804–1812. <https://doi.org/10.1128/AEM.06054-11>.
- (52) Bibb, M. J.; Janssen, G. R.; Ward, J. M. Cloning and Analysis of the Promoter Region of the Erythromycin Resistance Gene (ErmE) of Streptomyces Erythraeus. *Gene* **1985**, *38* (1), 215–226. [https://doi.org/10.1016/0378-1119\(85\)90220-3](https://doi.org/10.1016/0378-1119(85)90220-3).