Published in final edited form as:
Nat Chem Biol. 2015 August ; 11(8): 564–570. doi:10.1038/nchembio.1856.

A Prevalent Peptide-Binding Domain Guides Ribosomal Natural Product Biosynthesis

Brandon J. Burkhart[1,2], **Graham A. Hudson**[1,2], **Kyle L. Dunbar**[1,2], and **Douglas A. Mitchell**[1,2,3,*]

[1]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[2]Institute for Genomic Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

[3]Department of Microbiology, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA

Abstract

Ribosomally synthesized and posttranslationally modified peptides (RiPPs) are a rapidly growing natural product class. RiPP precursor peptides can undergo extensive enzymatic tailoring, yielding structurally and functionally diverse products, and their biosynthetic logic makes them attractive bioengineering targets. Recent work suggests that unrelated RiPP modifying enzymes contain structurally similar precursor peptide-binding domains. Using profile hidden Markov model comparisons, we discovered related and previously unrecognized peptide-binding domains in proteins spanning the majority of known prokaryotic RiPP classes; thus, we named this conserved domain the RiPP precursor peptide recognition element (RRE). Through binding studies, we verify the role of the RRE for three distinct RiPP classes: linear azole-containing peptides, thiopeptides, and lasso peptides. Because numerous RiPP biosynthetic enzymes act on peptide substrates, our findings have powerful predictive value as to which protein(s) drive substrate binding, laying a foundation for further characterization of RiPP biosynthetic pathways and the rational engineering of new peptide-binding activities.

Natural products, in particular polyketides and non-ribosomal peptides, have provided a wealth of pharmaceutically important molecules[1] and chemical probes for biology[2]. More recently, ribosomally synthesized and posttranslationally modified peptides (RiPPs) have been recognized as another major natural product sector[3] with a similar capacity for probing biological systems[4] and for serving as new drug scaffolds[5]. RiPPs can be exceedingly complex in structure and are further categorized based on the set of modifying enzymes encoded by their biosynthetic gene clusters and the corresponding structural features present

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html

[*]Correspondence and requests for materials should be addressed to: D.A.M., douglasm@illinois.edu, phone: 1-217-333-1345, fax: 1-217-333-0508.

Author contributions
Experiments were designed by D.A.M. and B.J.B. and performed by B.J.B. Critical reagents were provided by K.L.D. and G.A.H. The manuscript was written by D.A.M. and B.J.B. The overall study was conceived and managed by D.A.M.

Competing financial interests
The authors declare no competing financial interests.

Supplementary information is available online at http://www.nature.com/naturechemicalbiology/.

within the final products. Consequently, RiPP gene clusters are not identified or defined by a shared biosynthetic enzyme; rather, they are unified by a common logic of posttranslationally modifying an unrestrained, ribosomal peptide[3] (Fig. 1a). RiPP precursor peptides are usually bipartite, being composed of an *N*-terminal leader and *C*-terminal core regions. RiPP biosynthetic enzymes recognize conserved sequence motifs in the leader region and install residue-specific modifications to the core, yielding structurally complex compounds[3,6]. Because of the bipartite nature of RiPP precursor peptides, the modifying enzymes can be substrate specific while also displaying a high tolerance towards unnatural core sequences. Combined with their gene-encoded nature, RiPPs are attractive biosynthetic engineering targets[3,7–9]. Previous efforts relied on maintenance of native leader peptide sequences; omission of the conserved binding motifs significantly limits processing[3,10]. Although some residues of the leader peptide that are recognized by the biosynthetic enzymes have been identified[6], there are only a few RiPP classes where it is known which protein(s) physically associate with the leader and which residues of these proteins govern the interaction.

Our group and others have made progress towards the general mechanistic understanding of substrate binding and catalysis for the cyclodehydratase involved in thiazole/oxazole-modified microcin (TOMM) biosynthesis[11]. TOMM cyclodehydratases are found in >1,000 biosynthetic gene clusters[12] and span three RiPP classes including linear azole-containing peptides (LAPs), azole-containing cyanobactins, and thiopeptides[3,11]. Occurring as fused or discrete polypeptides[13–15], the TOMM cyclodehydratase consists of an adenosine 5′-triphosphate (ATP)-utilizing YcaO domain (D protein)[12,16,17] and a member of the E1 ubiquitin activating-like (E1-like) superfamily (C protein)[18]. In collaboration, the C/D protein(s) form azoline heterocycles from Cys, Ser, and Thr residues of the core peptide (Fig. 1b)[13–15]. Often a third component (B protein) is present, which is a flavin mononucleotide (FMN)-dependent dehydrogenase that oxidizes select azolines to azoles.

Originally enigmatic, work with the discrete TOMM cyclodehydratase from *Bacillus* sp. Al Hakam[19] (Balh) has revealed that BalhD catalyzes azoline formation[16] while BalhC protein mediates peptide substrate recognition[12]. Although the leader peptide-binding site remains unknown, the *N*-terminal winged helix-turn-helix (wHTH) domain has been implicated based on its homology to the “peptide clamp” of MccB[20], which is another E1-like RiPP modifying enzyme found in microcin C7 biosynthesis. Even though MccB and other related E1-like enzymes (*e.g.* PaaA and Rv3196) are not part of TOMM biosynthetic pathways, these E1-like enzymes all possess an *N*-terminal wHTH domain and produce ribosomal peptide-derived compounds[18], prompting the hypothesis that the wHTH domain guides the peptide to the active site. In support of this hypothesis, two recently solved structures, LynD (PDB: 4V1T, a fused cyclodehydratase; cyanobactin biosynthesis[14,21]) and NisB (PDB: 4WD9, lanthipeptide dehydratase; nisin biosynthesis[22]), show that their leader peptides are bound by a structurally similar wHTH motif. This suggests that additional, seemingly unrelated RiPP biosynthetic enzymes might employ the same strategy for substrate binding. However, the connection between these “recognition” domains and which RiPP enzymes directly engage the leader peptide remain unexplored.

Here, we bioinformatically identified a conserved domain linking the homologous wHTH motifs in NisB, LynD, and MccB. Through reevaluation of reported structures and experimental validation of additional RiPP enzymes, we demonstrated that this conserved domain is present in the majority of prokaryotic RiPP classes. Given its broad distribution and functional role, we named this domain the RiPP precursor peptide recognition element (RRE). These findings revealed previously unidentified similarities between disparate RiPP gene clusters and will guide future investigations of RiPP biosynthesis.

RESULTS

RREs are related to a PQQ biosynthetic protein

We began our investigation of RiPP precursor peptide recognition by searching for a link between the peptide-binding wHTH domains of NisB and LynD. Amino acid sequence homology was not detected using routine methods (*e.g.* BLAST[23]), likely due to high divergence and lack of similarity to any annotated domains in the Conserved Domain Database (CDD)[24]. However, HHpred[25], a highly sensitive homology detection tool based on profile hidden Markov model (HMM) comparisons[26], revealed that the wHTH domains of LynD (residues 2–81) and NisB (residues 153–223) were related to the protein PqqD (probability > 90%). HHpred probabilities give the most relevant representation of significance with > 90% usually being considered a true positive[25,27]. PqqD itself is a small protein (~90–100 residues) involved in the biosynthesis of another RiPP, pyrroloquinoline quinone (PQQ)[28]. PQQ is posttranslationally derived from Glu and Tyr residues of the PqqA precursor peptide. Although PQQ biosynthesis is not fully understood, PqqA maturation likely relies on PqqB (putative oxygenase), PqqC (oxygenase), PqqD ("peptide chaperone"), and PqqE (radical SAM)[28]. Prior to a very recent report of PqqD binding to PqqA[29], its function was unknown, but this finding is consistent with its homology to the peptide binding domains of NisB, LynD and MccB.

Because the structure of PqqD has been solved[30], we queried the Dali server[31] to assess structural similarity between these proteins as an additional verification of homology. Searches with MccB, LynD, and NisB did not return PqqD as a structural homolog; however, when submitting only the PqqD-like region of these proteins, LynD matched PqqD with a relatively weak similarity score (Z-score 5.3, with Z < 2.0 being viewed as spurious). The reverse search using PqqD as the query returned LynD and MccB with similar Z-scores (5.4 and 3.0, respectively). Visual comparison of the PqqD structure (PDB entry: 3G2B) with these proteins corroborates the HHpred/Dali alignment results (Fig. 2). Thus, structural homology to PqqD links the structurally similar domains of NisB, LynD, and MccB.

RREs are present in > 50% of prokaryotic RiPP classes

Building on the finding that at least three enzymatically unique RiPP biosynthetic enzymes harbor PqqD-like domains with distant homology, we hypothesized same could be true for additional RiPP biosynthetic proteins. Indeed, PqqD-like domains have been previously identified in AlbA[32] from subtilosin A (sactipeptide) biosynthesis and a protein of unknown function for lariatin (lasso peptide) biosynthesis[3,33]; however, the role of these domains remained largely enigmatic. We speculate that PqqD-like domains function as a RiPP precursor peptide recognition element (RRE). To determine the prevalence of such domains, a sequence similarity network of the ~4000 members of the PqqD InterPro family (IPR008792) was constructed at an E-value threshold of 10^{-16} (Supplementary Results, Supplementary Fig. 1). Analysis of the network revealed proteins from sactipeptide and lasso peptide biosynthesis, but homologs of lanthipeptide dehydratases and TOMM cyclodehydratases were absent (*i.e*. they are not members of IPR008792). As an alternative approach, we queried the Conserved Domain Architecture Retrieval Tool (CDART)[34] for proteins containing an annotated PqqD-like domain and found over 40 unique architectures (Supplementary Table 1). While some architectures were related to known RiPP biosynthetic proteins (*e.g.* AlbA), the few returned sequences did not appear to be RiPP related, indicating that CDART[34] was insufficient to identify RRE-containing proteins, which is likely a consequence of their general lack of annotation.

To overcome these shortcomings, we again returned to HHpred to search RiPP biosynthetic gene clusters for RREs (via PqqD homology). A representative gene cluster for previously defined RiPP classes[3] was chosen and each protein was individually queried using HHPred (Supplementary Table 2). Because eukaryotic RiPP biosynthesis is less bioinformatically tractable, the current analysis included only prokaryotes. Our results indicated that over half of all known prokaryotic RiPP classes harbor a PqqD-like domain, underscoring a broad distribution of RREs (Fig. 3; detailed results in Supplementary Table 3).

As described above, the RRE for canonical discrete (*e.g*. BalhC/D) and fused (*e.g*. LynD) cyclodehydratases is the *N*-terminal wHTH domain of the C protein portion, which itself is not annotated in the CDD[24]. However, we recently identified a third cyclodehydratase architecture, confined largely to the thiopeptides and heterocycloanthracins (Hca; LAP)[35,36]. These fused cyclodehydratases are *N*-terminally truncated by ~150 residues with the RRE being ablated. Almost invariably, TOMM clusters that contain “truncated” cyclodehydratases also encode a partner protein (annotated in the CDD as “ocin_ThiF_like”; hereafter, TOMM F protein), which contains the RRE (Fig. 3). As predicted, we have confirmed that the responsibility of binding the leader peptide resides with the RRE-containing TOMM F protein (*i.e*. HcaF) rather than the fused cyclodehydratase[36]. Consequently, we refer to these enzymes as F-dependent cyclodehydratases.

Via shared homology to the E1 ubiquitin-activating superfamily, the TOMM C protein is related to the microcin C7 adenylase MccB. The RRE present within MccB has previously been defined as a “peptide clamp” based on its interaction with the MccA peptide substrate[20]. The gene encoding MccA is one of the shortest known and produces 7 amino acid, leaderless peptide substrate[3]. Perhaps this explains why the MccA binding mode is somewhat different than that displayed by NisB and LynD (Fig. 2). Regardless, the MccB RRE interacts with the *N*-terminal residues of MccA and is essential for substrate binding. We predict the same for PaaA (pantocin biosynthesis[3]) owing to its similarity to MccB (microcin C7 biosynthesis)[18,37] (Fig. 3).

In class I lanthipeptides, the LanB and LanC enzymes (NisB and NisC, respectively for nisin biosynthesis) form the lanthionines. LanB enzymes are Ser/Thr dehydratases composed of a glutamylation and elimination domain, which occur as separate (“split”) polypeptides in thiopeptide gene clusters (*e.g.* PbtB and PbtC)[22]. Subsequently, LanC cyclases catalyze the Michael-like addition between the dehydrated Ser/Thr residues and particular Cys residues. In class II lanthipeptides, dehydration and cyclization are performed by a bifunctional synthetase composed of an *N*-terminal dehydratase domain (with no homology NisB) and a *C*-terminal LanC-like cyclase domain. Class III and IV lanthipeptide synthetases are trifunctional with an *N*-terminal lyase, central kinase, and variable *C*-terminal cyclase domains[3]. HHpred was unable to identify RREs in class II–IV lanthipeptide synthetases; however, RREs appear to be ubiquitous in class I LanB dehydratases and in their "split" counterparts in thiopeptide biosynthesis (Fig. 2c and Fig. 3).

While the majority of RREs are not explicitly annotated in public databases, the CDD identifies "PqqD" in a subset of lasso peptide gene clusters[33]. Canonical lasso peptide clusters harbor a precursor peptide, transglutaminase-like protease, asparagine synthetase-like protein, and an ABC transporter. In many recently discovered lasso peptide gene clusters, the protease appears to be encoded by two discrete, adjacent genes[38] (Supplementary Fig. 2). In these cases, the *C*-terminal portions are identified by their homology to transglutaminases (*e.g.* LarD, annotated as "Trans_glutcore_3", Fig. 3) and the *N*-terminal domain is annotated PqqD (*e.g.* LarC). Consistent with the hypothesis that the larger lasso proteases are fusions of the PqqD-like and transglutaminase-like proteins, the *N*-terminus of the fused protease involved in the biosynthesis of the lasso peptide astexin-1 (AtxB)[3] is similar to PqqD (probability > 90%). Surprisingly, some lasso peptide proteases (*e.g.* McjB, microcin J25) do not display detectable homology to PqqD (HHpred probability < 20%); however, because AtxB and McjB fulfill the same role in their respective gene clusters, we predict that McjB also possesses an RRE although HHpred cannot identify it.

Another example where RREs are annotated as PqqD-like are the sactipeptide radical SAM (rSAM) enzymes[32]. Sactipeptide rSAMs, (*e.g.* AlbA, subtilosin), install C_{α}-S thioether crosslinks and contain auxiliary [4Fe-4S] clusters within a *C*-terminal "SPASM" domain (named for its presence in <u>s</u>ubtilosin, <u>P</u>QQ, <u>a</u>naerobic <u>s</u>ulfatase, and <u>m</u>ycofactocin clusters)[39,40]. To investigate if RREs are prevalent among other rSAM/SPASM-containing gene clusters with short peptide substrates, the mycofactocin gene cluster (mft) from *Thermomicrobium roseum* was scanned with HHpred. Notable homology to PqqD (probability > 90%) was found in MftB, giving it a genetic arrangement similar to PqqD/E (Fig. 3). We also detected RREs in two additional small mycofactocin-like gene clusters[40]: the SCIFF (six cysteines in forty-five residues) and KxxxW rSAM families[41]. Consistent with the proposed role of the RRE in RiPP biosynthesis, anaerobic sulfatase-maturating enzymes (anSMEs) lack bioinformatically or structurally (PDB entry: 4K39)[42] detectable RREs and are not associated with RiPP biosynthesis. These assignments are in agreement with a recent independent report, which identified PqqD-like domains in rSAMs[29]. Additional examples that further implicate RREs with RiPP rSAM enzymes include two enzymes involved in proteusin biosynthesis. RREs were located in the *C*-termini of the rSAM epimerase (*e.g.* PoyD) and the B12-dependent methyltransferase (PoyB) (Fig. 3)[3]. PoyD has been characterized and is known to be leader peptide-dependent[43], as would be expected for an RRE-bearing protein. Although a proteusin B12-dependent methyltransferase has not been reconstituted, we hypothesize that it will similarly be a leader peptide-dependent enzyme.

Lastly, we identified RREs in proteins involved in bottromycin[3] and trifolitoxin biosynthesis[44,45]. These two gene clusters are characterized by the presence of a discrete TOMM D protein[12], which apparently acts in the absence of a TOMM C protein. Intriguingly, these "stand-alone" D proteins do not bear a discernable RRE; however, HHpred found an RRE (probability > 90%) in the *N*-terminus of two trifolitoxin oxidoreductases (TfxB and TfxC) and the *C*-terminus of three bottromycin rSAMs (BmbB, BmbF, and BmbJ), the latter of which is architecturally similar to the proteusin *C*-terminal methyltransferases (Fig. 3). Unusually, the bottromycin precursor peptide harbors a "follower" peptide, rather than a leader peptide[3]. Despite this unconventional arrangement, we predict that the follower peptide of bottromycin would be engaged by the rSAM RRE.

Validation of bioinformatically predicted RREs

Our findings have revealed a broad distribution of bioinformatically-identified RREs among various RiPP classes and enzymes with strong support for its role gleaned from the substrate-bound structures of the LynD cyclodehydratase, NisB dehydratase[22], and MccB adenylase (MccB)[20]. To provide further evidence that predicted RREs recruit precursor peptides during RiPP biosynthesis, we corroborated the function of additional RREs through site-directed mutagenesis and fluorescence polarization (FP) binding assays. We surmised that specific residues within the RRE would engage the peptide and thus employed site-directed mutagenesis to explore the nature of the interaction in five proteins deriving from three additional RiPP classes (LAPs, thiopeptides, and lasso peptides). Based on analysis of the known structures, in addition to sequence alignments and homology models, the targeted area for mutagenesis was focused on residues in β1–β3 and α1–3 (Fig. 2, Supplementary Fig. 3). The purity and integrity of each mutant protein was visually evaluated by Coomassie-stained SDS-PAGE gel (Supplementary Fig. 4).

LAP discrete cyclodehydratase

With some exceptions, the discrete cyclodehydratases (separate C and D proteins) tend to be found within LAP clusters[13]. Having previously narrowed leader peptide binding to the C protein using an FP assay[12], we installed alanine substitutions to the BalhC RRE and compared binding constants using a fluorescently-labeled BalhA1 leader peptide (FITC-BalhA1-LP). The resulting data showed that most Ala substitutions within the RRE substantially (>10 fold) reduced the interaction between BalhA1 and BalhC (Table 1; Supplementary Fig. 3). Three BalhC mutations (K20A, D23A, and E66A) had a moderate effect while two mutations (D32A and H38A) had a minor effect. For comparison, residues outside of the RRE were also substituted. These had essentially no effect on binding except for those that appeared to weaken the structural stability of BalhC, rendering it more susceptible to proteolytic degradation during expression (Supplementary Table 4; Supplementary Fig. 4).

Because an extended region of the BalhC RRE was sensitive to substitution, we examined a second LAP cyclodehydratase from *Corynebacterium urealyticum* DSM 7109 (Cur), which is a member of the plantazolicin subclass[46] (Supplementary Fig. 5). When CurC was tested for binding to the FITC-labeled CurA leader (FITC-CurA-LP), a dissociation constant (K_d) of 7 ± 1 μM was obtained. CurD did not have an appreciable binding response, nor did it enhance CurA binding by CurC (Supplementary Fig. 6). This result is consistent with previous studies with BalhC[12] and C proteins involved in LAP cytolysin biosynthesis[10]. In order to assess the importance of the putative RRE in CurA binding, various residues were substituted with alanine. The binding data indicated that the CurC RRE was not as sensitive as BalhC to mutation, as only CurC^{Y31A} in β3 and CurC^{R68A} and CurC^{I75A} in α3 resulted in substantial (≥10 fold) reduction in affinity (Table 1; Supplementary Fig. 3). The dissociation constants for peptide binding for the remaining substitutions in β1, β2, and α1 were within two-fold of CurC^{WT}. Collectively, these data indicate that the RREs of discrete cyclodehydratases, especially residues lining β3 and α3, are essential for peptide binding and suggest a binding mode similar to LynD/NisB.

F protein-dependent cyclodehydratase

Recent work from our group[36] has revealed the dependence of truncated cyclodehydratases found in thiopeptide[3] and heterocycloanthracin (Hca) gene clusters[35] on a newly recognized biosynthetic protein (TOMM F protein). For these “F-dependent” cyclodehydratases, it has been demonstrated that the F protein (HcaF) binds the precursor peptide (HcaA), but the involvement of the RRE remained undetermined. Therefore, we probed the HcaF protein from *Bacillus* sp. Al Hakam by substituting numerous residues of the RRE with Ala. The results indicate that residues lining the groove between β3 and α3 (Phe35, Arg73, Ile80) were important for binding (Table 2; Supplementary Fig. 3). Mutation of nearby residues (HcaF^{D38A}, HcaF^{N72A}, and HcaF^{E79A}) had minimal effects on FITC-labeled HcaA leader peptide (FITC-HcaA-LP) affinity, as they are likely oriented away from the probable binding groove owing to their relative positioning. Asp38 is on the opposite side of β3 compared to Phe35 while Asn72 and Glu79 are orthogonally oriented relative to Arg73 and Ile80, respectively. Given that HcaF^{D21A} (β1) also had nearly WT-like binding, HcaF most probably engages HcaA in a manner analogous to LynD/NisB.

With HcaF being a LAP-related protein, we chose to examine a non-LAP F-dependent cyclodehydratase. *Thermobispora bispora* (Tbt) is bioinformatically predicted to produce a GE2270A-like (thiopeptide) compound and, like Hca, harbors an F-dependent cyclodehydratase (TbtG)[47]. Initial experiments with TbtF (F protein) demonstrated binding to FITC-labeled TbtA leader peptide (FITC-TbtA-LP) while TbtG had no discernable interaction with the peptide (Supplementary Fig. 7). Mutation of the TbtF RRE gave results similar to HcaF, with residues in β3 and the *N*-terminal portion of α3 being important for binding (Val30, Phe68, and Arg71). TbtF^{D19A} and TbtF^{L24A} in β2 had no change in binding compared to Tbt^{WT}, suggesting that the peptide binds between β3 and α3. Based on comparison with LynD, the side chains of Gln74 and Arg79 are not expected to make significant contact with the peptide, and consistent with this model, substitution of these residues had no measurable effect on the affinity towards FITC-TbtA-LP (Table 2; Supplementary Fig. 3).

Lasso peptide split protease

With the role of the RRE established for TOMM cyclodehydratases, we next turned to the lasso peptides to test a RiPP class where little is known about precursor peptide recognition. Recently, our group reported on an unusual lasso peptide, streptomonomicin (STM), from *Streptomonospora alba*[48]. Genome sequencing revealed that the STM gene cluster contained a "split" protease (StmE and StmB; Supplementary Fig. 2). Initial binding studies using size exclusion chromatography (SEC) indicated that the RRE-containing StmE bound the StmA precursor peptide, while the separate protease domain (StmB) had no interaction with either StmA or StmE under the tested conditions and appeared substantially degraded after expression (Supplementary Fig. 4 and 8). The interaction between StmA and StmE was further corroborated by an FP binding experiment with FITC-labeled StmA leader peptide (FITC-StmA-LP; Supplementary Fig. 9). Encouraged by the correct functional prediction for StmE, we prepared eight Ala substituted proteins and evaluated binding to FITC-StmA-LP. In accord with earlier examples, substitution of residues along the side of α3 nearest to β3 (D69A and L73A) resulted in dramatically (>500-fold) reduced peptide binding while Ala replacements of the adjacent Asp68 and Asp75, with their relative positions on the opposite side of α3, had no effect (Table 3). Unlike previous examples, StmC^{V12A} and StmC^{N22A} in β1 and β2, respectively, also displayed reduced binding while StmE^{Y28A} and StmE^{Q76A}, which reside near the groove between β3 and α3, displayed nearly WT affinity. While these homology model-based/mutant protein-binding results imply that lasso peptide proteases employ a slightly different subset of RRE residues to bind the leader peptide, the fact remains that irrespective of RiPP class or the protein's function, the RRE is implicated in governing substrate recognition.

DISCUSSION

In this study, we implicate a conserved PqqD-like domain in precursor peptide recognition in over half of all currently known RiPP classes (reflecting multiple thousands of gene clusters; Supplementary Table 2 and 3). Leveraging highly sensitive bioinformatics, known crystal structures (Fig. 2), and FP binding assays, the role of the RRE in leader peptide binding has substantial experimental support. Further, albeit indirect, evidence for the importance of RREs derive from two additional observations (*i*) RREs are routinely identified in RiPP biosynthetic proteins but are lacking in homologs not involved in processing RiPPs (*e.g.* RiPP rSAMs, E1-like enzymes, *vide supra*) and (*ii*) RiPP enzymes which act after leader peptide cleavage do not harbor RREs (e.g. methyltransferases involved in PZN[49] and linaridin[3] biosynthesis). Based on these findings, we predict that PqqD-like domains associated with peptide-modifying enzymes will function as RREs whenever present, as has been just reported for *bona fide* PqqD[29].

In addition to demonstrating the role of RREs, our work also provided insight into how RREs bind their respective peptides. Binding experiments with BalhC (LAP) indicated that nearly all of the RRE was affected by mutation. Accordingly, the precise orientation of BalhA1 could not be inferred; however, CurC (52% similar), which displays considerably improved solution-phase behavior, had reduced peptide-binding affinity only upon substitution of residues in β3 and α3. Given that β3 and α3 were the primary affected regions for TOMM F proteins (*i.e.* HcaF and TbtF), and that LynD interacts with its peptide substrate using the same strategy (Supplementary Fig. 3), TOMM cyclodehydratases appear to bind their substrates in a similar orientation irrespective of domain organization. However, there are other RREs that interact with the peptide substrate in differing orientations (*e.g.* MccB). Perhaps this should be unsurprising, since RREs are as divergent as the proteins and pathways they are found within. Other sequence divergent, but structurally similar, peptide-binding domains also have varied binding modes[50]. Logically, RREs must bind their substrates in a manner conducive to posttranslational modification; therefore, proper presentation of the peptide to the modifying enzymes is essential. Perhaps this explains why residues in β1 and β2 of the StmE RRE were important for binding, owing to the mechanics necessary for forming the lasso topology. It is important to note, however, that bioinformatics and binding assays alone do not reveal atomic-level details of the substrate-binding interactions.

The real power of our findings resides in the ability to accurately predict what protein(s) will recruit the precursor peptide to the modifying enzymes during RiPP biosynthesis, even for proteins of unknown function and pathways that require numerous enzymes. As a case in point, the identification of an RRE within a small protein (StmE) implicated in lasso peptide biosynthesis allowed us to assign a function where none had previously been proposed[3,38]. Similarly, the biosynthetic logic and inclusion of an “ocin_ThiF_like” (TOMM F) protein of certain TOMM biosynthetic gene clusters can be understood through identification of the RRE-containing protein. However, such insights are limited to systems where RREs are detectable, and accordingly, leader recognition remains enigmatic in many RiPP classes (Supplementary Table 2). It is possible that as homology detection algorithms become more sensitive and more sequences become available for building HMMs, additional RREs will be found. Another possibility is that RiPP pathways lacking identifiable RREs have evolved other peptide recognition modules which are unrelated to the RRE in both sequence and structure. Nonetheless, RREs appear to be the most prevalent solution to precursor peptide recognition among prokaryotic RiPPs and they functionally link a diverse array of biosynthetic platforms and enzymes (Fig. 3).

These findings suggest several future applications, including the engineering of new precursor peptide specificities via RRE swapping and utilizing the RRE as a marker to discover new RiPP classes. Indeed, it is difficult to predict the peptide substrates and modifying enzymes of a novel RiPP class *a priori*[3], which is why most first-in-class RiPPs have been found through phenotypic screens. However, if an RRE is bioinformatically-identified in a novel genomic context, the flanking regions could be scanned for potential biosynthetic proteins and precursor peptides. As a starting point, we note that many PqqD homologs found herein belong to uncharacterized gene clusters and appear ripe for the above approach (Supplementary Fig. 1 and Supplementary Table 1).

Overall, this work advances our understanding of RiPP biosynthesis and should serve to guide future studies. More work will be required to unequivocally establish the function of RREs in additional RiPP biosynthetic pathways, but importantly, our study lays a foundation on which proteins to prioritize such efforts.

ONLINE METHODS

General methods

All materials were purchased from Fisher Scientific or Sigma-Aldrich unless indicated otherwise. DNA sequencing was performed by the Roy J. Carver Biotechnology Center (University of Illinois at Urbana-Champaign) or ACGT Inc. DNA oligonucleotides were purchased from Integrated DNA Technologies (IDT). All fluorescently-labeled leader peptides were purchased from GenScript (>90% purity) with an *N*-terminal FITC-Ahx (fluorescein isothiocyanate with an amino hexyl linker) conjugated to a single glycine spacer before the leader peptide, except for the FITC-labeled leader peptides for HcaA and TbtA, which were prepared as previously described[36]. Full sequences are given in Supplementary Table 5.

Cloning

The genes encoding CurC/D and StmA/B/E were amplified from the genomic DNA of their respective host organisms. PCR was performed with Platinum Taq DNA Polymerase High Fidelity (Invitrogen) or Pfu DNA polymerase using the primers listed in Supplementary Table 6. BamHI-HF, NotI-HF, SalI, or HindIII (New England Biolabs, Inc., NEB) restriction sites were designed into these primers to avoid internal restriction enzyme cut sites. The amplified genes were PCR purified using DNA-binding spin columns (Qiagen, following manufacture's instruction) and digested with appropriate restriction enzymes. After an additional PCR purification step, the digested inserts were ligated with T4 DNA ligase (NEB) into similarly digested and purified pET28 vector in frame with an *N*-terminal, maltose-binding protein (MBP) tag. Codon optimized TbtF and TbtG were synthesized by Genscript with 5′ BamHI and 3′ XhoI cut sites, and after digestion with BamHI and XhoI, the excised genes were purified from a 1% agarose gel using a gel extraction kit (Qiagen). His_6-tagged HcaF was previously generated[36].

Site-directed mutagenesis

Site-directed mutagenesis was carried out using the Quikchange (Agilent) method or a modified method where primers were designed to overlap ~11 bp upstream and ~27 bp downstream of the targeted codon for mutation (to minimize primer-primer annealing). All mutagenesis primers are given in Supplementary Table 6.

Protein expression and purification

All proteins were expressed and purified as tobacco etch virus (TEV) protease-cleavable fusions with MBP as previously described[16] except for HcaF and its mutants which were purified by a N-terminal His6 tag[36]. The purity of all expressed proteins was examined by SDS-PAGE (Supplementary Fig. 4)

Fluorescence polarization (FP) binding assay

The interaction between fluorescently-labeled leader peptides and various RiPP biosynthetic proteins was quantified using an FP assay. To maximize the polarization signal, all proteins remained MBP-tagged. In general, protein was serially diluted into binding buffer (50 mM

HEPES, pH 7.5, 300 mM NaCl, 2.5% (*v/v*) glycerol, 0.5 mM TCEP) and mixed with 25 nM of the appropriate fluorescently-labeled leader peptide. Binding assays were carried out in non-binding-surface, 384-black-well polystyrene microplates (Corning) and measured using a FilterMax F5 multi-mode microplate reader (Molecular Devices) with λ_{ex} = 485 nm and λ_{em} = 538 nm. Prior to measurement, the dilutions were equilibrated with shaking for 15 to 30 min at 23 °C. Dissociation constant (K_d) values were calculated from the 50% saturation point using a dose-response curve fit in Origin Pro 9.1 with three independent titrations. Background fluorescence from the proteins alone was subtracted from the fluorescence polarization signal obtained with the fluorophore.

Sequence alignments

Initial sequence alignments of proteins within the same RiPP class were made using ClustalW2 using default parameters[51]. Due to divergence between different RiPP classes, homology model sequence alignments were generated using HHpred as described below. (http://toolkit.tuebingen.mpg.de/hhpred)[25,26].

Detection of protein homologs

BLAST[23] was used to identify closely related homologs of proteins. To identify more distantly related, yet homologous, regions of RiPP biosynthetic proteins, we employed HHpred[25], a sensitive homology detection tool based on profile hidden Markov model (HMM) comparisons[26]. HHpred queries were performed using the pdb70_30Apr15 database. The multiple sequence alignment (MSA) for the input sequence was generated with HHblits (iterative HMM-HMM comparison, maximum of 3 iterations), and secondary structure was scored with local alignment mode. All other options were left on their default settings except the “max number of hits in hit list” was increased to 500 so all hits above the 20% probability threshold would be returned. The hit list was then searched for PqqD or matches to the *N*-terminus of MccB, LynD, or NisB; however, we discovered that PqqD was a superior marker of RREs as the others frequently had lower probability scores if they appeared at all in the results. Only hits ≥ 90% probability were considered true hits. Because the MSA is dependent on the input sequence and can affect the probability score, all discovered PqqD-like domains were verified by re-submitting smaller sections of the input sequence, centered on the identified PqqD-like region. In practice, this was only necessary for the RiPP methyltransferases because the HHblits MSA generation tended to include non-RiPP rSAM enzymes (due to the similarity of their iron-sulfur binding domains), diluting the PqqD signature in the HMM. A representative gene cluster for each RiPP class (where the DNA sequence was available) was chosen from a comprehensive review of known RiPPs[3], and the biosynthetic proteins were individually assessed for homology with PqqD (Supplementary Table 2).

Homology model generation

Models were created using the HHpred interface, using LynD (PDB entry: 4V1T) as the manually selected template. In this method, the HHpred based alignment is automatically used as input for MODELLER[52,53] to generate a 3D model.

Sequence similarity network

A sequence similarity network of InterPro family IPR008792 was generated with the Enzyme Function Initiative Enzyme (EFI) Similarity Tool (http://www.enzymefunction.org/)[54]. The resulting network was visualized using the organic layout of Cytoscape[55], with an e-value threshold of 10^{-16}, and each node represents sequences with 80% or greater sequence identity. The function of different clusters with at least 3 nodes was investigated using the EFI Genome Neighborhood Tool[56]. A neighborhood size of 7 was chosen with 20% co-occurrence limit. Most nodes appeared to be part of uncharacterized gene clusters, but some lasso peptide gene clusters were identified by to co-occurrence with “Transglut_core3” and “Asn_synthase”.

Size-exclusion chromatography

Size-exclusion chromatography (SEC) was performed on a Flexar HPLC (Perkin Elmer) with analytical Yarra SEC-3000 (300 × 4.6 mm, Phenomenex). The column was pre-equilibrated with 5 column volumes (CVs) of binding buffer (50 mM HEPES, pH 7.5, 300 mM NaCl, 2.5% (*v/v*) glycerol) at 4 °C. Protein samples (50 μM) were made up on ice in the same buffer, injected (5 μL), and monitored at 280 nm. Traces were exported to Microsoft Excel for analysis and normalized to the highest absorbance value. Oligomeric state and apparent mass were determined based on a standard curve generated from the elution times of molecular weight standards 12–200 kDa (Sigma).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to A. Maniak, and R. Dowen for cloning several of the TOMM biosynthetic proteins and to C. Cox for bioinformatics consultation. We also thank C. Deane for critical review of this manuscript. This work was supported by the US National Institutes of Health (NIH) (1R01 GM097142 to D.A.M. and 2T32 GM070421 to B.J.B. and K.L.D). Additional financial support came from the UIUC Department of Chemistry (Robert C. and Carolyn J. Springborn Endowment to B.J.B. and Harold R. Snyder Fellowship to K.L.D. and G.A.H.). B.J.B was also funded by a National Science Foundation Graduate Research Fellowship (DGE-1144245).

References

1. Newman DJ, Cragg GM. Natural products as sources of new drugs over the 30 years from 1981 to 2010. J Nat Prod. 2012; 75:311–35. [PubMed: 22316239]
2. Carlson EE. Natural products as chemical probes. ACS Chem Biol. 2010; 5:639–53. [PubMed: 20509672]
3. Arnison PG, et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat Prod Rep. 2013; 30:108–60. [PubMed: 23165928]
4. Bindman NA, van der Donk WA. A general method for fluorescent labeling of the N-termini of lanthipeptides and its application to visualize their cellular localization. J Am Chem Soc. 2013; 135:10362–71. [PubMed: 23789944]
5. Cotter PD, Ross RP, Hill C. Bacteriocins – a viable alternative to antibiotics? Nat Rev Micro. 2013; 11:95–105.
6. Oman TJ, van der Donk WA. Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat Chem Biol. 2010; 6:9–18. [PubMed: 20016494]

7. Ruffner DE, Schmidt EW, Heemstra JR. Assessing the Combinatorial Potential of the RiPP Cyanobactin tru Pathway. ACS Synth Biol. 2015; 4:482–92. [PubMed: 25140729]

8. Goto Y, Ito Y, Kato Y, Tsunoda S, Suga H. One-pot synthesis of azoline-containing peptides in a cell-free translation system integrated with a posttranslational cyclodehydratase. Chem Biol. 2014; 21:766–774. [PubMed: 24856821]

9. Deane CD, Melby JO, Molohon KJ, Susarrey AR, Mitchell DA. Engineering unnatural variants of plantazolicin through codon reprogramming. ACS Chem Biol. 2013; 8:1998–2008. [PubMed: 23823732]

10. Mitchell DA, et al. Structural and functional dissection of the heterocyclic peptide cytotoxin. J Biol Chem. 2009; 284:13004–12. [PubMed: 19286651]

11. Melby JO, Nard NJ, Mitchell DA. Thiazole/oxazole-modified microcins: complex natural products from ribosomal templates. Curr Opin Chem Biol. 2011; 15:369–78. [PubMed: 21429787]

12. Dunbar KL, et al. Discovery of a new ATP-binding motif involved in peptidic azoline biosynthesis. Nat Chem Biol. 2014; 10:823–9. [PubMed: 25129028]

13. Lee SW, et al. Discovery of a widely distributed toxin biosynthetic gene cluster. Proc Natl Acad Sci USA. 2008; 105:5879–84. [PubMed: 18375757]

14. Schmidt EW, et al. Patellamide A and C biosynthesis by a microcin-like pathway in Prochloron didemni, the cyanobacterial symbiont of Lissoclinum patella. Proc Natl Acad Sci USA. 2005; 102:7315–7320. [PubMed: 15883371]

15. Li YM, Milne JC, Madison LL, Kolter R, Walsh CT. From peptide precursors to oxazole and thiazole-containing peptide antibiotics: microcin B17 synthase. Science. 1996; 274:1188–1193. [PubMed: 8895467]

16. Dunbar KL, Melby JO, Mitchell DA. YcaO domains use ATP to activate amide backbones during peptide cyclodehydrations. Nat Chem Biol. 2012; 8:569–75. [PubMed: 22522320]

17. McIntosh JA, Schmidt EW. Marine molecular machines: heterocyclization in cyanobactin biosynthesis. Chembiochem. 2010; 11:1413–21. [PubMed: 20540059]

18. Burroughs AM, Iyer LM, Aravind L. Natural history of the E1-like superfamily: implication for adenylation, sulfur transfer, and ubiquitin conjugation. Proteins. 2009; 75:895–910. [PubMed: 19089947]

19. Melby JO, Dunbar KL, Trinh NQ, Mitchell DA. Selectivity, directionality, and promiscuity in peptide processing from a Bacillus sp Al Hakam cyclodehydratase. J Am Chem Soc. 2012; 134:5309–16. [PubMed: 22401305]

20. Regni CA, et al. How the MccB bacterial ancestor of ubiquitin E1 initiates biosynthesis of the microcin C7 antibiotic. EMBO J. 2009; 28:1953–64. [PubMed: 19494832]

21. McIntosh JA, Lin Z, Tianero MD, Schmidt EW. Aestuaramides, a natural library of cyanobactin cyclic peptides resulting from isoprene-derived Claisen rearrangements. ACS Chem Biol. 2013; 8:877–83. [PubMed: 23411099]

22. Ortega MA, et al. Structure and mechanism of the tRNA-dependent lantibiotic dehydratase NisB. Nature. 2015; 517:509–12. [PubMed: 25363770]

23. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990; 215:403–410. [PubMed: 2231712]

24. Marchler-Bauer A, et al. CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res. 2013; 41:D348–52. [PubMed: 23197659]

25. Soding J, Biegert A, Lupas AN. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res. 2005; 33:W244–8. [PubMed: 15980461]

26. Soding J. Protein homology detection by HMM-HMM comparison. Bioinformatics. 2005; 21:951–60. [PubMed: 15531603]

27. Lopes A, Amarir-Bouhram J, Faure G, Petit MA, Guerois R. Detection of novel recombinases in bacteriophage genomes unveils Rad52, Rad51 and Gp2.5 remote homologs. Nucleic Acids Res. 2010; 38:3952–62. [PubMed: 20194117]

28. Klinman JP, Bonnot F. Intrigues and intricacies of the biosynthetic pathways for the enzymatic quinocofactors: PQQ, TTQ, CTQ, TPQ, and LTQ. Chem Rev. 2014; 114:4343–65. [PubMed: 24350630]

29. Latham JA, Iavarone AT, Barr I, Juthani PV, Klinman JP. 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. 2015; 290:12908–18. [PubMed: 25817994]

30. Tsai TY, Yang CY, Shih HL, Wang AH, Chou SH. Xanthomonas campestris PqqD in the pyrroloquinoline quinone biosynthesis operon adopts a novel saddle-like fold that possibly serves as a PQQ carrier. Proteins. 2009; 76:1042–8. [PubMed: 19475705]

31. Holm L, Rosenstrom P. Dali server: conservation mapping in 3D. Nucleic Acids Res. 2010; 38:W545–9. [PubMed: 20457744]

32. Wecksler SR, et al. Interaction of PqqE and PqqD in the pyrroloquinoline quinone (PQQ) biosynthetic pathway links PqqD to the radical SAM superfamily. Chem Commun (Camb). 2010; 46:7031–3. [PubMed: 20737074]

33. Li, Y.; Zirah, S.; Rebuffat, S. Lasso Peptides. Springer; New York: 2015. Biosynthesis, Regulation and Export of Lasso Peptides; p. 81-95.

34. Geer LY, Domrachev M, Lipman DJ, Bryant SH. CDART: protein homology by domain architecture. Genome Res. 2002; 12:1619–23. [PubMed: 12368255]

35. Haft DH. A strain-variable bacteriocin in Bacillus anthracis and Bacillus cereus with repeated Cys-Xaa-Xaa motifs. Biol Direct. 2009; 4:15. [PubMed: 19383135]

36. Dunbar KL, Tietz JI, Cox CL, Burkhart BJ, Mitchell DA. Identification of an Auxiliary Leader Peptide-Binding Protein Required for Azoline Formation in Ribosomal Natural Products. J Am Chem Soc. 2015 in press.

37. Bantysh O, et al. Enzymatic synthesis of bioinformatically predicted microcin C-like compounds encoded by diverse bacteria. MBio. 2014; 5:e01059–14. [PubMed: 24803518]

38. Hegemann JD, Zimmermann M, Zhu S, Klug D, Marahiel MA. Lasso peptides from proteobacteria: Genome mining employing heterologous expression and mass spectrometry. Biopolymers. 2013; 100:527–42. [PubMed: 23897438]

39. Haft DH. Bioinformatic evidence for a widely distributed, ribosomally produced electron carrier precursor, its maturation proteins, and its nicotinoprotein redox partners. BMC Genomics. 2011; 12:21. [PubMed: 21223593]

40. Haft DH, Basu MK. Biological systems discovery in silico: radical S-adenosylmethionine protein families and their target peptides for posttranslational modification. J Bacteriol. 2011; 193:2745–55. [PubMed: 21478363]

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

42. Goldman PJ, et al. X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification. Proc Natl Acad Sci USA. 2013; 110:8519–24. [PubMed: 23650368]

43. Morinaka BI, et al. Radical S-adenosyl methionine epimerases: regioselective introduction of diverse D-amino acid patterns into peptide natural products. Angew Chem Int Ed Engl. 2014; 53:8503–7. [PubMed: 24943072]

44. Breil BT, Ludden PW, Triplett EW. DNA sequence and mutational analysis of genes involved in the production and resistance of the antibiotic peptide trifolitoxin. J Bacteriol. 1993; 175:3693–702. [PubMed: 8509324]

45. Breil B, Borneman J, Triplett EW. A newly discovered gene, tfuA, involved in the production of the ribosomally synthesized peptide antibiotic trifolitoxin. J Bacteriol. 1996; 178:4150–6. [PubMed: 8763943]

46. Molohon KJ, et al. Structure determination and interception of biosynthetic intermediates for the plantazolicin class of highly discriminating antibiotics. ACS Chem Biol. 2011; 6:1307–13. [PubMed: 21950656]

47. Morris RP, et al. Ribosomally synthesized thiopeptide antibiotics targeting elongation factor Tu. J Am Chem Soc. 2009; 131:5946–55. [PubMed: 19338336]

48. Metelev M, et al. Structure, bioactivity, and resistance mechanism of streptomonomicin, an unusual lasso Peptide from an understudied halophilic actinomycete. Chem Biol. 2015; 22:241–50. [PubMed: 25601074]

49. Lee J, et al. Structural and functional insight into an unexpectedly selective N-methyltransferase involved in plantazolicin biosynthesis. Proc Natl Acad Sci USA. 2013; 110:12954–9. [PubMed: 23878226]

50. Fedorov AA, Fedorov E, Gertler F, Almo SC. Structure of EVH1, a novel proline-rich ligand-binding module involved in cytoskeletal dynamics and neural function. Nat Struct Mol Biol. 1999; 6:661–665.

51. Sievers F, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011; 7:539. [PubMed: 21988835]

52. Sali A, Blundell TL. Comparative protein modelling by satisfaction of spatial restraints. J Mol Biol. 1993; 234:779–815. [PubMed: 8254673]

53. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M. Evaluation of comparative protein modeling by MODELLER. Proteins. 1995; 23:318–26. [PubMed: 8710825]

54. Atkinson HJ, Morris JH, Ferrin TE, Babbitt PC. Using Sequence Similarity Networks for Visualization of Relationships Across Diverse Protein Superfamilies. PLoS ONE. 2009; 4:e4345. [PubMed: 19190775]

55. Shannon P, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504. [PubMed: 14597658]

56. Zhao S, et al. Prediction and characterization of enzymatic activities guided by sequence similarity and genome neighborhood networks. Elife. 2014; 3:e03275.

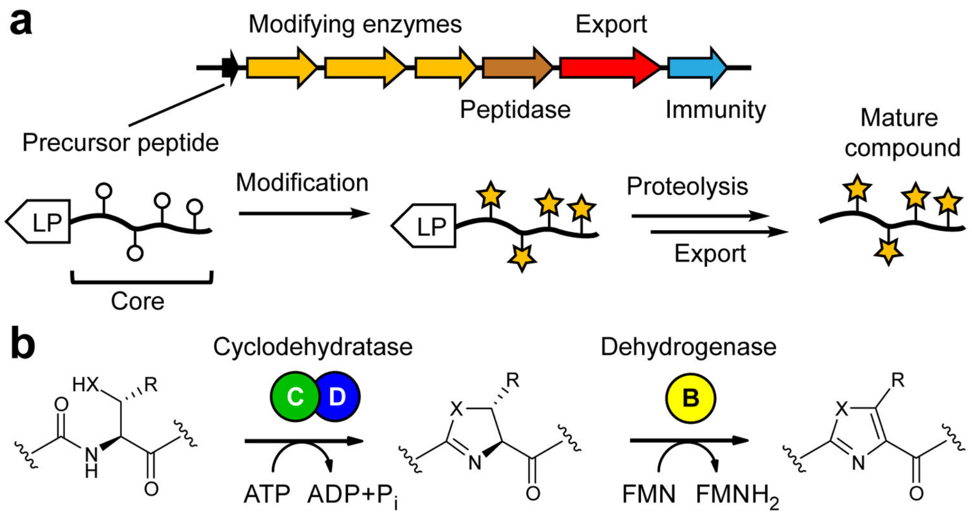


Figure 1. Overview of RiPP biosynthesis and the TOMM subclass

(**a**) A generic RiPP biosynthetic gene cluster is displayed. The precursor peptide is composed of a leader peptide (LP) and core. While the LP contains binding motifs for the modifying enzymes, the core contains residues that undergo enzymatic processing to diverse functional groups. After removal of the LP and any additional tailoring processes, the mature RiPP is exported. (**b**) One RiPP biosynthetic class, the thiazole/oxazole-modified microcins (TOMMs), installs azoline/azole heterocycles. TOMMs arise from the action of an ATP-dependent cyclodehydratase (C and D proteins) and a flavin mononucleotide (FMN)-dependent dehydrogenase (B protein). X = S or O; R = H or CH_3.

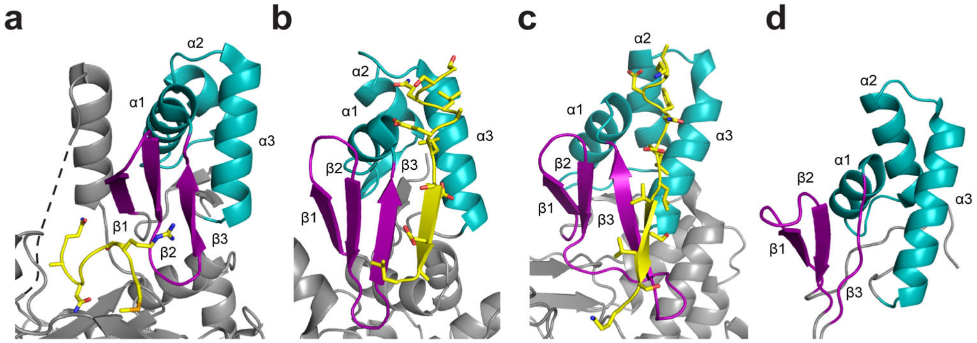
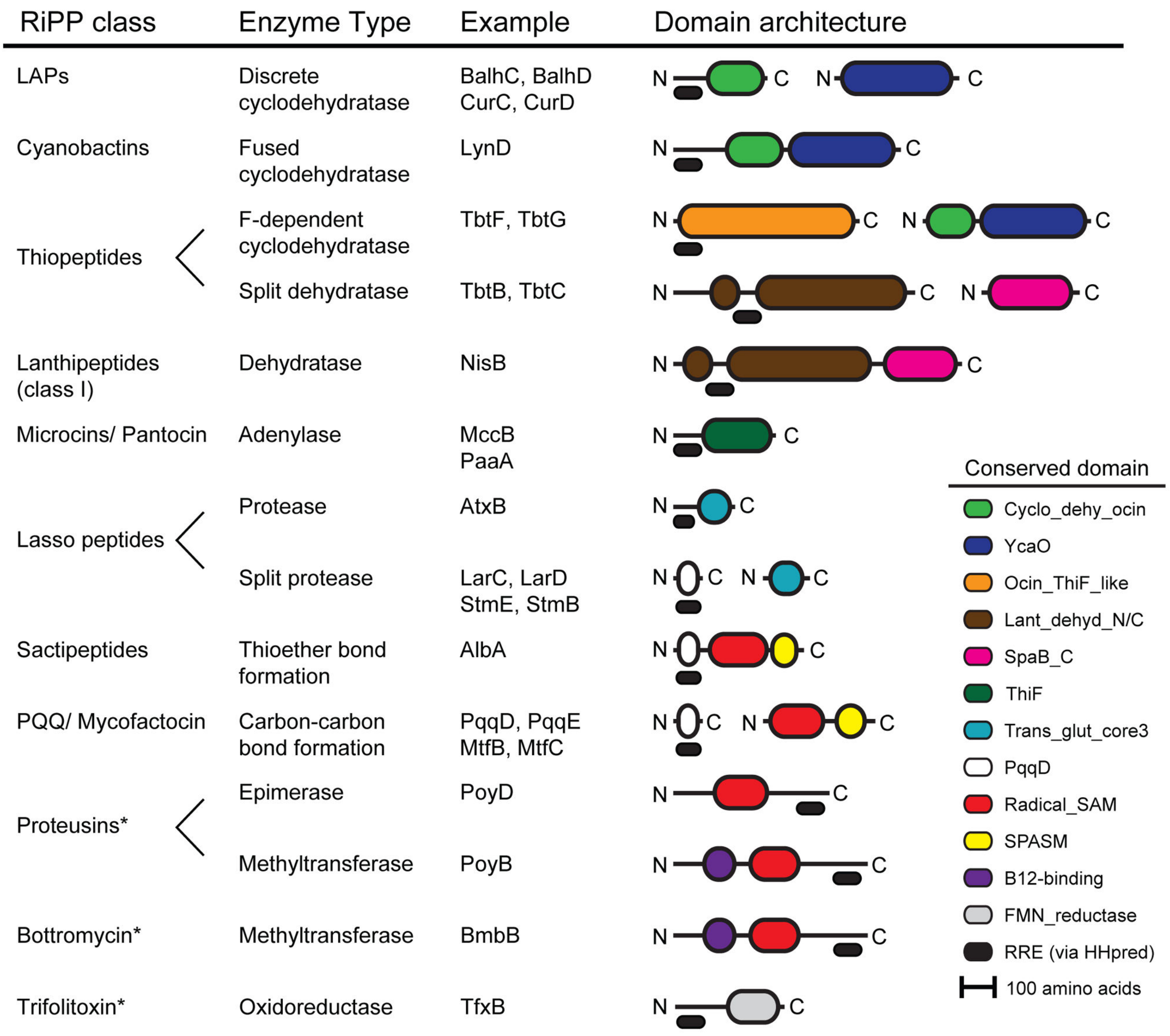


Figure 2. Structural comparison of four RiPP modifying enzymes

Shown are structurally equivalent sections of RiPP biosynthetic proteins involved in the biosynthesis of (**a**) the Trojan horse antibiotic microcin C7 (MccB, an adenylating enzyme [PDB entry 3H9J]), (**b**) antitumor cyanobactins (LynD, a cyclodehydratase [4V1T]), (**c**) the lantibiotic nisin (NisB, dehydratase [4WD9]), and (**d**) the bacterial dehydrogenase cofactor PQQ (PqqD, rSAM-associated [3G2B]). The purple β-sheets and cyan α-helices constitute a conserved RiPP precursor peptide recognition element (RRE). In (**a–c**), the precursor peptide is shown in yellow stick format. Dashed lines indicate missing electron density.



RiPP class	Enzyme Type	Example	Domain architecture
LAPs	Discrete cyclodehydratase	BalhC, BalhD CurC, CurD	
Cyanobactins	Fused cyclodehydratase	LynD	
Thiopeptides	F-dependent cyclodehydratase	TbtF, TbtG	
	Split dehydratase	TbtB, TbtC	
Lanthipeptides (class I)	Dehydratase	NisB	
Microcins/ Pantocin	Adenylase	MccB PaaA	
Lasso peptides	Protease	AtxB	
	Split protease	LarC, LarD StmE, StmB	
Sactipeptides	Thioether bond formation	AlbA	
PQQ/ Mycofactocin	Carbon-carbon bond formation	PqqD, PqqE MtfB, MtfC	
Proteusins*	Epimerase	PoyD	
	Methyltransferase	PoyB	
Bottromycin*	Methyltransferase	BmbB	
Trifolitoxin*	Oxidoreductase	TfxB	

Figure 3. RREs are present in diverse RiPP biosynthetic proteins

RREs were found in a myriad of RiPP biosynthetic proteins using HHpred to search for PqqD homology (thick black bars). Solid lines represent protein sequences with *N*/*C*-termini labeled as “N” and “C.” Colored sections indicate the annotations for the conserved domains identified by the Conserved Domain Database[24]. Asterisks denote RRE assignments based solely on HHpred findings. Abbreviations: LAP, linear azole-containing peptide; PQQ, pyrroloquinoline quinone; rSAM, radical SAM. More details can be found in Supplementary Table 3.

Table 1

Discrete LAP cyclodehydratase RREs bind the leader peptide

BalhC and CurC were tested for binding to FITC-BalhA1-LP and FITC-CurA-LP, respectively. Binding was abolished (K_d >150 μM) for many BalhC mutants (D14A, Y16A, F18A, E21A, F29A, R31A, Y34A, I35A, Y71A, I75A, L78A, and K82A) and these cases are not listed in the table. Error on K_d values represents the s.e.m. from curve fitting analysis of three independent replicates.

Protein	Kd (μM)	Location of mutation
BalhC WT	10 ± 1	
BalhC K20A	52 ± 4	β1
BalhC D23A	50 ± 4	β2
BalhC D32A	21 ± 3	β3
BalhC H38A	21 ± 3	β3
BalhC E66A	60 ± 7	α2
CurC WT	7 ± 1	
CurC D18A	3.3 ± 0.2	β1
CurC Q21A	8 ± 1	β2
CurC E25A	4.3 ± 0.3	β2
CurC Y31A	>150	β3
CurC E36A	3.1 ± 0.2	α1
CurC R68A	79 ± 6	α3
CurC I75A	66 ± 6	α3

Abbreviation: WT, wild-type.

Table 2

F-dependent cyclodehydratase RREs bind the leader peptide

Protein components of the Hca and Tbt F-dependent cyclodehydratases were tested for binding to FITC-HcaA-LP and FITC-TbtA-LP, respectively. Error on K_d values represents the s.e.m. from curve fitting analysis of three independent replicates.

Protein	Kd (nM)	Position of mutation
HcaF WT	89 ± 9	
HcaF D21A	130 ± 15	β1
HcaF F35A	1,700 ± 300	β3
HcaF D38A	77 ± 7	β3
HcaF N72A	79 ± 7	α3
HcaF R73A	>10,000	α3
HcaF E76A	250 ± 50	α3
HcaF E79A	75 ± 6	α3
HcaF I80A	500 ± 60	α3
TbtF WT	66 ± 5	
TbtG WT	>10,000	
TbtF/G	66 ± 5	
TbtF D19A	57 ± 2	β2
TbtF L24A	65 ± 3	β2
TbtF V30A	210 ± 15	β3
TbtF F68A	7,700 ± 400	α3
TbtF R71A	190 ± 22	α3
TbtF Q74A	49 ± 3	α3
TbtF R79A	79 ± 5	α3

Abbreviation: WT, wild-type.

Table 3

Split lasso peptide protease RRE binds the leader peptide

StmC binding to FITC-labeled StmA leader peptide was investigated by FP. Error on K_d values represents the s.e.m. from curve fitting analysis of three independent replicates.

Protein	Kd (nM)	Position of mutation
StmC WT	35 ± 10	
StmC V12A	160 ± 50	β1
StmC N22A	410 ± 120	β2
StmC Y28A	28 ± 5	β3
StmC D68A	38 ± 7	α3
StmC D69A	2600 ± 400	α3
StmC L73A	>10,000	α3
StmC D75A	35 ± 9	α3
StmC Q76A	75 ± 25	α3

Abbreviation: WT, wild-type.