Published in final edited form as:

Proteins. 2014 October ; 82(10): 2691–2702. doi:10.1002/prot.24635.

Analysis of the Linker Region Joining the Adenylation and Carrier Protein Domains of the Modular Non-Ribosomal Peptide Synthetases

Bradley R. Miller[‡], Jesse A. Sundlov[‡], Eric J. Drake, Thomas A. Makin, and Andrew M. Gulick^{*}

Hauptman-Woodward Medical Research Institute and Department of Structural Biology, University at Buffalo. Buffalo, NY 14203

Abstract

Non-Ribosomal Peptide Synthetases (NRPSs) are multi-modular proteins capable of producing important peptide natural products. Using an assembly-line process the amino acid substrate and peptide intermediates are passed between the active sites of different catalytic domains of the NRPS while bound covalently to a peptidyl carrier protein (PCP) domain. Examination of the linker sequences that join the NRPS adenylation and PCP domains identified several conserved proline residues that are not found in standalone adenylation domains. We examined the roles of these proline residues and neighboring conserved sequences through mutagenesis and biochemical analysis of the reaction catalyzed by the adenylation domain and the fully reconstituted NRPS pathway. In particular, we identified a conserved LPxP motif at the start of the adenylation-PCP linker. The LPxP motif interacts with a region on the adenylation domain to stabilize a critical catalytic lysine residue belonging to the A10 motif that immediately precedes the linker. Further, this interaction with the C-terminal sub-domain of the adenylation domain may coordinate movement of the PCP with the conformational change of the adenylation domain. Through this work, we extend the conserved A10 motif of the adenylation domain and identify residues that enable proper adenylation domain function.

Keywords

Structural Biology; Siderophores; Acyl-Carrier Proteins; Enzyme Structure; Enterobactin; Natural Product Biosynthesis; Secondary Metabolism

Introduction

Many bacteria and fungi contain fascinating multi-domain proteins involved in the biosynthesis of novel natural products. [1, 2] One class of these proteins are the Non-Ribosomal Peptide Synthetases (NRPSs) [3, 4], which produce unusual peptides such as the antibiotics vancomycin and gramicidin, as well as the siderophores enterobactin [5],

^{*}To whom correspondence should be addressed at the Hauptman-Woodward Medical Research Institute and Department of Structural Biology, State University of New York at Buffalo, 700 Ellicott St. Buffalo, NY 14203-1102, Phone (716) 898-8619, FAX (716) 898-8660, gulick@hwi.buffalo.edu.

[‡]These authors contributed equally.

yersiniabactin [6], and pyoverdine. [7] NPRSs use a modular architecture with multiple catalytic domains joined as a single protein. While structures of most individual domains have been determined over the last decade [8], only recently have multi-domain structures been determined that begin to provide insights into the interaction and coordination between domains. [9–14]

Most NRPS modules contain a set of core domains that consist of a condensation domain, an adenylation domain, and a peptidyl carrier protein (PCP) domain. The PCP is post-translationally modified with the phosphopantetheine group of coenzyme A (CoA). [15] Bound to this pantetheine through a thioester linkage, amino acid and peptide intermediates are delivered to adjacent catalytic domains in an assembly line fashion. The enterobactin NRPS system (Figure 1A) of *Escherichia coli* has served as a model system for structural and functional studies. [16–18]

The activity of each NRPS module begins with the adenylation domain that loads the amino acid substrate onto the carrier protein cofactor. These adenylation domains belong to a superfamily of acyl-CoA synthetases, NRPS adenylation domains, and beetle luciferase (ANL) adenylating enzymes that all catalyze an initial adenylation partial reaction. [19] In the NRPS adenylation half reaction an amino acyl-adenylate is formed by reaction of the amino acid substrate with Mg•ATP. In the subsequent thioester-forming half reaction, the phosphopantetheine from the carrier protein enters the active site and displaces AMP to form a thioester with the amino acid substrate (Figure 1B). The adenylation and carrier protein domains are joined by a short peptide linker that connects the two domains.

The adenylation domain is composed of a larger N-terminal sub-domain of 400–500 residues and a ~120 residue C-terminal sub-domain. This smaller sub-domain contains a two-stranded β -sheet, and a larger three-stranded β -sheet that is surrounded by α -helices. The C-terminal subdomain of the adenylation domain undergoes a large conformational change and rotates by 140° to adopt the two conformations that are used for the two partial reactions. [20] This rotation allows the active site of the enzyme to change without moving the substrate. Within the ANL superfamily there are ten conserved motifs designated as A1–A10 [19, 21] that play both structural and catalytic roles. The primary binding features of the active site are located in the N-terminal sub-domain, while a catalytic lysine found on the A10 motif (PxxxxGK) of the C-terminal sub-domain is required for acyl-adenylate formation. [22–25] This catalytic lysine binds to both the amino acid substrate and the ATP. [26] When mutated to an alanine catalytic efficiency reduced by at least two orders of magnitude. As a result of the 140° domain rotation, the A10 motif is removed from the active site during the thioester-forming reaction. [19, 27] We have referred to this structural rearrangement as *domain alternation*.

This conformational change of the adenylation domain is centered on a conserved hinge residue found in the A8 motif. The hinge residue, virtually always an aspartic acid or lysine, links the two sub-domains of the adenylation domain and, through changes in the main chain ϕ/ψ angles, enables the proper orientation for both the adenylate and thioester-forming conformations. The importance of the flexibility of this hinge residue has been shown by mutation of the aspartic acid hinge in 4-chlorobenzoyl CoA ligase (CBL) to a proline. This

mutant enzyme is trapped in the adenylate-forming conformation as the hinge would have to pass through unfavorable ϕ/ψ angles during the conformational change. As a result, the activity was reduced 200-fold compared to that of the wild-type CBL. [25] A similar result has been observed with acetyl-CoA synthetase. [24]

Recent crystal structures of multi-domain NRPS enzymes provide an opportunity to examine the linkers that join the various domains. The crystal structure of SrfA-C (2VSQ), a four domain termination module from the surfactin biosynthetic cluster of *Bacillus subtilis*, contains a condensation-adenylation-PCP-thioesterase domain architecture. [13] The two N-terminal domains, the condensation and adenylation domain, share a large interface that likely forms a relatively stable complex that will not change during the NRPS catalytic cycle. The authors referred to this as a “platform” around which the other domains migrate during the NRPS catalytic cycle.

The linkers that surround the PCP are likely to play a dynamic role to allow the PCP to migrate between active sites on adjacent catalytic domains. Downstream of the PCP is usually the condensation domain of the following module or, in the case of a termination module such as SrfA-C and EntF, the thioesterase domain. Upstream however is the C-terminal sub-domain of the adenylation domain that undergoes the 140° domain rotation.

In the SrfA-C structure [13], which has the PCP domain positioned to interact with the condensation domain, the adenylation domain approximates the adenylate-forming conformation. Two crystal structures of adenylation-PCP complexes, EntE-B [12, 14] and PA1221 [11], are both in the conformation that catalyzes the thioester-forming reaction. The EntE-B protein contained two fused proteins from enterobactin synthesis, the standalone adenylation domain EntE and its substrate aryl carrier protein (ArCP) from EntB. [12] This chimeric protein was made as a single protein and thought to be more likely to crystallize than the transient complex between the two separate proteins. The design of the linker between the two proteins was modeled on the natural linker in other adenylation-PCPs by the addition of four residues. While the EntE-B chimera was expected to crystallize in an intra-molecular interaction between the fused adenylation and carrier protein domains, the protein adopted a domain-swapped dimer (Figure 1C) in two independent crystal forms. [12, 14] Standalone adenylation domains such as EntE contain a long α -helix at the C-terminus that folds against the central β -sheet. Because an α -helix of 16 residues is significantly shorter (in distance) than a random coil of the same number of amino acids, the EntE-B linker was not long enough to allow the intra-molecular interaction to occur (Figure 1). Since multi-domain adenylation-PCP proteins can contain linkers of the same number of residues as EntE-B, this C-terminal helix either unwinds or does not exist in the multi-domain NRPSs.

We therefore initiated a structure-based bioinformatic analysis of the linker region between NRPS adenylation and carrier domains. Presented here are the results of this analysis, followed by biochemical examination of conserved residues that were identified. To biochemically assess this region, a series of potentially important linker residues was examined in the multi-domain NRPS enzyme EntF of the enterobactin system in *E. coli*. The results revealed a new motif within this region that interacts with the neighboring

adenylation domain. This interaction facilitates the orientation of the upstream A10 motif for acyl adenylate formation, and may also play a role in coordinating the movement of the carrier protein with the conformational change of the adenylation domain. These results provide evidence that consideration of the linker sequences is required for successfully engineering of NRPS proteins through domain or module shuffling [28] to produce novel peptides.

Materials and Methods

Cloning and Mutagenesis of *entF* and *ybdZ*

The *entF* gene was amplified from the genomic DNA of *E. coli* strain JM109, and cloned into a modified pET15b vector encoding an N-terminal 5x His-tag sequence and TEV protease recognition site [29], resulting in the plasmid pAR233. Site-directed mutagenesis was carried out using the QuikChange Mutagenesis Kit (Agilent) following the manufacturer's protocol. Oligonucleotides were designed for each of the four individual proline mutations in the EntF adenylation-PCP linker region. All plasmids were sequenced to confirm the introduction of the desired mutations only.

The *ybdZ* gene was chemically synthesized (GenScript) and provided in a pUC57 plasmid. The *ybdZ* gene was subcloned into a pET15b vector encoding an N-terminal 5x His-tag sequence and TEV protease recognition site (pBM2).

Expression of EntF, EntE, EntB, and YbdZ

The *entF* expression plasmid was transformed into BL21-DE3 *ybdZ* cells (kindly provided by Dr. Michael G. Thomas, University of Wisconsin) for protein expression. Cells were grown in LB media at 37°C to an OD₆₀₀ of 0.6. Protein expression was induced with 1 mM IPTG and cells were incubated for an additional 18 hours at 16°C. Initial purification of EntF was obtained using two nickel-affinity chromatographic steps. Cells were lysed by sonication in 50 mM Tris pH 7.5, 400 mM NaCl, 0.2 mM TCEP, and 10 mM imidazole. The lysate was passed over a 5mL Ni²⁺•HiTrap Chelating HP column (GE Healthcare); bound proteins were eluted with 300 mM imidazole. Fractions showing EntF by SDS-PAGE analysis were dialyzed over night with TEV protease in cleavage buffer (50 mM Tris pH 7.5, 400 mM NaCl, 0.2 mM TCEP, and 0.5 mM EDTA). At the same time as the TEV protease cleavage, 200 nM Sfp (the promiscuous phosphopantetheinyl transferase from *B. subtilis*), 100 μM CoA, and 1 mM MgCl₂ were added to phosphopantetheinylate the PCP domain. After an overnight dialysis, imidazole was added to 10 mM and the dialyzed protein was passed over a nickel affinity column a second time. Flow-through fractions were collected for dialysis into 50 mM EPPS pH 8.0, 150 mM NaCl, 1 mM MgCl₂ and 0.2 mM TCEP. SDS-PAGE was performed on the final sample to ensure purity. Mutant EntF proteins were purified using the same protocol as the wild-type and produced comparable yields of protein. EntE and EntB were also purified following the same protocol.

The *ybdZ* gene in plasmid pBM2 was transformed into BL21-DE3 cells. Cells were grown in LB media at 37°C to an OD₆₀₀ of 0.6. Protein expression was induced with 500 μM IPTG

and incubated overnight at 16°C. YbdZ was purified using the same buffers and protocol as EntF.

Generation of Adenylation-PCP Domain Database

To identify NRPS adenylation-PCP sequences, the UniProtKB/TreMBL and UniProtKB/Swiss-Prot databases were probed through the PROSITE interface ([HTTP://prosite.expasy.org](http://prosite.expasy.org)) with the following search pattern:

$$[ST]-[STG]-G-[ST]-x(200, 250)-D-x(13, 15)-R-x-[DK]-x(30, 70)-x(40)-P-x(5)-K-x(40, 65)-G-x-x-S-x(20, 30)$$

In the PROSITE definition, x can be any amino acid and x(N,M) represents a gap of residues between N and M amino acids in length. Searching the UniProt database for matches in fully sequenced proteins only (excluding protein fragments), an initial database containing 12,820 hits in 7,689 protein sequences was generated. The large number of hits compared to protein sequences reflects large modular NRPSs that contain multiple adenylation-PCP motifs. To remove equivalent proteins from different strains of the same species, or otherwise highly similar proteins, entries with sequence identities over 90% were clustered and only one representative example of each cluster was used for the following steps. As a result of the clustering, the database was trimmed to 3,626 full length proteins. These 3,626 proteins were then inspected for the adenylation-PCP search pattern, resulting in 6,512 pattern hits. Sequences that contained unresolved residues (sequences with “X” residues) were further omitted, leaving 3,606 full-length proteins containing 6,476 pattern matches. To remove false positives from the database, every entry was required to have both adenylation domain and carrier protein identifiers. The final database contained 6,374 pattern hits in 3,529 full-length proteins.

Construction of Self-Standing Adenylation Domain Database

A second search of the UniProtKB/TreMBL and UniProtKB/Swiss-Prot databases through the PROSITE interface was performed to identify standalone adenylation domains of the ANL superfamily. [19] A more selective motif was used in order to limit the number of results to a reasonable number:

$$[ST]-X-X-[EQ]-X(30, 60)-[RKF]-X-G-X(60, 100)-[ST]-[STG]-G-[ST]-X(200, 250)-D-X(13, 15)-R-X-[DK]-X(75, 100)-P-X(5)-K-X(0, 20).$$

The procedures for removing duplicate proteins were the same as for analysis of adenylation-PCP sequences. To prune the database to contain only standalone domains, further processing steps were taken. First, any protein over one thousand residues was excluded. Next, the list of accession numbers for the adenylation-PCP-containing proteins was compared to the list of adenylation-containing proteins and any duplicates were removed. The remaining selection contains any protein under one thousand residues with an adenylation domain not followed by a PCP domain. This list was manually curated to

remove anything other than standalone adenylation domains. The final database of standalone adenylation domain sequences contained 1,773 proteins.

In Vitro Reconstitution of Enterobactin Synthesis

The biochemical activity of enterobactin biosynthesis was monitored by HPLC. [18] Briefly, 10 μ M EntE and 10 μ M EntB were added to 75 mM EPPS pH 8.0, 0.2 mM TCEP, 10 mM $MgCl_2$, 10 mM ATP, 1.5 mM L-serine (10 mM for L958D), and 1.0 mM 2,3-dihydroxybenzoic acid (DHB) and incubated at 37° C for 10 minutes. At the same time, EntF and YbdZ were incubated together at 37° C. The reaction was initiated with the addition of the EntF/YbdZ to the reaction. 100 μ L of the reaction was quenched at 0.5 to 5 minutes with 150 μ L of 1 N HCl. Enterobactin and intermediates were extracted with ethyl acetate. After evaporating the ethyl acetate, enterobactin was resuspended in 250 μ L of 30% acetonitrile and run on a Zorbax Eclipse Plus C-18 column (4.6 $\times$ 100 mm, 3.5 μ m). Enterobactin and intermediates were eluted using a 12 minute 10% to 50% acetonitrile gradient. The migration of DHB, the DHB-Ser monomer, dimer, trimer, and cyclized enterobactin were detected via UV absorption at 254 nm. Enterobactin (Sigma-Aldrich) was used to produce a standard curve for quantification of enterobactin for each EntF construct.

Pyrophosphate Exchange Assay

To measure serine dependent acyl-adenylate formation by EntF, we used the pyrophosphate (PPi) exchange assay. [30] Briefly, 100 μ L reactions were performed with 1 μ M EntF, 2 mM ATP, 200 μ M Na_4PPi , 0–1 mM L-serine, and 0.15 μ Ci [^{32}P] Na_4PPi . The reaction was carried out at 37° C for five minutes and quenched with 500 μ L of 1.2% activated charcoal (wt/v), 0.1 M Na_4PPi , and 0.35 M perchloric acid. The samples were centrifuged and the charcoal pellet was washed with 500 μ L dH_2O three times. After the final wash the charcoal was re-suspended in 500 μ L dH_2O and transferred into 10 mL of liquid scintillation fluid. Radiolabeled nucleotide was quantified using a Packard Tri-Carb 1900 TR liquid scintillation counter. Because of the higher K_M values, the Y908S mutant enzyme was probed with serine concentrations to 5 mM and the L958D mutant was probed with concentrations to 10 mM.

Computer software

The homology model of EntF was created using the SrfA-C module (PDB 2VSQ) as a molecular model with MODELLER v9.12.[31] The Sequence Logo of linker regions was generated with WEBLOGO3.[32]

Results

Hinge residue

We previously probed the N- and C-terminal sub-domain rotation in ANL adenyating enzymes by examining the impact of mutating the hinge residue to a proline, which traps the enzyme in the adenylate-forming conformation. [24, 25] While this has confirmed the conformational change for acyl-CoA synthetases, the role of the hinge in an NRPS adenylation domain in the context of a multi-domain protein has not been tested. We therefore mutated the hinge residue of EntF, located at Asp857, to a proline. We tested the

effect of the D857P mutation on enterobactin production in a full biochemical reconstitution assay and also specifically on the adenylation partial reaction using the PP_i exchange assay. The hinge mutant failed to produce enterobactin during the *in vitro* reconstitution (Figure 2). After increasing the reconstitution reaction time from 5 to 20 minutes trace amounts of DHB-Ser monomer and dimer were detected. The PP_i exchange results showed that the EntF D857P mutant was competent to catalyze acyl-adenylate formation at rates comparable to wild-type enzyme with only a slight decrease in catalytic efficiency (Table I). The impact of the proline mutation on the overall throughput of the enterobactin biosynthesis pathway confirmed that the domain rotation of the NRPS adenylation domain played an important role and led us to analyze the linker residues located between the adenylation and carrier protein domain.

Bioinformatic Analysis of the Adenylation-PCP Domain Linkers in Comparison with the Homologous C-terminal Regions of Standalone Adenylation Domains

The crystal structure of the chimeric EntE-B protein showed a domain-swapped dimer with the ArCP of each chain interacting with the adenylation domain of the other (Figure 1C). The formation of the domain-swapped dimer prompted a closer look at the linker regions between these two domains and the corresponding regions of standalone adenylation domains (Figure 3A). The examination of an initial panel of several dozen sequences identified a number of proline residues that were found in the linker region of native adenylation-PCP constructs.

To further explore this phenomenon, the search was expanded by scanning a Uniprot database of bacterial proteins using a search pattern that incorporated several conserved motifs of adenylation domains, along with the conserved GxxS motif of PCPs. To capture standalone adenylation domains, a search without the PCP conserved motif was performed and manually curated to remove the di-domain protein sequences. Next, for each protein the sequence directly after the conserved A10 motif was analyzed for proline content. For adenylation-PCP proteins this included the 20 amino acids beginning three residues after the catalytic lysine of the A10 motif. For standalone domains this included all the remaining residues after the A10 motif, which, because the protein sequence terminated, averaged 14 residues. Of the 6,374 adenylation-PCP sequences, 96% had at least one proline following the A10 motif, 88% had more than two prolines, and 75% of the proteins had more than three proline residues in the linker. In contrast, only 34% of the 1,773 standalone adenylation domains contained at least one proline and 7% of the standalone adenylation domains contained two or more. (Figure 3B).

We then examined the sequence of this region more closely (Figure 3D). Position 1 is either a lysine or arginine 61% of the time while the hydrophobic residues alanine and leucine occupy positions 2 and 3 76% and 95% of the time, respectively. The initial observation of the prevalence of Pro residues in adenylation-PCP linkers is borne out at positions 4 and 6, where they are present in 78% and 71% of the sequences. Position 7 is either an aspartic or glutamic acid 45.7% of the time. Positions 8–15 have less dramatic sequence conservation although alanine is the most common residue at every position, except the 14th position, where arginine and alanine are equally likely to appear. Starting with position 15 and

continuing through position 20, there is an increase once again in the propensity for proline. It is the second most common residue at positions 18 and at positions 17, 19, and 20, proline is the most common amino acid and is found in 19%, 32%, and 17% of the sequences.

Biochemical Analysis of Mutations to Residues in the EntF Adenylation-PCP Linker

We used the biochemical reconstitution assay of enterobactin biosynthesis to examine the role of these conserved residues in the adenylation-PCP linker. This reaction contained all of the protein components for enterobactin synthesis as well as necessary substrates Mg•ATP, serine, and DHB. Recent studies have shown that some NRPS adenylation domains are activated through the incubation of a small protein that binds to the adenylation domain. [33–36] These proteins are named after the MbtH protein of mycobactin biosynthesis and are referred to as MbtH-like proteins (MLP). The MLPs are small ~70 residues proteins that contain several conserved tryptophan residues that interact with an alanine on the adenylation domain. [36, 37] Immediately upstream of the *entF* gene, *E. coli* encodes an MLP called YbdZ that binds to EntF and increases adenylation domain activity by reducing the K_M for serine by an order of magnitude. [33] We therefore also included YbdZ, an MLP that binds and activates the adenylation domain of EntF.

We created a series of proline to alanine mutants at the four linker prolines (Pro959, Pro961, Pro968, and Pro972) in EntF (Figure 3C). We mutated each of the four proline residues individually and in combination and tested the mutations in the reconstitution and PP_i exchange assays (Table I). The rate of enterobactin production (measured in $\mu\text{M}/\text{min}$) was unaffected by mutating Pro959, Pro968, and Pro972. Interestingly mutating Pro961 alone dropped the rate of enterobactin formation from 4.00 $\mu\text{M}/\text{min}$ to 2.49 $\mu\text{M}/\text{min}$. This was the only single proline mutation to negatively affect enterobactin production. Mutating the prolines in any combination resulted in a significant decrease in the production rate. Since mutating Pro961 alone reduced the rate drastically compared to the other single proline mutations, this prompted further examination of this region.

As the A10 motif with the catalytic lysine ends and the linker region begins, the linker passes across the surface of the C-terminal sub-domain of the adenylation domain in the SrfA-C crystal structure (Figure 4A). In the majority of adenylation-PCP NRPSs this portion of the linker contains an LPxP motif (Figure 3D) seen at position 3–6 of the linker (Pro959 and 961 in EntF). Located on the central β -sheet of the C-terminal sub-domain where this LPxP motif crossed is a tyrosine residue positioned between the two proline residues of the LPxP motif (Figure 4A). The hydroxyl of this tyrosine forms a hydrogen bond with the carbonyl of the first proline of the LPxP motif and the ring stacks against the leucine. This LPxP motif is present in 70% of the sequences of the adenylation-PCP database. In this subset of proteins, the interacting tyrosine is present 93% of the time.

To biochemically determine the importance of this interaction in EntF, Tyr908 was mutated to a phenylalanine and a serine (Table I). Phenylalanine was chosen to test the role of the hydrogen bond between Tyr908 and the amine of Pro959, while the serine was chosen to disrupt the hydrophobic interaction as well. The rate of enterobactin production for the EntF Y908F mutant was only decreased to 3.01 $\mu\text{M}/\text{min}$. However, the Y908S mutant produced barely detectable amounts of enterobactin and the rate was reduced by over two orders of

magnitude down to 0.02 $\mu\text{M}/\text{min}$. This indicated that while both the hydroxyl and the hydrophobic ring of Tyr908 helped facilitate natural product biosynthesis, the hydrophobic ring had a more significant contribution to enterobactin production.

Exploration of Which Step in Enterobactin Synthesis Effectuated is by the Mutations

Having demonstrated that the Y908S mutant can severely impair EntF function, while the proline mutations have a modest impact on the enterobactin biosynthesis, we asked what step in the synthesis is impaired. Are the mutations preventing delivery of the PCP domain to different catalytic domains of EntF? We therefore used the PPi exchange assay to monitor adenylate formation in the adenylation domain. This reaction serves to demonstrate that the adenylation domain is competent to catalyze the initial reaction.

We determined apparent kinetic constants for serine for the EntF adenylation domain. For the mutant proteins containing multiple proline to alanine substitutions, as well as the Y908F mutant, the effect on the serine adenylation reaction were modest as observed in the biochemical reconstitution reaction. The effect on k_{cat} for all mutants was less than 2-fold and $k_{\text{cat}}/K_{\text{M}}$ values showed at most a 3-fold effect. These results demonstrate that the proline mutations have a limited effect on the active site of the adenylation domain for catalysis of the initial adenylation of serine.

Examination of the rate of the adenylation reaction catalyzed by the Y908S mutant showed a larger effect than the other mutant enzymes. The $k_{\text{cat}}/K_{\text{M}}$ value was reduced by 20-fold compared to wild-type, which was a result of an increase in the K_{M} for serine. This suggested that the interaction between Tyr908 and the leucine of the LPxP motif was required to properly configure the adenylation domain active site. As the Tyr908 residue is more than 15 Å from the active site of the adenylation domain, based on SrfA-C [13] and the phenylalanine activating domain of GrsA [26] (Figure 4B), the most likely cause of this change in the $k_{\text{cat}}/K_{\text{M}}$ value results from changes in the A10 motif which is immediately upstream of the LPxP motif.

To further examine the effect of the LPxP residues on the adenylation reaction, we made a dramatic mutation in which the leucine residue was mutated to an aspartic acid. We chose this substitution because it is roughly isosteric with the leucine side chain but is expected to prevent the LPxP linker region from forming a tight interface with the core of the C-terminal subdomain. This L958D mutant is able to catalyze both assays, at least modestly, demonstrating that we have not affected the overall protein folding or conformation. Nonetheless, it is severely compromised in the PPi exchange assay with over a 1000-fold reduction in catalytic efficiency. In the full enterobactin reconstitution reaction, the serine concentration was increased to 10 mM to accommodate the increased K_{M} of the L958D mutant. Even with this increased serine concentration the L958D mutant could only produce enterobactin at 0.32 $\mu\text{M}/\text{min}$. These results suggest that the integrity of the loop that contains the A10 catalytic lysine is compromised in the L958D mutant enzyme and therefore that the LPxP motif is critical for the proper positioning of this catalytic residue. Although we cannot rule out large changes to the fold of the C-terminal sub-domain, for example, Leu958 is expected to be more than 15 Å from the active site and, like the Y908S mutant, we expect that its effect is largely driven by changes to the A10 motif.

Discussion

We present here the analysis of the role played by residues within the NRPS linker that joins the adenylation and PCP domains. Examination of structures of recent multi-domain NRPS proteins led to an examination of the sequences of this region in a large database of proteins. This identified a common sequence motif at the start of the linker region that we examined through biochemical studies of enterobactin biosynthesis. These experiments demonstrated important interactions that occur between the residues at the start of this linker and the C-terminal subdomain of the NRPS adenylation domain.

Previous studies of members of the ANL family of adenyating enzymes showed that replacing the hinge residue with a proline could trap the enzyme in the adenylate-forming conformation and prevent the conformational change necessary for thioester formation. [19, 24, 25] To demonstrate the same effect in a multi-domain NRPS, the aspartic acid hinge of EntF was mutated to a proline. This D857P mutant was unable to produce enterobactin while still being able to form the adenylate in the PPi exchange reaction. This confirmed the two catalytic conformations for the adenylation domain are required for the NRPS catalytic cycle as predicted from the crystal structures of multi-domain proteins. [11–14] The crystal structures show, however, that the PCP domain and the C-terminal sub-domain do not move as a rigid body but rather there are two components to the movement. [11] This raised the possibility that the sole purpose of the linker was to tether the two domains to bring the PCP close to the adenylation domain interface that is formed when the enzyme adopts the thioester-forming conformation.

The results of our mutagenesis study of EntF then identify a new motif that is common in the N-terminal region of the linker that joins NRPS adenylation and PCP domains. This LPxP motif interacts with a hydrophobic patch of residues on the mobile C-terminal sub-domain of the adenylation domain and appears to play two distinct roles in the NRPS catalytic cycle.

Members of this family of adenyating enzymes contain a universally conserved lysine residue that is part of the A10 motif, within the sequence of PxxxxGK. [19, 21] This lysine residue interacts with oxygen atoms on the acyl substrate carboxylate and the α -phosphate of the nucleotide, and helps to delocalize the negative charge on the transition state for the adenylation reaction. It appears that part of the role of the LPxP motif is to properly orient the loop containing the A10 lysine residue for the adenylation partial reaction. This is shown most clearly with the Y908S and the L958D mutants that are severely compromised for the adenylation reaction despite being 15 Å from the adenylation domain active site.

We created a homology model of the EntF protein based on the structure of SrfA-C (Figure 4D). The two proteins share 26% sequence identity over the full 1293 residues and ~35% identity over the adenylation C-terminal sub-domain and linker region that are most relevant to our analysis. In this homology model, the linker region downstream of A10 begins with the chain reversing course and tracing a path over the surface of the C-terminal sub-domain. The first two residues project their sidechains into solvent and do not make contact with the body of the adenylation domain. Leu958 of the LPxP motif is conserved in 95% of the

sequences in the database. The reason for the conservation is evident upon analysis of the three available adenylation-PCP structures. The leucine sidechain projects towards the internal β -sheet of the C-terminal sub-domain, and is enveloped by hydrophobic residues, including the universally conserved Pro946 of the A10 motif. This explains why the EntF L958D mutant was unable to form the acyl-adenylate. The polar nature of aspartic acid does not allow it to interact with the hydrophobic pocket of the C-terminal sub-domain normally occupied by the leucine. This prevented the A10 motif from adopting a catalytic conformation where the lysine of the A10 motif is positioned in the active site of the adenylation domain.

The A10 motif of NRPS adenylation domains, and the ANL superfamily in general, has been defined as PxxxxGK(V/L). [19, 21] Interestingly, the residues that precede the proline also are well conserved, as these residues form part of the hydrophobic patch that interacts with the LPxP motif at the start of the linker. The residue in front of the A10 proline is a leucine residue 57% of the time and a hydrophobic residue (leucine, isoleucine, methionine, valine, or phenylalanine) in 96 % of the NRPS sequences within our database. Three residues in front of this is another hydrophobic position, which is a leucine in 70% of the sequences and a hydrophobic residue 95% of the time. These positions are held by Leu942 and Leu945 in the EntF protein. In the homology model of EntF, these two residues cradle the Leu958 of the LPxP motif.

Structures of other adenylation domains share this hydrophobic interaction between the linker and adenylation domain. The structure of the truncated phenylalanine activating domain of gramicidin synthetase in the adenylate-forming conformation [26] contains the LPxP motif (Figure 4B). In crystal structures of standalone adenylation domains and acyl-CoA ligases that have been solved in both the adenylate and thioester-forming conformations and that lack the LPxP motif, a short α -helix is formed after the A10 motif. The natural adenylation-PCP di-domain PA1221 that has been structurally characterized also lacks the LPxP motif. [11] However, in PA1221, the short α -helix interacts with the C-terminal sub-domain in a similar manner as seen in the LPxP proteins followed by a coiled linker (Figure 4C). In place of the leucine and proline residues on the coil, the PA1221 amphipathic helix contains three leucine residue that are directed toward the C-terminal subdomain and two arginine residues that point towards solvent. This appears to fulfill the same roles of the LPxP motif in keeping the loop containing the A10 lysine properly oriented. This interaction in PA1221 in the thioester-forming conformation are similar to those observed in DhbE [38] and DltA [39], which crystallized in the adenylate-forming conformation, suggesting the stabilizing hydrophobic interactions between the helix and the C-terminal sub-domain are maintained in both conformations. This stabilizing interaction preserves the conformation of the A10 loop allowing the catalytic lysine to enter the active site without having to undergo a reorganization of the A10 loop.

The biochemical analysis of the LPxP motif and interacting residues showed required interaction for the stabilization of the A10 motif. A second potential role for this interaction between the LPxP motif and the C-terminal sub-domain of the adenylation domain is to couple the movement of the PCP with the 140° rotation of the adenylation domain. This seems reasonable since the interaction between the LPxP motif and the adenylation domain

does not appear to be transient but rather maintained throughout both conformations. This would effectively shorten the length of the linker to those residues following the LPxP motif allowing the PCP to be pulled by the rotating adenylation domain.

The LPxP motif of EntF consists of Pro959 and Pro961. While the P961A mutant resulted in a decreased enterobactin reconstitution rate, P959A did not decrease activity. Mutating both simultaneously did not further decrease the rate significantly. While mutating P968A and P972A individually did not affect the reconstitution rate, the P961A/P968A and P968A/P972A did have a combinatorial affect on the reconstitution rate. Since Pro968 and Pro972 do not interact with the C-terminal sub-domain like the LPxP motif does, it is possible that mutating Pro968 and Pro972 simultaneously disrupted inherent rigidity in the linker that assists in coordinating the movement of the PCP with the rotation of the C-terminal sub-domain. Another possibility is that changing consecutive proline residues to alanine increased the propensity for α -helix formation. This could render the linker too short, as was the case with the chimeric EntE-B.

The results presented here suggest that the small C-terminal subdomain of the NRPS adenylation domains terminates at the LPxP motif and that this sequence defines the end of the adenylation domain. This suggests that the ten residues that follow the A10 lysine residue should be characterized as part of the adenylation domain. The linker region that joins the adenylation and carrier protein domains therefore starts following the LPxP motif and continues to the start of the first α -helix of the following PCP domain.

An understanding of the linker requirements is important for efforts to engineer NRPS proteins by combining domains from different sources. A recent study swapped various PCP domains in the single module IndC from *Photorhabdus luminescens* that produces the blue pigment indigoidine. [40] Replacing the PCP of IndC with that of BpsA, a closely related indigoidine synthetase from *Streptomyces lavedulae* disrupted pigment production. Surprised by this the authors turned to the PCP linkers. By including the adenylation-PCP linker along with the PCP from BpsA, pigment production was restored. The IndC and BpsA adenylation-PCP linkers share 35% sequence identity and 54% similarity and lack the LPxP motif. This seems to suggest that a similar strategy as PA1221 is used, in which a short helical anchor is followed by a coiled linker. Furthermore, it can be assumed that because the BpsA linker-PCP swap had similar pigment production as wild-type IndC, the BpsA linker was able to retain a functional IndC A10 loop. This points to a possible incompatibility between the IndC linker and the BpsA PCP. Finally, when the BpsA PCP-thioesterase linker was included in the swap pigment production was again disrupted. While our results presented here shed light on the required interaction between the adenylation-PCP linker and the C-terminal sub-domain, it is clear that for successful combinatorial biosynthesis a better understanding of how both PCP linkers contribute to PCP movement is needed.

Conclusion

During peptide biosynthesis the PCP must visit each catalytic domain of the NRPS. As suggested through crystal structures of multi-domain NRPSs, this requires substantial

movements and domain reconfiguration. The opening and closing of the condensation domain was suggested based on the closed conformation of the standalone condensation domain CDA-C1 [41], and the 140° domain alternation of the adenylation domain is well characterized. [20, 25] However, exactly how the PCP is able to migrate between active sites has remained a mystery. While the results presented here do not explicitly confirm that the PCP movement is linked to domain alternation, we now have insight into this possible coordination. Our analysis suggests that the A10 motif must be anchored to the C-terminal sub-domain for efficient natural product biosynthesis. This anchoring of the linker to the C-terminal sub-domain would also directly affect the downstream linker and thus the PCP during domain alternation since the C-terminal sub-domain cannot rotate independently of the LPxP motif of the linker.

Combinatorial biosynthesis of novel NRPS systems opens up the possibility for novel, potentially potent, drug candidates. [42, 43] We contend that the linker regions, whether the stable linker of the condensation-adenylation interface or the more dynamic linkers around the PCP, must be taken into consideration. Specifically the adenylation-PCP linker must be long enough to reach all the necessary catalytic active sites. The introduction of proline residues helps to accomplish this by not only disrupting secondary structure formation, but also by adding inherent rigidity. Finally an anchoring component of the linker to the C-terminal sub-domain of the adenylation domain is also employed. Whether the use of a short α -helical region that can hydrophobically interact with the C-terminal sub-domain, as is the case with PA1221, or the use of the more common LPxP motif, this interaction is vital for stabilization of the A10 motif and likely the coordination of the C-terminal sub-domain with the PCP.

Acknowledgments

We thank Dr. Albert S. Reger for the cloning of the *entF* gene from *E. coli* and Dr. Michael G. Thomas for the mutant *ybdZ* expression strain. This research was funded in part by the National Institutes of Health (GM-068440 to A.M.G.). Bradley R. Miller is supported in part by a Stafford Fellowship (Hauptman-Woodward Institute).

References

1. Meier JL, Burkart MD. The chemical biology of modular biosynthetic enzymes. *Chem Soc Rev.* 2009; 38(7):2012–2045. [PubMed: 19551180]
2. Walsh CT. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science.* 2004; 303(5665):1805–1810. [PubMed: 15031493]
3. Conurso HL, Bruner SD. Structure and noncanonical chemistry of nonribosomal peptide biosynthetic machinery. *Nat Prod Rep.* 2012; 29(10):1099–1110. [PubMed: 22729219]
4. Hur GH, Vickery CR, Burkart MD. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep.* 2012; 29(10):1074–1098. [PubMed: 22802156]
5. Raymond KN, Dertz EA, Kim SS. Enterobactin: an archetype for microbial iron transport. *Proc Natl Acad Sci U S A.* 2003; 100(7):3584–3588. [PubMed: 12655062]
6. Bearden SW, Fetherston JD, Perry RD. Genetic organization of the yersiniabactin biosynthetic region and construction of avirulent mutants in *Yersinia pestis*. *Infect Immun.* 1997; 65(5):1659–1668. [PubMed: 9125544]
7. Visca P, Imperi F, Lamont IL. Pyoverdine siderophores: from biogenesis to biosignificance. *Trends Microbiol.* 2007; 15:22–30. [PubMed: 17118662]
8. Strieker M, Tanovic A, Marahiel MA. Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol.* 2010; 20(2):234–240. [PubMed: 20153164]

9. Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G. Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature*. 2008; 454(7206): 903–906. [PubMed: 18704088]
10. Liu Y, Zheng T, Bruner SD. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem Biol*. 2011; 18(11):1482–1488. [PubMed: 22118682]
11. Mitchell CA, Shi C, Aldrich CC, Gulick AM. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry*. 2012; 51(15):3252–3263. [PubMed: 22452656]
12. Sundlov JA, Shi C, Wilson DJ, Aldrich CC, Gulick AM. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem Biol*. 2012; 19(2):188–198. [PubMed: 22365602]
13. Tanovic A, Samel SA, Essen LO, Marahiel MA. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science*. 2008; 321(5889):659–663. [PubMed: 18583577]
14. Sundlov JA, Gulick AM. Structure determination of the functional domain interaction of a chimeric nonribosomal peptide synthetase from a challenging crystal with noncrystallographic translational symmetry. *Acta Crystallogr D Biol Crystallogr*. 2013; 69 (Pt 8):1482–1492. [PubMed: 23897471]
15. Lambalot RH, Gehring AM, Flugel RS, Zuber P, LaCelle M, Marahiel MA, Reid R, Khosla C, Walsh CT. A new enzyme superfamily - the phosphopantetheinyl transferases. *Chem Biol*. 1996; 3(11):923–936. [PubMed: 8939709]
16. Drake EJ, Nicolai DA, Gulick AM. Structure of the EntB multidomain nonribosomal peptide synthetase and functional analysis of its interaction with the EntE adenylation domain. *Chem Biol*. 2006; 13(4):409–419. [PubMed: 16632253]
17. Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT. The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc Natl Acad Sci U S A*. 2000; 97(6):2509–2514. [PubMed: 10688898]
18. Gehring AM, Mori I, Walsh CT. Reconstitution and characterization of the *Escherichia coli* enterobactin synthetase from EntB, EntE, and EntF. *Biochemistry*. 1998; 37(8):2648–2659. [PubMed: 9485415]
19. Gulick AM. Conformational dynamics in the acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol*. 2009; 4:811–827. [PubMed: 19610673]
20. Reger AS, Wu R, Dunaway-Mariano D, Gulick AM. Structural characterization of a 140° domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry*. 2008; 47(31):8016–8025. [PubMed: 18620418]
21. Marahiel MA, Stachelhaus T, Mootz HD. Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chem Rev*. 1997; 97(7):2651–2674. [PubMed: 11851476]
22. Branchini BR, Murtiashaw MH, Magyar RA, Anderson SM. The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry*. 2000; 39(18):5433–5440. [PubMed: 10820015]
23. Horswill AR, Escalante-Semerena JC. Characterization of the propionyl-CoA synthetase (PrpE) enzyme of *Salmonella enterica*: residue Lys592 is required for propionyl-AMP synthesis. *Biochemistry*. 2002; 41(7):2379–2387. [PubMed: 11841231]
24. Reger AS, Carney JM, Gulick AM. Biochemical and Crystallographic Analysis of Substrate Binding and Conformational Changes in Acetyl-CoA Synthetase. *Biochemistry*. 2007; 46(22): 6536–6546. [PubMed: 17497934]
25. Wu R, Reger AS, Lu X, Gulick AM, Dunaway-Mariano D. The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase. *Biochemistry*. 2009; 48(19):4115–4125. [PubMed: 19320426]
26. Conti E, Stachelhaus T, Marahiel MA, Brick P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J*. 1997; 16 (14):4174–4183. [PubMed: 9250661]

27. Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC. The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5'-propylphosphate and coenzyme A. *Biochemistry*. 2003; 42(10):2866–2873. [PubMed: 12627952]
28. Giessen TW, Marahiel MA. Ribosome-independent biosynthesis of biologically active peptides: Application of synthetic biology to generate structural diversity. *FEBS Lett*. 2012; 586(15):2065–2075. [PubMed: 22273582]
29. Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD, Waugh DS. Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng*. 2001; 14(12):993–1000. [PubMed: 11809930]
30. Rusnak F, Faraci WS, Walsh CT. Subcloning, expression, and purification of the enterobactin biosynthetic enzyme 2,3-dihydroxybenzoate-AMP ligase: demonstration of enzyme-bound (2,3-dihydroxybenzoyl)adenylate product. *Biochemistry*. 1989; 28(17):6827–6835. [PubMed: 2531000]
31. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A. Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct*. 2000; 29:291–325. [PubMed: 10940251]
32. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res*. 2004; 14(6):1188–1190. [PubMed: 15173120]
33. Felnagle EA, Barkei JJ, Park H, Podevels AM, McMahon MD, Drott DW, Thomas MG. MbtH-like proteins as integral components of bacterial nonribosomal peptide synthetases. *Biochemistry*. 2010; 49(41):8815–8817. [PubMed: 20845982]
34. Boll B, Taubitz T, Heide L. Role of MbtH-like proteins in the adenylation of tyrosine during aminocoumarin and vancomycin biosynthesis. *J Biol Chem*. 2011; 286(42):36281–36290. [PubMed: 21890635]
35. Zolova OE, Garneau-Tsodikova S. Importance of the MbtH-like protein TioT for production and activation of the thiocoraline adenylation domain of TioK. *Med Chem Comm*. 2012; 3(8):950.
36. Herbst DA, Boll B, Zocher G, Stehle T, Heide L. Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes. *J Biol Chem*. 2013; 288(3):1991–2003. [PubMed: 23192349]
37. Drake EJ, Cao J, Qu J, Shah MB, Straubinger RM, Gulick AM. The 1.8 Å crystal structure of PA2412, an MbtH-like protein from the pyoverdine cluster of *Pseudomonas aeruginosa*. *J Biol Chem*. 2007; 282(28):20425–20434. [PubMed: 17502378]
38. May JJ, Kessler N, Marahiel MA, Stubbs MT. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A*. 2002; 99:12120–12125. [PubMed: 12221282]
39. Du L, He Y, Luo Y. Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from *Bacillus cereus*. *Biochemistry*. 2008; 47(44):11473–11480. [PubMed: 18847223]
40. Beer R, Herbst K, Ignatiadis N, Kats I, Adlung L, Meyer H, Niopek D, Christiansen T, Georgi F, Kurzawa N, Meichsner J, Rabe S, Riedel A, Sachs J, Schessner J, Schmidt F, Walch P, Niopek K, Heinemann T, Eils R, Di Ventura B. Creating functional engineered variants of the single-module non-ribosomal peptide synthetase IndC by T domain exchange. *Mol Biosyst*. 2014
41. Bloudoff K, Rodionov D, Schmeing TM. Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *J Mol Biol*. 2013; 425(17):3137–3150. [PubMed: 23756159]
42. Doekel S, Coeffet-Le Gal MF, Gu JQ, Chu M, Baltz RH, Brian P. Non-ribosomal peptide synthetase module fusions to produce derivatives of daptomycin in *Streptomyces roseosporus*. *Microbiology*. 2008; 154(Pt 9):2872–2880. [PubMed: 18757821]
43. Nguyen KT, Ritz D, Gu JQ, Alexander D, Chu M, Miao V, Brian P, Baltz RH. Combinatorial biosynthesis of novel antibiotics related to daptomycin. *Proc Natl Acad Sci U S A*. 2006; 103(46):17462–17467. [PubMed: 17090667]

Figure 1.

Enterobactin biosynthesis. A. The enterobactin biosynthesis pathway uses three NRPS enzymes. A molecule of 2,3-dihydroxybenzoate is activated by the standalone adenylation domain of EntE and loaded onto the aryl-carrier protein domain of EntB, while serine is similarly loaded by EntF adenylation domain to the integrated PCP. The EntF condensation domain then transfers the molecule of DHB to the serine, forming an amide. Three iterations of this synthesis allow the thioesterase domain to combine the DHB-Ser amides to form a cyclized enterobactin molecule. B. The two-step reaction catalyzed by the integrated adenylation domain of EntF is shown. The reaction is catalyzed by the adenylation domain (yellow) and results in the serine residue placed on the downstream carrier protein (CP) domain. C. The crystal structure of the chimeric EntE-B protein (PDB 3RG2) illustrated a domain-swapped dimeric architecture in which each molecule donates the aryl-carrier protein to the alternate adenylation domain. Asp522, at the start of the adenylation-PCP linker is shown with a yellow sphere, while Ser540 is shown in orange and red spheres for the residue within the same chain and the domain-swapped chain, respectively. The orange and red dashed lines then demonstrate the longer length that would be needed for the helical linker to reach to form an intramolecular interface between EntE and EntB. D. Sequence alignment of the adenylation-PCP linker region of several multi-domain NRPS proteins, as well as the chimeric EntE-B protein. The EntE sequence terminates at RASA and the EntB carrier protein begins at IPAS. The GRAS insertion was introduced after similarity with the linker of EntF (GRAP).

Figure 2. Biochemical Reconstitution of enterobactin biosynthesis. The enterobactin biosynthetic NRPS enzymes were incubated with substrates and the MbtH-like protein YbdZ. The reaction was quenched and separated on HPLC. Peaks of enterobactin and 2,3-DHB are labeled. The trace of the wild-type (blue) and the D857P mutant (red) are shown, demonstrating that the proline substitution at the hinge residue severely compromises enterobactin biosynthesis. The chromatographic traces are offset slightly to better visualize. The peaks at 6 and 8 min are linear dimers and trimers of the DHB-serine amide.

Figure 3.

Bioinformatic analysis of the linker region. A. A sequence alignment of the linker that spans the adenylation and PCP domains of multi-domain NPRSs shows the prevalence of proline residues in the linker of multi-domain sequences. The catalytic lysine at the C-terminus of the adenylation domain (blue) and the pantetheinylation site (GGHS, green) of the carrier protein domain are highlighted. B. A histogram of the number of proline residues in the 20 residues that follow the catalytic lysine of multi-domain and standalone adenylation highlights the number of proline residues in multi-domain sequences. C. Sequence alignment and residue numbering of EntF and SrfA-C spanning the A10 catalytic motif to the pantetheine binding site. D. Sequence Logo representation of the linker in multi-domain

NRPS proteins illustrates conserved elements, highlighting the LPxP motif at the start of the linker and the higher proline propensity at the C-terminus.

Figure 4.

Structural analysis of linker regions of NRPS proteins. A. The linker of SrfA-C (PDB 2VSQ) is shown. The C-terminal sub-domain of the adenylation domain is shown in green and the carrier protein is shown in pink. The A10 motif is highlighted in blue and the linker is in darker grey. The LPXP motif is highlighted and the interaction between Tyr906 and the PXP residues are shown. B. A ribbon diagram of the genetically truncated adenylation domain of gramicidin synthetase (PDB 1AMU) is shown, illustrating the conserved LPXP motif of this enzyme and the similar interaction with the Tyr residue. The A10 and C-terminal sub-domains are colored as in panel A. C. The ribbon diagram of PA1221 (PDB 4DG9) is shown with a C-terminal sub-domain, adopting the thioester-forming conformation, colored tan. The short helix at the start of the linker is followed by several disordered residues (dashed line). The PA1221 protein lacks the LPXP motif, but still contains a leucine residue that makes similar hydrophobic interactions with residues on the C-terminal sub-domain. D. Ribbon diagram of the homology model of EntF illustrates the interaction of LPXP motif around the side chain of Tyr908 and in close proximity to Leu945 and Pro946 of the A10 motif. The side chain of Leu960, the x of the LPXP motif, projects away from the C-terminal subdomain and is not shown for clarity.

Table I

Kinetic parameters of mutant EntF proteins

EntF Construct	Adenylation Partial Reaction Apparent Kinetic Constants (Ser)			Reconstitution Rate ($\mu\text{M}/\text{min}$)
	k_{cat} (s^{-1})	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ ($\text{M}^{-1}\text{s}^{-1}$)	
Wild-type	$(5.0 \pm 0.3) \times 10^{-1}$	91 ± 14	5.5×10^3	4.00
D857P	$(6.1 \pm 0.2) \times 10^{-1}$	379 ± 27	1.6×10^3	0.0
P959A	<i>a</i>	–	–	4.75
P961A	–	–	–	2.49
P968A	–	–	–	4.95
P972A	–	–	–	6.06
P959A/P961A	$(3.1 \pm 0.03) \times 10^{-1}$	205 ± 8	1.5×10^3	2.22
P961A/P968A	–	–	–	1.20
P968A/P972A	$(3.6 \pm 0.2) \times 10^{-1}$	89 ± 18	4.0×10^3	2.38
P959A/P961A/P968A/P972A	$(2.4 \pm 0.2) \times 10^{-1}$	179 ± 36	1.3×10^3	1.70
Y908F	$(4.9 \pm 0.1) \times 10^{-1}$	107 ± 11	4.6×10^3	3.01
Y908S	$(4.6 \pm 0.2) \times 10^{-1}$	1720 ± 223	2.7×10^2	0.02
L958D	$(0.083 \pm 0.01) \times 10^{-1}$	2361 ± 708	3.5	0.32^b

^a –, not performed^b assessed with excess serine (10 mM)