The synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase

Janice M. Reimer[1]*, Martin N. Aloise[1]*, Paul M. Harrison[2] & T. Martin Schmeing[1]

[1]Department of Biochemistry, McGill University, 3649 Prom. Sir-William-Osler, Montréal, QC H3G 0B1, Canada; [2]Department of Biology, McGill University, 1205 Dr. Penfield Ave, Montréal, QC, Canada H3A 1B1

*Equal contribution.

Correspondence to T.M.S: martin.schmeing@mcgill.ca

Nonribosomal peptide synthetases (NRPSs) are massive proteins that produce small peptide molecules with wide-ranging biological activities, including green chemicals and many widely-used therapeutics[1]. NRPSs are true macromolecular machines, with modular assembly-line logic, a complex catalytic cycle, moving parts and many active sites[2,3]. In addition to the core domains required to link the substrates, they often include specialized tailoring domains, which introduce chemical modifications and allow the product to access a large expanse of chemical space[3,4]. It is still unknown how any of the NRPS tailoring domains are structurally accommodated into the megaenzymes or how they have adapted to function in nonribosomal peptide synthesis. Here we present a series of crystal structures of the initiation module of an antibiotic-producing NRPS, linear gramicidin synthetase[5,6]. This module includes the specialized tailoring formylation domain, and we capture states that represent every major step of the assembly-line synthesis in the initiation module. The transitions between conformations are staggering, with both the peptidyl carrier protein and the adenylation subdomain undergoing huge movements to transport substrate between distal active sites. The structures highlight the great versatility of NRPSs, as small domains repurpose and recycle their limited interfaces to interact with their various binding partners. Understanding tailoring domains is important if NRPSs are to be exploited for production of novel therapeutics.

Tailoring domains embedded within NRPSs are vital for the production and bioactivity of these synthetases' nonribosomal peptide (NRP) products[4]. Tailoring domains exist in addition to the core NRPS adenylation (A), peptidyl carrier protein (PCP) and condensation (C) domains, which a module requires to add an amino acid to the growing NRP: the A domain selects, activates and transfers the substrate amino acid to the PCP domain, which transports it to the C domain for peptide bond formation[1] (Fig. 1, Extended Data Fig. 1). Tailoring domains are common in NRPSs[3,4]. For example, cyclosporin synthetase contains methyltransferase domains[7]; daptomycin ("Cubicin") synthetase, epimerization domains[8]; bactitracin ("BACiiM") synthetase, a heterocyclization domain[9]; valinomycin synthetase, ketoreductase domains[10]; bleomycin synthetase, an oxidase domain[11]; and soframycin synthetase, a reductase domain[12]. These domains impart key functionalities into the NRP, by, for example, providing protease resistance, enabling novel interactions, improving

affinity by limiting NRP conformational flexibility, or allowing the NRP to assume its active conformation. Linear gramicidin synthetase (LgrA-D) was found by Marahiel and colleagues to contain an active formylation (F) domain as the first domain of its F-A-PCP initiation module[5,6] (Fig. 1). F domains are homologous to formyltransferase (FT) proteins that modify substrates in three diverse pathways: ribosomal translation[13], purine anabolism[14] and bacterial outer membrane synthesis[15]. The LgrA initiation module must formylate its substrate for linear gramicidin synthesis to proceed[6] (Fig. 1), and this formyl group is essential for gramicidin's clinically important antibacterial activity[16]. Gramicidin molecules form head-to-head dimers through the formyl group to make a β helical pore in gram-positive bacterial membranes. This pore freely allows passage of monovalent cations, destroying the ion gradient and killing the bacteria.

We have determined four independent crystal structures of the initiation module of LgrA, at 2.5, 2.6, 2.8 and 3.8 Å resolutions (Extended Data Table 1, Extended Data Fig. 2), that show four different functional conformations: the A domain open (substrate binding), A domain closed (adenylation), thiolation and formylation states (Fig. 2, Extended Data Fig. 3, Supplemental Video 1). This augments the excellent existing structural knowledge of NRPSs, (reviewed by Weissman[2]) by visualizing the structure of an NRPS module that includes a tailoring domain, showing how the tailoring domain is incorporated into and used as part of an NRPS, observing several functional states (open, closed, thiolation) in a single protein, rather than over different excised mono- and di-domains[2,17-21], and visualizing a novel functional state (formylation).

The F domain is connected to the rest of the F-A-PCP LgrA initiation module through an interface with the A domain (Fig. 3) that buries 830 Å² of surface area. This is distinct from the C-A interface in C-A-PCP elongation and termination modules[22] (Extended Data Fig. 4). The F-A interface appears sufficient to maintain these domains in a very elongated conformation (Fig. 2). Across all nonequivalent molecules in the crystals, the relative orientation between the two domains varies only by ~5°, and our small angle X-ray scattering analysis indicates that this extended conformation is representative of the initiation module in solution (Extended Data Fig. 5). This architecture means that the adenylation active site and the formylation

active site are always ~50 Å apart, necessitating that valine substrate travel a large distance between subsequent steps in synthesis. Accordingly, positions of the PCP domain and the A_{sub} domain (C-terminal portion of the A domain) change markedly in the module's progression through functional states.

The NRPS assembly-line process (Fig. 1, Extended Data Fig. 1, Supplemental Video 2) begins with ATP and valine binding to an open conformation of the A domain[20] (Fig. 2a, Supplemental Video 1). The A domain closes upon substrate binding by rotating the A_{sub} by ~30° to catalyze formation of the valine adenylate[17,18] (Fig. 2b). Next, the thiolation reaction transfers the valine from the adenylate to the thiol of the PCP domain's phosphopantetheine arm (PPE). We accessed this state by attaching a non-hydrolyzable analog[23] of the product of the reaction, valine-NH-PPE, to the PCP domain. The resulting structure shows the known 140° rotation of the A_{sub}[19,21] and the product valine-NH-PCP still bound to the active site (Fig. 2c). The PCP must now transport its valine 50 Å between A and F domain active sites to accept a formyl group. Our next structure (Fig. 2d) shows that to achieve this, the PCP domain makes a massive movement of a rigid ~75° rotation and 61 Å translocation (Fig. 2e). The ~10 residue linker between A and PCP domains is not nearly sufficient to span the 55 Å travelled by the first residue of the PCP domain; accordingly, the A_{sub} domain undergoes a full 180° rotation and 21 Å translocation to allow PCP domains to bind the F domain (Fig. 2f). There, valine-PCP accepts a formyl group from the donor cofactor formyl-tetrahydrofolate (fTHF) (Extended Data Fig. 2f) onto its amino group[5,6]. The PCP will then move the formyl-valine to the next module, where that module's condensation (C) domain will catalyze peptide bond formation between fVal-PCP and its glycine-PCP2, making the first peptide bond of linear gramicidin and liberating the PCP to participate in the next round of reactions[6] (Supplemental Video 2).

How did the F domain become a functional NRPS domain? LgrA's F domain was fused into an existing NRPS[5], and we suggest that the pre-transfer source was a single domain FT from a distantly related bacterium with a signature of missing helix α2 and strand β3. As the high incidence of horizontal transfer (Extended Data Fig. 6) is consistent with conferring a competitive advantage, and bacteria possessing FTs similar to the F domain also have canonical tRNA and phosphoribosylglycinamide

FTs, it is likely that the pre-transfer FT performed the remaining known FT function, sugar formylation for cell wall synthesis. After fusion, the F-A-PCP initiation module evolved rapidly (Extended Data Fig. 6, 7). The fold of the first 171 amino acids of LgrA is conserved with the sugar FTs, leaving only residues 172-179, including a single α-helix, as a new structural element and link to the A domain (Extended Data Fig. 8). A "landing pad" evolved to include a hydrophobic patch for binding the PCP domain, and positive residues and hydrogen bond donors to interact with the PPE phosphate (Fig. 3, Extended Data Fig. 8). The F-PCP interaction places the PPE attachment point, Ser729, an ideal 16 Å away from the fTHF in the conserved FT active site. Interestingly, this positions the valine-PPE exactly in the sugar-dTDP binding site of sugar FTs[5,15,24] (Fig. 4a). The similar length and hydrophilic nature of the sugar-dTDP and valine-PPE likely enabled the F domain to formylate valine-PCP soon after the fusion event, before formylation was absolutely required for downstream peptide synthesis to proceed.

The PCP domain interaction with the F domain is quite minimal, and accordingly, the A_{sub} domain donates an additional binding interaction in the formylation state[5] (Fig. 3b). This is very reminiscent of methionyl-$tRNA_f^{Met}$ formyltransferase (FMT), the essential bacterial two-domain FT that uses its C-terminal domain (FMT_{CTD}) to present the methionyl-$tRNA_f^{Met}$ to the FT active site[13] (Fig. 4b,c). This functional convergent evolution presents yet another interesting parallel to the completely separate macromolecular system that synthesizes peptides, ribosomal translation. In both LgrA and the ribosome, a mobile carrier macromolecule (PCP domain / tRNA) covalently (through thioester / ester bonds) transports an amino acid to a formyltransferase enzyme (F domain/FMT), where the carrier is oriented by a positioning domain (A_{sub}/FMT_{CTD}) to allow formylation, before acting as the first donor substrate for a peptidyl transferase enzyme (C domain/large ribosomal subunit).

Observing the same protein all these conformations, including the novel formylation conformation, highlights and further demonstrates the great versatility of the small domains in NRPSs[2,3,18,23]. The small ~100 residue A_{sub} has three distinct roles in the cycle: providing catalytic residues for the adenylation reaction[18]; positioning the PCP for the thiolation reaction and later for the formylation reaction, and bridging the distance between the active sites the PCP must visit[21,22]. The A_{sub} uses different

surfaces for each of these roles (Extended Data Fig. 9a). In addition, the F domain adds to the long list of partners with which the equally small PCP domain must interact (A, F, C, TE, all tailoring domains), and it performs its tasks with overlapping surfaces[25] (Extended Data Fig. 9b).

By co-opting and adapting an FT, nature has further increased the functionality of NRPSs. The formyl functionality seems to be very useful in nonribosomal peptides, as F domains have been incorporated into NRPSs multiple independent times: The F domains in kolossin A synthetase[26], anabaenopeptilide synthetases[24], the oxazolomycin synthetase[27] and a dozen orphan NRPSs and NRPS-PKSs arose from a separate fusion event with an FMT and display a different FMT- FMT_{CTD} -$C_{partial}$-A-PCP domain sequence in their initiation modules[5]. Sampling additional chemical space can lead to novel or improved activity in nonribosomal peptides, which inspires many bioengineering experiments on NRPSs[28] aimed to meet the dynamic challenges to human health. NRPSs are well placed to be engineered for production of new compounds because their synthetic scheme is conceptually straightforward, and NRPSs already naturally produce many therapeutics, as well as promising NRPs like teixobactin[29] and piperidamycin[30], two recently discovered, first-in-class compounds with strong antibacterial activity. The structures presented here reveal the interface between the F and A domains and show all the interactions that the PCP domain makes in the LgrA initiation module. This knowledge could substantially facilitate our ability to introduce an F domain into a foreign NRPS, and make formylation an accessible tool in the NRPS bioengineering toolkit.

References

1 Walsh, C. T. Polyketide and nonribosomal peptide antibiotics: modularity and versatility. *Science* **303**, 1805-1810 (2004).

2 Weissman, K. J. The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* **11**, 660-670 (2015).

3 Hur, G. H., Vickery, C. R. & Burkart, M. D. Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* **29**, 1074-1098 (2012).

4 Walsh, C. T. *et al.* Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol* **5**, 525-534 (2001).

5 Kessler, N., Schuhmann, H., Morneweg, S., Linne, U. & Marahiel, M. A. The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J Biol Chem* **279**, 7413-7419 (2004).

6 Schoenafinger, G., Schracke, N., Linne, U. & Marahiel, M. A. Formylation domain: an essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin. *J Am Chem Soc* **128**, 7406-7407 (2006).

7 Lawen, A. & Zocher, R. Cyclosporin synthetase. The most complex peptide synthesizing multienzyme polypeptide so far described. *J Biol Chem* **265**, 11355-11360 (1990).

8 Robbel, L. & Marahiel, M. A. Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J Biol Chem* **285**, 27501-27508 (2010).

9 Konz, D., Klens, A., Schorgendorfer, K. & Marahiel, M. A. The bacitracin biosynthesis operon of Bacillus licheniformis ATCC 10716: molecular characterization of three multi-modular peptide synthetases. *Chem Biol* **4**, 927-937 (1997).

10 Cheng, Y. Q. Deciphering the biosynthetic codes for the potent anti-SARS-CoV cyclodepsipeptide valinomycin in Streptomyces tsusimaensis ATCC 15141. *Chembiochem* **7**, 471-477 (2006).

11 Schneider, T. L., Shen, B. & Walsh, C. T. Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42**, 9722-9730 (2003).

12 Koketsu, K., Minami, A., Watanabe, K., Oguri, H. & Oikawa, H. Pictet-Spenglerase involved in tetrahydroisoquinoline antibiotic biosynthesis. *Curr Opin Chem Biol* **16**, 142-149 (2012).

13 Schmitt, E., Panvert, M., Blanquet, S. & Mechulam, Y. Crystal structure of methionyl-tRNAfMet transformylase complexed with the initiator formyl-methionyl-tRNAfMet. *EMBO J* **17**, 6819-6826 (1998).

14 Almassy, R. J., Janson, C. A., Kan, C. C. & Hostomska, Z. Structures of apo and complexed Escherichia coli glycinamide ribonucleotide transformylase. *Proc Natl Acad Sci U S A* **89**, 6114-6118 (1992).

15 Thoden, J. B., Goneau, M. F., Gilbert, M. & Holden, H. M. Structure of a sugar N-formyltransferase from Campylobacter jejuni. *Biochemistry* **52**, 6114-6126 (2013).

16 Wallace, B. A. Common structural features in gramicidin and other ion channels. *Bioessays* **22**, 227-234 (2000).

17 Conti, E., Stachelhaus, T., Marahiel, M. A. & Brick, P. Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J* **16**, 4174-4183 (1997).

18 Yonus, H. *et al.* Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J Biol Chem* **283**, 32484-32491 (2008).

19 Reger, A. S., Wu, R., Dunaway-Mariano, D. & Gulick, A. M. Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry* **47**, 8016-8025 (2008).

20 Gulick, A. M. Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem Biol* **4**, 811-827 (2009).

21 Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* **51**, 3252-3263 (2012).

22 Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659-663 (2008).

23 Liu, Y. & Bruner, S. D. Rational manipulation of carrier-domain geometry in nonribosomal peptide synthetases. *Chembiochem* **8**, 617-621 (2007).

24 Rouhiainen, L. *et al.* Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in Anabaena strain 90. *Molecular Microbiology* **37**, 156-167 (2000).

25 Lohman, J. R. *et al.* The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. *Proteins* **82**, 1210-1218 (2014).

26 Bode, H. B. *et al.* Structure elucidation and activity of kolossin A, the D-/L-pentadecapeptide product of a giant nonribosomal peptide synthetase. *Angewandte Chemie International Edition* **54**, 10352-10355 (2015).

27 Zhao, C. *et al.* Oxazolomycin biosynthesis in Streptomyces albus JA3453 featuring an "acyltransferase-less" type I polyketide synthase that incorporates two distinct extender units. *J Biol Chem* **285**, 20097-20108 (2010).

28 Clardy, J., Fischbach, M. A. & Walsh, C. T. New antibiotics from bacterial natural products. *Nature biotechnology* **24**, 1541-1550 (2006).

29 Ling, L. L. *et al.* A new antibiotic kills pathogens without detectable resistance. *Nature* **517**, 455-459 (2015).

30 Hosaka, T. *et al.* Antibacterial discovery in actinomycetes strains with mutations in RNA polymerase or ribosomal protein S12. *Nature biotechnology* **27**, 462-464 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature

Acknowledgements We thank Thierry Mintya, Daina Avizonis, Marie-Christine Tang and Alexandra Furtos for experimental support, Robert Zamboni, John Collucci and Kimiaka Guerard for small molecule synthesis assistance, Jessie Jiang, Diego Alonzo and Dmitry Rodinov for experimental advice and assistance, Shaun Labuik and Pawel Grochulski (Canadian Light Source) for diffraction data collection, Richard Gillian (CHESS SAXS beamline G1), Monica Pillon and Alba Guarne for SAXS help, Kristjan Bloudoff, Michael Tarry, Asfarul Haque and Albert Beghuis for helpful discussions and Jerry Pelletier, Nancy Rogerson and Ali Nahvi for critical reading of the manuscript. This work was supported by CIHR grant 106615, a HFSP

CDA and a Canada Research Chair in Macromolecular Machines to T.M.S.. J.M.R. is supported by an NSERC Alexander Graham Bell studentship, and M.N.A. by a studentship from the CIHR Training Grant in Chemical Biology.

Author Contributions J.M.R. and M.N.A. cloned, expressed, purified, assayed, crystallized and determined the structure of LgrA F-A and F-A-PCP proteins. J.M.R. performed the CoA syntheses and prepared figures. P.M.H. performed the bioinformatics analyses. T.M.S designed the study, analysed data and wrote the manuscript with input from the other authors.

Author Information Crystallography data have been deposited in the Protein Data Bank under accession codes 5ES5, 5ES6, 5ES7, 5ES8 and 5ES9. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to T.M.S. (martin.schmeing@mcgill.ca).

Figures

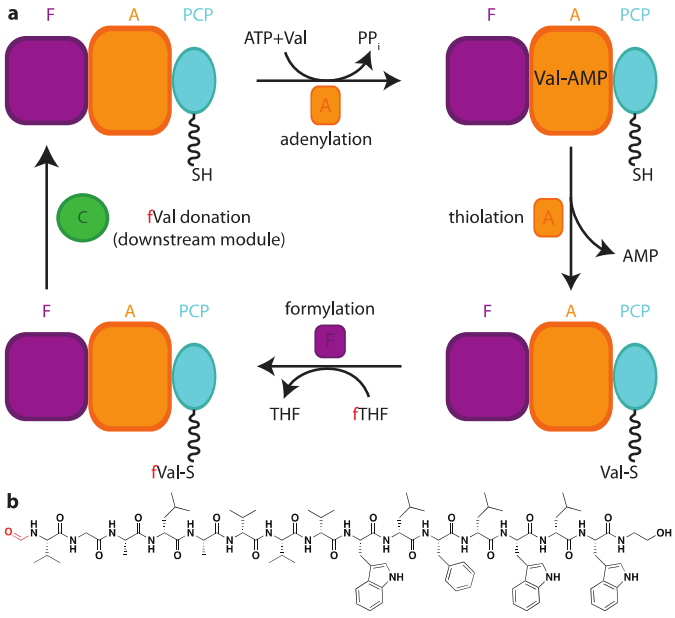


Figure 1: A schematic of the action of the linear gramicidin synthetase initiation module.

a, The F-A-PCP initiation module is the first module of LgrA, the dimodular F-A-PCP-C-A-PCP-E* NRPS protein in the LgrA-E synthetic cluster. The initiation cycle begins with valine selection and adenylation followed by thiolation onto the PPE arm

of the PCP domain. The F domain formylates PCP-PPE-Val before it is brought to the downstream module to be the donor in that module's condensation reaction. **b,** Linear gramicidin A.

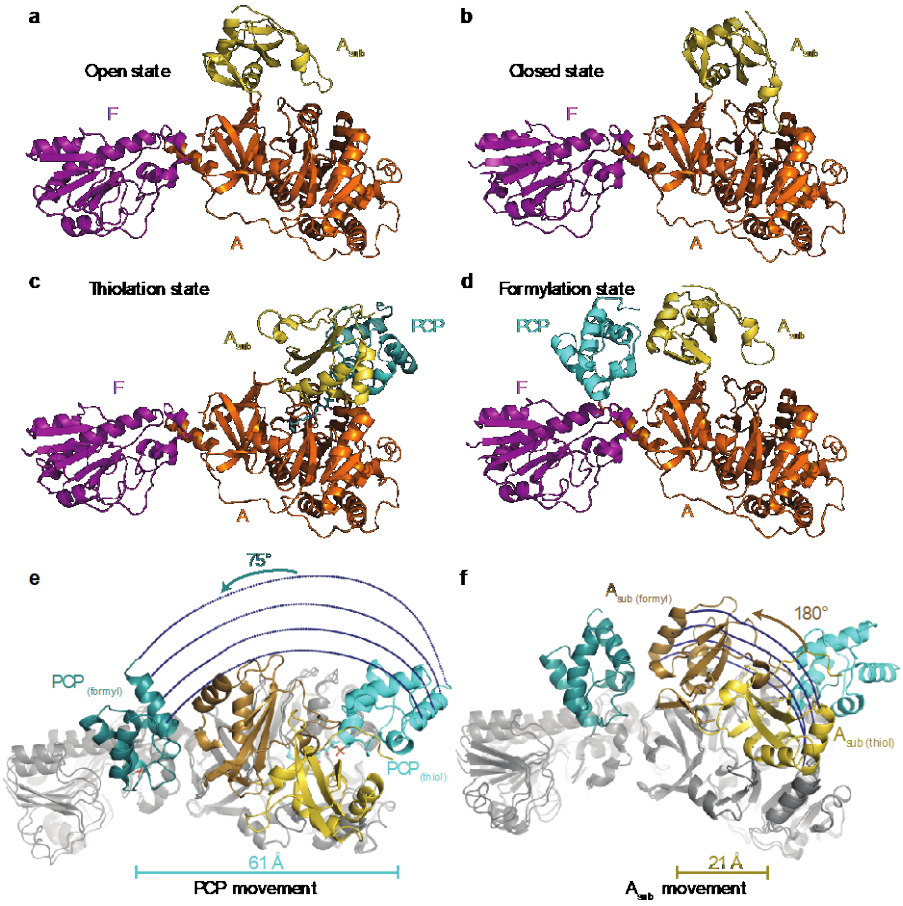


Figure 2: Crystal structures representing the steps of the synthesis cycle in the LgrA initiation module

The F-A-PCP LgrA initiation module in **a,** open **b,** closed **c,** thiolation and **d,** formylation states. (The PCP domain is not necessary for the open and closed states and is disordered in **b** and **c**.) The transition between thiolation and formylation states requires massive rigid body movements of both the A_{sub} and PCP domains. **e,** The PCP domain rotates 75° and translocates its center of mass by 61 Å. PPE arm attachment point, Ser729, moves 52 Å, and some residues move more than 80 Å. **f,** The A_{sub} domain rotates 180° and translocates its center of mass by 21 Å.

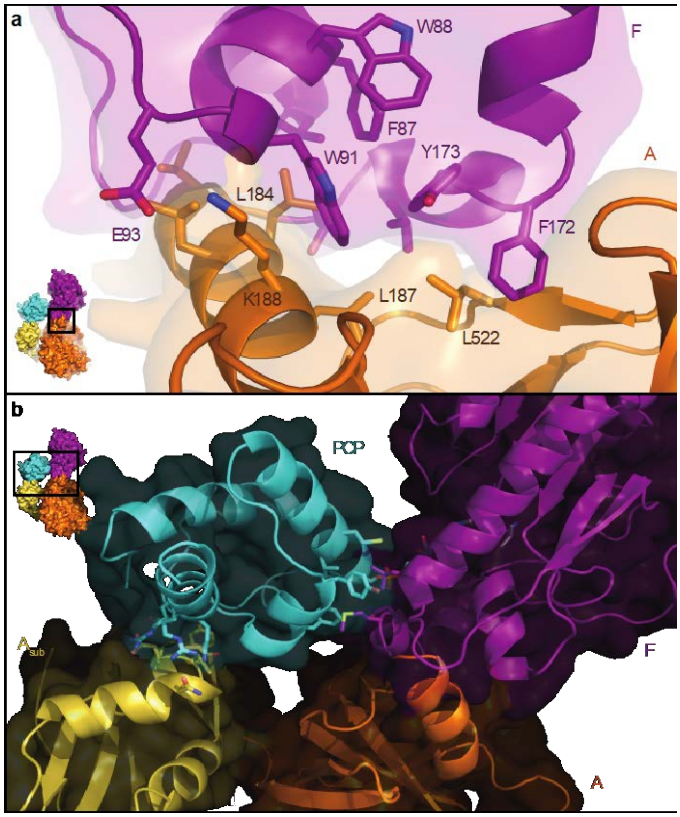


Figure 3: Interdomain interfaces of the initiation module.

a, The F domain is fused onto the A domain and forms a small hydrophobic core (Extended Data Fig. 8). **b,** Interaction of the PCP domain with A_{sub} and F domains in the formylation state. The A_{sub} domain creates an electrostatic platform for the PCP domain. The PCP domain binds to F domain hydrophobic residues Leu127 (often Lys or Glu in FTs) and Met178 (in the very C-terminus of the F domain that is not similar to FTs). The PPE phosphate interacts with Arg170 (often Glu, Ser or Asn in FTs) and Asn177 (usually Glu, Asp or Met in FTs).

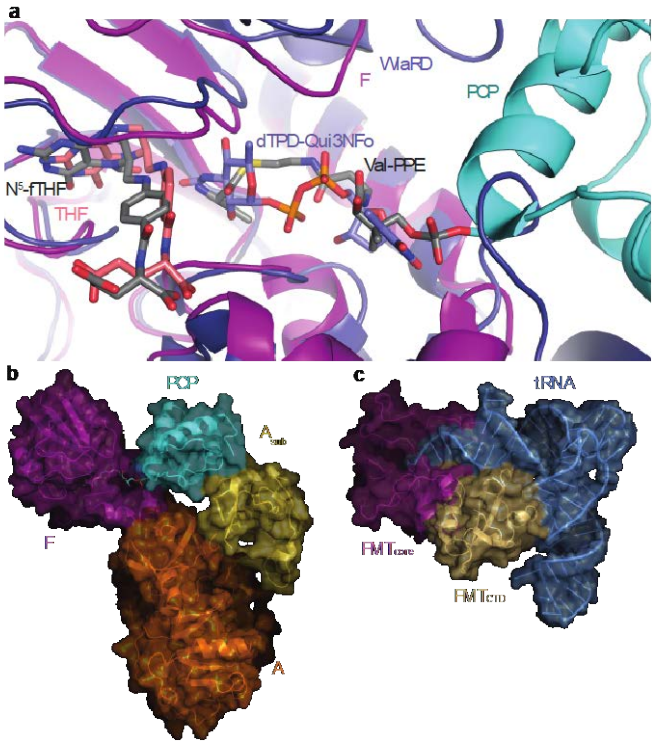


Figure 4: Comparisons of the F domain to sugar and tRNA FTs.

a, The binding mode for the PPE arm to the F domain is similar to that of sugar-dTDP in sugar FTs (protein WlaRD, PDB 4LY3[15]). Note that the valine and most of the PPE arm (carbons shown in grey) are modelled, as they are not visible in electron density maps at 3.8 Å resolution. **b,** The A_{sub} domain emulates the positioning role of the FMT_{CTD} in methionyl-$tRNA_f^{Met}$ formyltransferases (PDB 2FMT[13]). Excluding the PPE arm, the PCP domain buries only 279 $Å^2$ of F domain surface. The A_{sub} provides an additional 345 $Å^2$ of interaction surface to position the PCP domain.

Methods

Cloning of linear gramicidin synthetase initiation module constructs

Genomic DNA was isolated from *Brevibacillus brevis* ATCC 8185 (Cedarlane Laboratories) using a GenElute Bacterial Genomic DNA Kit (Sigma-Aldrich). Gene constructs comprising F and A domains (F-A) and all three domains (F-A-PCP) were amplified by PCR from the *lgrA* gene using the following primers, designed using sequence alignment with A and PCP domains of known structure and the study of Marahiel and coworkers: FA_fwd AATCATCCATGGGAAGAATACTATTCCTAACAACATTTATGAGCAAAG; FA_rev AATCATCTCGAGTTACGCATCGGCCTGCACGTCT; FAT_fwd TGACTACCATGGGGAGAATACTATTCCTAACAACATTTATGAGC; FAT_rev CGTTGAGCGGCCGCTTGCTCCGTAAGCAGACGTTT. PCR product for F-A-PCP was digested using NcoI and NotI (New England Biolabs) and ligated into a pET21-derived vector containing an N-terminal octa-histidine tag with a tobacco etch virus (TEV) protease cleavage site. PCR product for F-A-PCP was cloned between NcoI and NotI restriction sites into a **pET21**-derived vector containing an N-terminal TEV cleavable octa-histidine tag and a C-terminal TEV cleavable calmodulin binding peptide (CBP) tag. Point mutations were introduced into the construct of F-A-PCP using the QuikChange (Agilent) site-directed mutagenesis kit.

Expression and purification of proteins

The F-A protein was expressed in *E. coli* BL21 (DE3) cells. A 10 ml aliquot of overnight culture was used to inoculate 1L of LB medium supplemented with 350 μg ml^{-1} kanamycin. The culture was grown at 37 °C to an OD_{600} of 0.6, before inducing protein expression using 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and reducing the temperature to 16 °C for 18h. Cells were harvested by centrifugation at 4 °C and resuspended in nickel binding buffer (2 mM imidazole, 150 mM NaCl, 0.25mM TCEP, 50 mM Tris-HCl, pH 7.0). The cells were lysed by sonication on ice and centrifuged for 30 min at 20 000 x *g* at 4 °C. Clarified lysate was loaded onto a HiTrap IMAC FF column (GE Healthcare). F-A was eluted using a gradient of 2-250 mM imidazole. Fractions containing F-A were pooled, diluted 10-fold with ion exchange binding buffer (0.25 mM TCEP, 20 mM Tris, pH 8.0), loaded onto a HiTrap Q HP column and eluted using a gradient to 100% elution buffer (1M NaCl, 0.25 mM TCEP, 20 mM Tris-HCl, pH 8.0). The eluted protein was

concentrated using a 10k MWCO Amicon Ultra-15 filtration unit (EMD Millipore) and subjected to gel filtration chromatography using a HiLoad 16/600 Superdex 200 column (GE Healthcare) equilibrated with S200 buffer (150 mM NaCl, 0.25 mM TCEP, 20 mM Tris, pH 7.0). Protein purity was confirmed using SDS-PAGE and native PAGE. Pure F-A was concentrated in storage buffer (25% glycerol, 150 mM NaCl, 0.25 mM TCEP, 20 mM Tris, pH 7.0), flash frozen with liquid nitrogen and stored at -80 °C for later use.

F-A-PCP was expressed in *E. coli* BL21 EntD-(DE3) cells using the same protocol as above. Cells were pelleted, resuspended in CBP binding buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM imidazole pH 8.0, 2 mM $CaCl_2$, 2 mM β-mercaptoethanol (βME), and 0.1 mM phenylmethanesulfonyl fluoride (PMSF)), sonicated and clarified by centrifugation for 30 min at 20 000 x *g* at 4 °C. Clarified lysate was loaded onto a 30 ml calmodulin sepharose 4B column (GE Healthcare). F-A-PCP was eluted with elution buffer (25 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM EGTA, 2 mM βME, and 0.1 mM PMSF). For biochemical assays of F-A-PCP and mutants, protein was pooled, concentrated in storage buffer, flash-frozen in liquid nitrogen and stored at -80 °C. For crystallographic studies, F-A-PCP was further purified as follows. Protein was dialyzed against binding buffer for a minimum of 4 hours before being loaded onto a 5 ml HiTrap IMAC FF column (GE Healthcare) charged with Ni^{2+} and equilibrated in nickel binding buffer. F-A-PCP was eluted using a 60 mL gradient of 0 – 250 mM imidazole. Fractions containing F-A-PCP were pooled and affinity tags were removed by cleavage with TEV protease at room temperature overnight using a 1:4 mg ratio of TEV to F-A-PCP. Cleaved F-A-PCP was passed back over the nickel and calmodulin affinity columns, with the flow-through collected, concentrated, and applied to a HiLoad 16/600 Superdex 200 (GE Healthcare) in S200 buffer. Pure F-A-PCP was concentrated to 5.0 mg ml^{-1} in storage buffer, flash-frozen in liquid nitrogen and stored at -80 °C.

Substrate syntheses

Amino-coenzyme A (amino-CoA)[23] was prepared enzymatically starting from amino-pantetheine (WuXi AppTec) using a previously published protocol[31] with the following modifications: one-pot synthesis was carried out at pH 9.0, the amounts of DPCK and ATP were doubled to 9.8 mg and 30 mM, respectively; and the enzymes were removed using a 10k MWCO Amicon Ultra-15 filtration unit (EMD Millipore).

An ATP regeneration system using 0.1 mg/ml pyrophosphatase (Roche), 30 mM phosphoenolpyruvate, and 0.1 mg/ml pyruvate kinase (Roche) was also included. The filtrate containing amino-CoA was purified on a preparative reverse-phase C18 HPLC (35 ml/min; 0-4 min, 0% B; 4-9 min, 0-98% B, where A is 0.1% trifluoroacetic acid (TFA, Sigma-Aldrich) in H_2O and B is 0.1% in acetonitrile (ACN, Sigma-Aldrich)). Amino-CoA eluted at 7 min and was lyophilized to dryness.

Valine-amino-coenzyme A (Val-NH-CoA)[23] was synthesized by coupling 1 molar equivalent of amino-CoA with 8 molar equivalents of tert-Butoxycarbonyl-L-valine-N- hydroxysuccinimide ester (Boc-Val-OSu, Sigma-Aldrich) in N,N-dimethylformamide (DMF, Sigma-Aldrich) with 4 molar equivalents of N,N-diisopropylethylamine (DIPEA, Sigma-Aldrich) overnight with stirring. Boc-Val-NH-CoA was purified using the above chromatographic profile and lyophilized to dryness, then deprotected using 1.5 ml 95% TFA/2.5% H2O/2.5% triisopropylsilane (TIPS, Sigma-Aldrich). The deprotection mix was agitated for 2 h at 25 °C in a thermomixer at 700 rpm before being transferred to 20 ml ice-cold diethyl ether and incubated at -20 °C for 2 h. The solution was centrifuged and the pellet was redissolved in 5% aqueous ACN solution and purified with the same protocol as amino-CoA. Compound identity was verified by mass spectrometry and NMR (Supplemental Data 1).

Loading phosphopanetheinylates on the PCP domain

Unmodified F-A-PCP was converted to valine-NH-F-A-PCP by incubating 25 μM apo-F-A-PCP with 5 μM Sfp, 0.25 mM valine-NH-CoA, 10 mM $MgCl_2$ and 25 mM Tris pH 7.0 for a minimum of 4 h at 25 °C. To remove Sfp for subsequent crystallization trials, the reaction mix was loaded onto a Superdex S75 10/300 GL (GE Healthcare Life Sciences) equilibrated in 25 mM Tris pH 7.5, 150 mM NaCl, and 2 mM βME.

SAXS

Inline SEC-SAXS data was collected on the G1 beamline at the Macromolecular Diffraction Facility at the Cornell High Energy Synchrotron Source[32,33] at 9.963 keV (1.244 A) at $7.89x10^{11}$ photons/s. The X-ray beam was collimated to 250x250 μm and the sample cell path length was 2 mm. The G1 beamline was outfitted with a GE AKTA purifier with a GE Superdex 200 5/150 GL

column and 50 µl sample loop. The column was equilibrated in 25 mM Tris pH 7.5, 150 mM NaCl, and 2mM BME and the samples were centrifuged for 10 minutes prior to sample injection. Images were recorded on a Pilatus 100K-s detector and normalized using beam stop photodiode counts. F-A-PCP eluted in a single monomeric peak and eleven peak exposures were averaged using BioXTAS RAW software[34]. A buffer scattering curve was created by averaging the first eleven exposures after injection, and this scattering curve was subtracted from the F-A-PCP scattering curve to yield the corrected scattering curve for F-A-PCP. *Ab initio* models were generated by first creating pairwise distribution functions (P(r)) with GNOM[35], leading to twenty independent bead models produced by DAMMIF[36]. Models were aligned, averaged, and filtered using DAMAVER[37] assuming P1 symmetry. All DAMMIF models were included in the final DAMAVER model. They had a mean NSD value of 0.82 +/- 0.052. CRYSOL[38] was used to check how well the final model fit with our crystal structures. Flexibility was analyzed using EOM[39,40], whereby crystal structures of F, A_{core}, A_{sub}, and the PCP were used to generate a pool of 10,000 models.

Crystallography

To obtain the crystal structures described in this study, genes from 4 species, of up to 4 domain constructs each (F, F-A, F-A(ΔA_{sub}), F-A-PCP) were cloned and assayed for heterologous expression. Purification was undertaken for all well-expressing proteins and crystallization trials were performed, including trials using protein with affinity tags removed or retained, and in the presence or absence of a variety of ligands (ATP, AMPcPP, AMP, valine, THF, N^5-fTHF, phosphopantetheine, valine amino phosphopantetheine, valine vinyl sulfonamide adenylate, dead-end THF analog). Up to 4032 crystallization conditions were assayed per protein sample, and gave a total of ~50 “hits”, 6 of which were successfully optimized to allow structure determination. Together, 4 of these crystal structures (F-A in crystals of space group $P4_12_12$, F-A-PCP in R3:H, F-A-PCP-PPE-NH-Val in $P2_1$ and F-A-PCP-PPE in $P3_22$), plus an additional structure including ligands soaked into F-A $P4_12_12$ crystals, captured the states that represent every major step of the assembly-line synthesis in the LgrA initiation module and are presented here.

The final crystallization conditions were optimized in 24-well sitting drop plates, with 2 µl protein sample plus 2 µl reservoir solution in the drop and a 500µl

reservoir volume and are as follows: “F-A” and “F-A soak”: protein LgrA F-A (10 mg/ml) was crystallized using a precipitant solution of 2 M Na-formate, 0.1 M Na-Acetate pH 5.3 into space group $P4_12_12$. “F-A-PCP” (open and closed states): protein LgrA F-A-PCP (5 mg/ml) was crystallized using a precipitant solution of 0.92 M $AmSO_4$, 0.1 M bis-Tris pH 5.5, 1% PEG 3350 into space group R3:H. “F-A-PCP-NH-Val” (thiolation state): protein F-A-PCP-PPE-NH-Val (4.7 mg/ml) was crystallized using a precipitant solution of 12% PEG 20 000, 0.1 M MES pH 6.7 into space group $P2_1$. “F-A-PCP-PPE” (formylation state): protein F-A-PCP-PPE (5.5 mg/ml) was crystallized using a precipitant solution of 1 M $AmSO_4$, 0.1 M bis-Tris pH 5.5, 3% PEG 3350 into space group $P3_22$.

Solutions of mother liquor with increasing amounts of glycerol (5%, 10%, 25%) were used to replace the drop solution for cryoprotection. For soaking with the N^5-fTHF, valine and AMPcPP, 10mM of each was included in the final cryoprotection solution and incubated for 30 min. (Marahiel and coworkers showed that LgrA uses commercially available N^5-fTHF in addition to its natural substrate, N^{10}-fTHF[6].) Crystals were flash-cooled in liquid nitrogen and diffraction data sets collected at 200 K using beamline 8 of the CMCF at the Canadian Light Source (λ= 0.979 Å) in Saskatoon, Canada.

All datasets were integrated and scaled using the programs HKL-2000[41] and iMosflm[42]. Structure determination of F-A in the $P4_12_12$ space group was performed by molecular replacement using a search model of the A domain from gramicidin Soviet synthetase (1AMU[17]; note that linear gramicidin and gramicidin Soviet are made by different NRPSs) with the A_{sub} subdomain removed and side chains trimmed to the beta carbon, in the program Phaser[43]. Density for the F domain was visible in the resulting maps. Iterative building in the program COOT[44] and refinement in the program Phenix[45] produced the final F-A structure. This structure was then used as a search model to determine the structure of F-A-PCP in space groups $P3_22$, R3:H, and $P2_1$ by molecular replacement using the program Phaser, followed by iterative building in the program COOT and refinement in the programs Phenix and CNS[46]. The highest resolution shell CC* values are: $P4_12_12$ – 0.845; $P4_12_12$ (soak) – 0.897; $P3_22$ – 0.883; R3:H – 0.822, and $P2_1$ – 0.822. The quoted resolution of each structure represents the CC 1/2 of the diffraction data[42,47].

Bioinformatics

Multiple sequence alignments (MSAs) were constructed using Clustal Omega[48] (ebi.ac.uk/Tools/msa/clustalo) and PROMALS3D[49] (prodata.swmed.edu/promals3d), following database searches using BLAST[50] (ncbi.nlm.nih.gov/blast). MSAs were drawn/edited using Jalview[51] (www.jalview.org). PHYLIP (evolution.genetics.washington.edu) was used to make neighbour-joining trees bootstrapped with 100 replicates, and FigTree (tree.bio.ed.ac.uk) was used to draw them. WebLogo[52] (weblogo.berkeley.edu) was used to draw sequence logos of residue groupings of interest. AmiGO[53] (amigo.geneontology.org) was used to check for experimentally characterized proteins.

Online-only references

31 Nazi, I., Koteva, K. P. & Wright, G. D. One-pot chemoenzymatic preparation of coenzyme A analogues. *Anal Biochem* **324**, 100-105 (2004).

32 Acerbo, A. S., Cook, M. J. & Gillilan, R. E. Upgrade of MacCHESS facility for X-ray scattering of biological macromolecules in solution. *Journal of Synchrotron Radiation* **22**, 180-186 (2015).

33 Skou, S., Gillilan, R. E. & Ando, N. Synchrotron-based small-angle X-ray scattering of proteins in solution. *Nat. Protocols* **9**, 1727-1739 (2014).

34 Nielsen, S. S. *et al.* BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis. *Journal of Applied Crystallography* **42**, 959-964 (2009).

35 Svergun, D. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography* **25**, 495-503 (1992).

36 Franke, D. & Svergun, D. I. DAMMIF, a program for rapidab-initioshape determination in small-angle scattering. *Journal of Applied Crystallography* **42**, 342-346 (2009).

37 Volkov, V. V. & Svergun, D. I. Uniqueness ofab initioshape determination in small-angle scattering. *Journal of Applied Crystallography* **36**, 860-864 (2003).

38 Svergun, D., Barberato, C. & Koch, M. H. J. CRYSOL - A program to evaluate x-ray solution scattering of biological macromolecules from atomic coordinates. *Journal of Applied Crystallography* **28**, 768-773 (1995).

39 Tria, G., Mertens, H. D. T., Kachala, M. & Svergun, D. I. Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ* **2**, 207-217 (2015).

40 Bernado, P., Mylonas, E., Petoukhov, M. V., Blackledge, M. & Svergun, D. I. Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* **129**, 5656-5664 (2007).

41 Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* **276**, 307-326 (1997).

42 Leslie, A. G. W. & Powell, H. R. Processing diffraction data with MOSFLM. *Nato Sci Ser Ii Math* **245**, 41-51 (2007).

43 McCoy, A. J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).

44 Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallographica Section D-Biological Crystallography* **66**, 486-501 (2010).

45 Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213-221 (2010).

46 Brunger, A. T. Version 1.2 of the crystallography and NMR system. *Nat Protoc* **2**, 2728-2733 (2007).

47 Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta crystallographica Section D, Biological crystallography* **69**, 1204-1214 (2013).

48 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

49 Pei, J. & Grishin, N. V. PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol Biol* **1079**, 263-271 (2014).

50 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410 (1990).

51 Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189-1191 (2009).

52 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190 (2004).

53 Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25-29 (2000).

54 Allen, C. L. & Gulick, A. M. Structural and bioinformatic characterization of an Acinetobacter baumannii type II carrier protein. *Acta Crystallographica Section D* **70**, 1718-1725 (2014).

55 Lai, J. R., Fischbach, M. A., Liu, D. R. & Walsh, C. T. A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. *Proc Natl Acad Sci U S A* **103**, 5314-5319 (2006).

56 Jaremko, M. J., Lee, D. J., Opella, S. J. & Burkart, M. D. Structure and substrate sequestration in the pyoluteorin type II peptidyl carrier protein PltL. *J Am Chem Soc* **137**, 11546-11549 (2015).

57 Goodrich, A. C., Harden, B. J. & Frueh, D. P. Solution structure of a nonribosomal peptide synthetase carrier protein loaded with Its substrate reveals transient, well-defined contacts. *J Am Chem Soc* **137**, 12100-12109 (2015).

58 May, J. J., Kessler, N., Marahiel, M. A. & Stubbs, M. T. Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases. *Proc Natl Acad Sci U S A* **99**, 12120-12125 (2002).

59 Du, L., He, Y. & Luo, Y. Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from Bacillus cereus. *Biochemistry* **47**, 11473-11480 (2008).

60 Drake, E. J., Duckworth, B. P., Neres, J., Aldrich, C. C. & Gulick, A. M. Biochemical and structural characterization of bisubstrate inhibitors of BasE, the self-standing nonribosomal peptide synthetase adenylate-forming enzyme of acinetobactin synthesis. *Biochemistry* **49**, 9292-9305 (2010).

61 Henderson, J. C. *et al.* Antimicrobial peptide resistance of Vibrio cholerae results from an LPS modification pathway related to nonribosomal peptide synthetases. *ACS Chem Biol* **9**, 2382-2392 (2014).

62 Tan, X. F. *et al.* Structure of the adenylation-peptidyl carrier protein didomain of the Microcystis aeruginosa microcystin synthetase McyG. *Acta Crystallogr D Biol Crystallogr* **71**, 873-881 (2015).

63 Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem Biol* **19**, 188-198 (2012).

64 Sundlov, J. A. & Gulick, A. M. 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* **69**, 1482-1492 (2013).

65 Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A. & Holak, T. A. Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407-418 (2000).

66 Bruner, S. D. *et al.* Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301-310 (2002).

67 Koglin, A. *et al.* Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312**, 273-276 (2006).

68 Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L. O. Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781-792 (2007).

69 Frueh, D. P. *et al.* Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903-906 (2008).

70 Osman, K. T., Du, L., He, Y. & Luo, Y. Crystal structure of Bacillus cereus D-alanyl carrier protein ligase (DltA) in complex with ATP. *J Mol Biol* **388**, 345-355 (2009).

71 Liu, Y., Zheng, T. & Bruner, S. D. Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem Biol* **18**, 1482-1488 (2011).

72 Tufar, P. *et al.* Crystal structure of a PCP/Sfp complex reveals the structural basis for carrier protein posttranslational modification. *Chem Biol* **21**, 552-562 (2014).

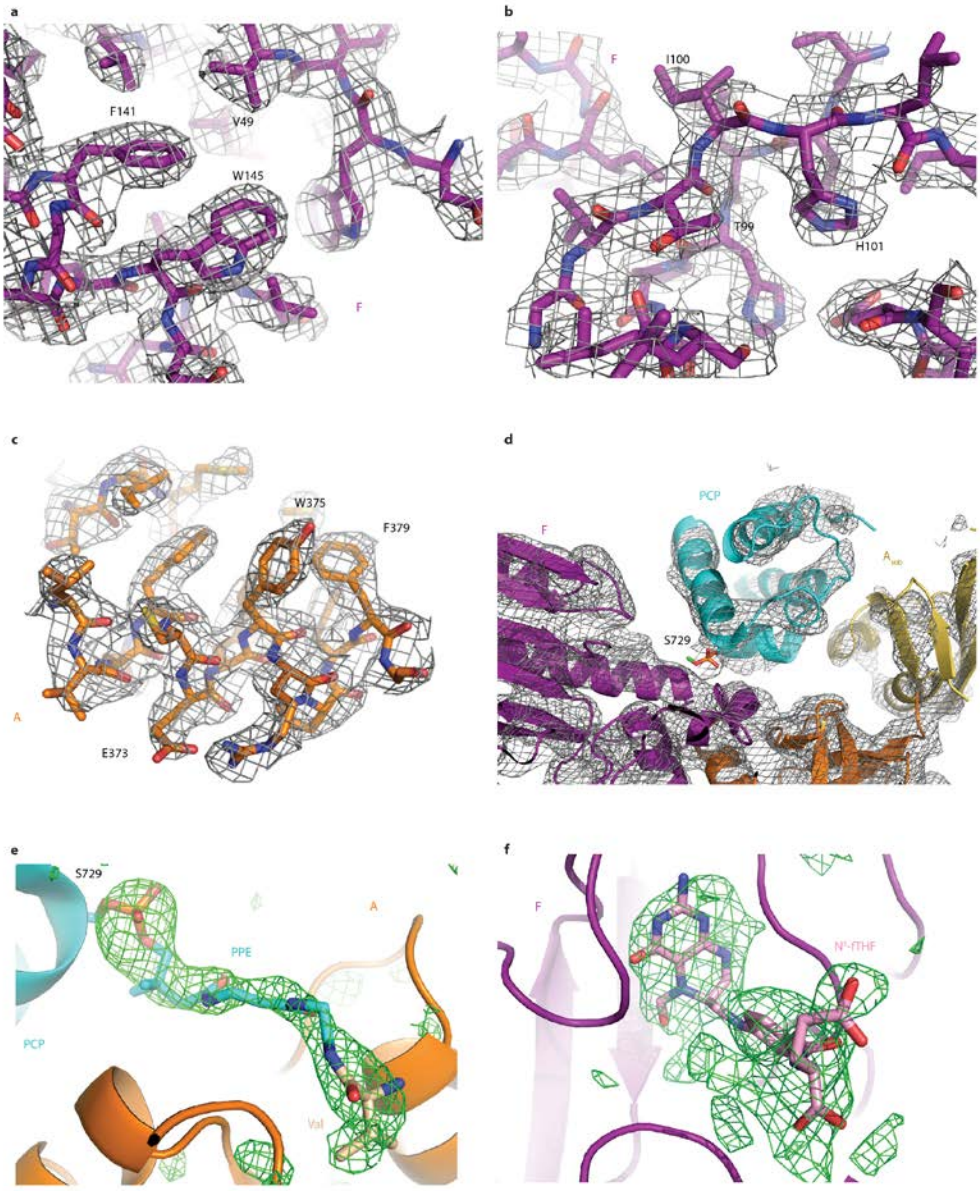
73 Haslinger, K., Redfield, C. & Cryle, M. J. Structure of the terminal PCP domain of the non-ribosomal peptide synthetase in teicoplanin biosynthesis. *Proteins* **83**, 711-721 (2015).

Extended and Supplemental Data

Extended Data Table 1: Crystallographic statistics.

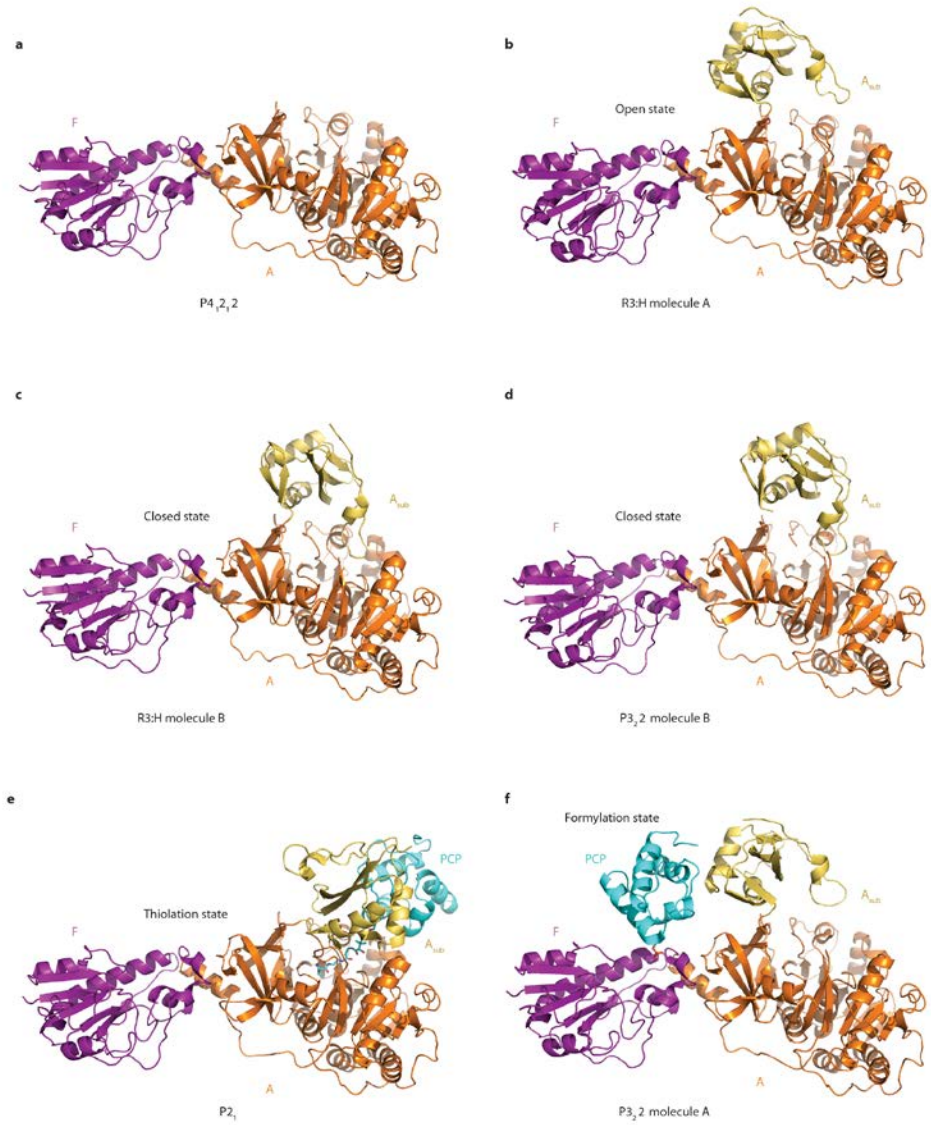
Extended Data Fig. 1: Synthetic cycles in canonical initiation, canonical elongation and LgrA initiation modules

Schematic diagrams comparing the synthetic cycle in canonical initiation and elongation modules with that in the LgrA initiation module.



Extended Data Fig. 2: Representative electron density.

$2F_O$-F_C density maps for protein in **a,** $P4_12_12$, **b,** R3:H, **c,** $P2_1$ and **d,** $P3_22$ crystal forms contoured at 1σ. Unbiased F_O-F_C density maps for **e**, the PPE-NH-Val arm in the $P2_1$ (thiolation state) contoured at 3.3σ and **f,** a $P4_12_12$ crystal soaked with N^5-fTHF, AMPcPP and valine contoured at 2.5σ.

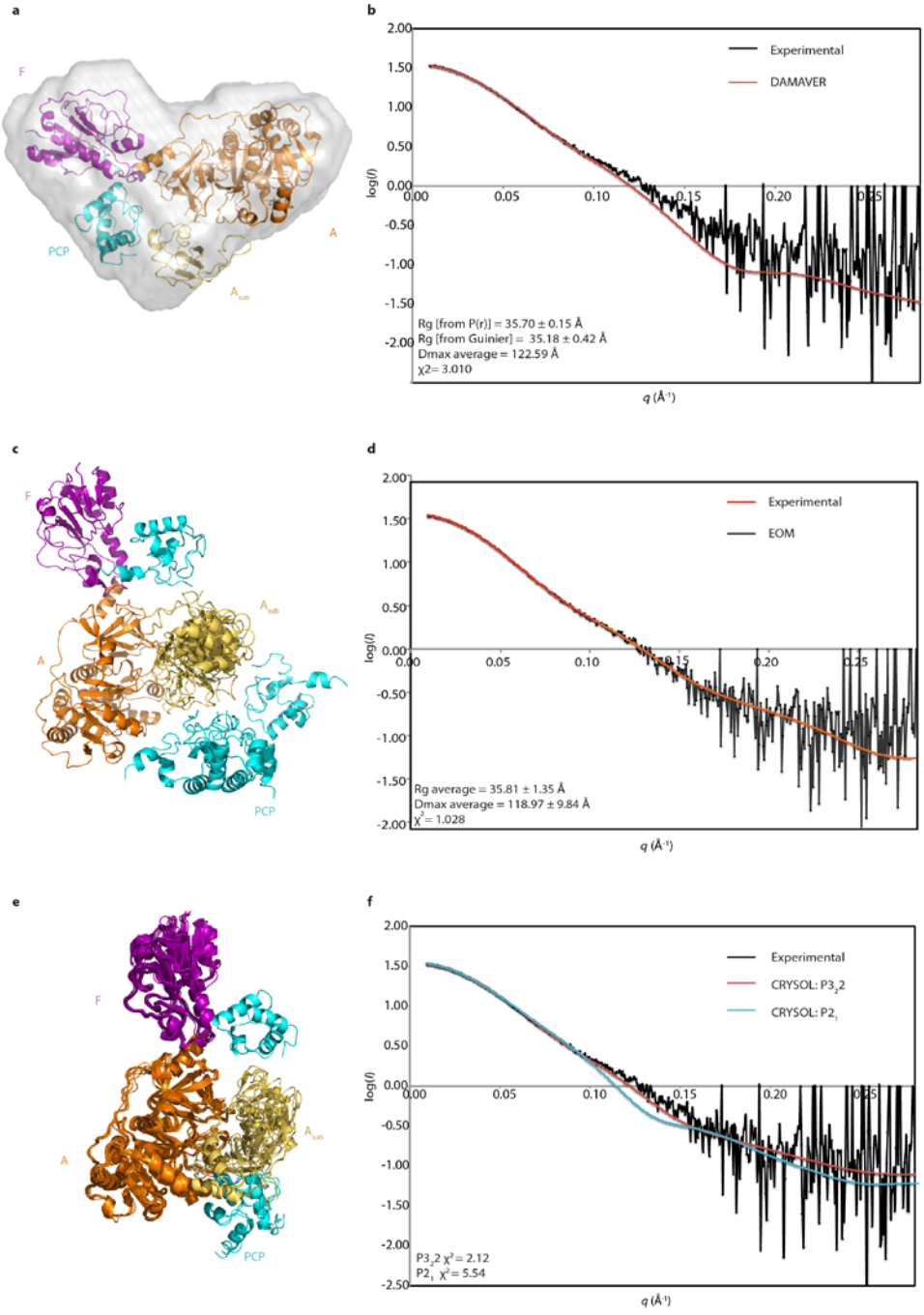


Extended Data Fig. 3: Crystal structures of the initiation module of linear gramicidin synthetase.

Models of **a,** F-A (A_{sub} disordered), **b, c, d,** F-A-PCP (PCP disordered) and **e, f,** F-A-PCP from the four independent crystals structures determined. The crystal with space group $P3_22$ diffracted anisotropically to ~3.8 Å resolution, but the other higher resolution structures enabled the building of high quality models shown in **d** and **f**.

Extended Data Fig. 4: Comparison between the LgrA initiation module and the SrfAC termination module.

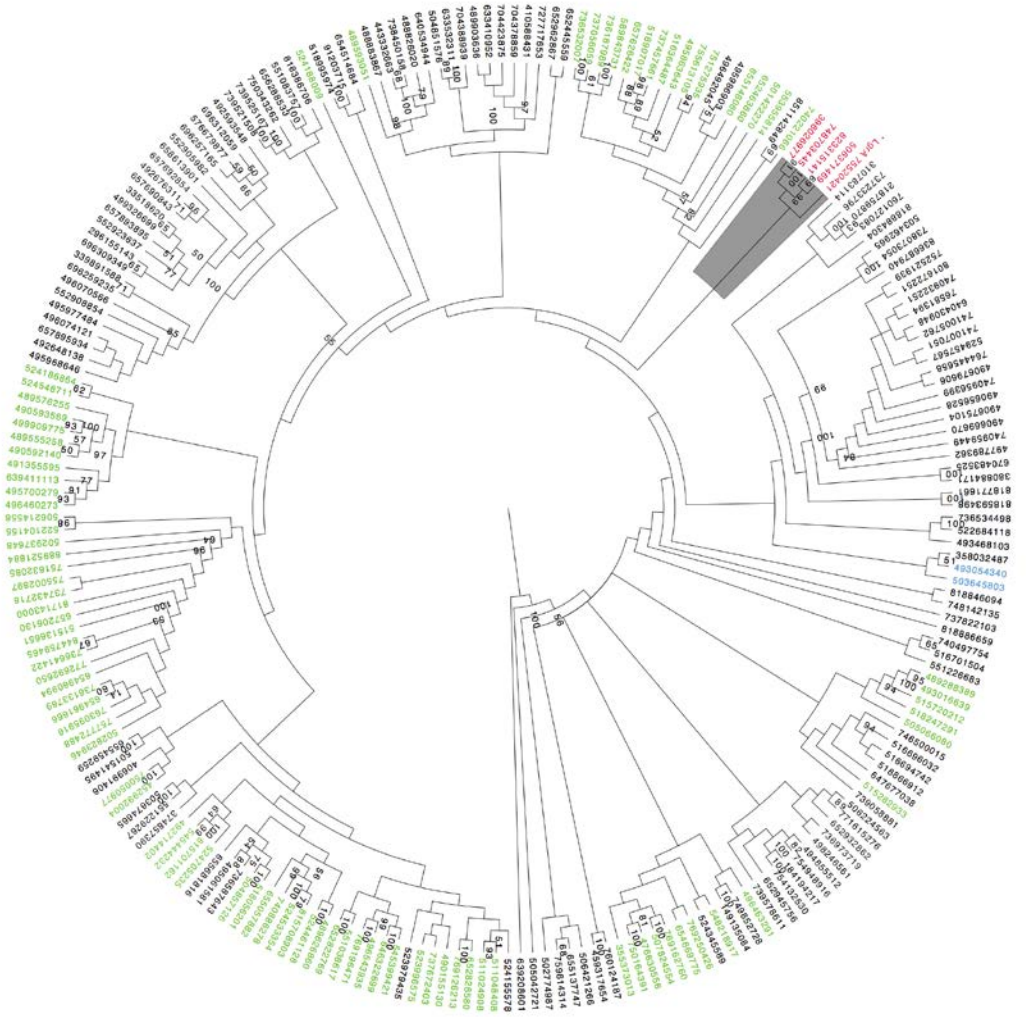
a, The LgrA initiation module in the formylation state and **b,** the SrfAC termination module[22] in the state where aminoacyl-PCP would be positioned to act as an acceptor substrate in the condensation reaction (PPE arm not present). The F and C domains are each positioned directly N-terminal of their A domains and bury similar amounts of A domain surface area (829 $Å^2$ and 903 $Å^2$; contributing residues shown in spheres), each forming "stable platforms"[22]. Both modules use very large movements of their PCP and A_{sub} domains to bring the module's aminoacyl-PCP to distant active sites to act as the acceptor substrate in an amide bond forming reaction. However, the F-A and C-A interfaces are distinct, and **c,** if the A domains are superimposed, the F and C domains are only partially overlapping. This places their active sites in dissimilar locations, necessitating that A_{sub} and PCP assume very different positions to deliver their substrate. The PCP domain in the formylation state completely overlaps with the position of the C domain.



Extended Data Fig. 5: SAXS analysis of F-A-PCP.

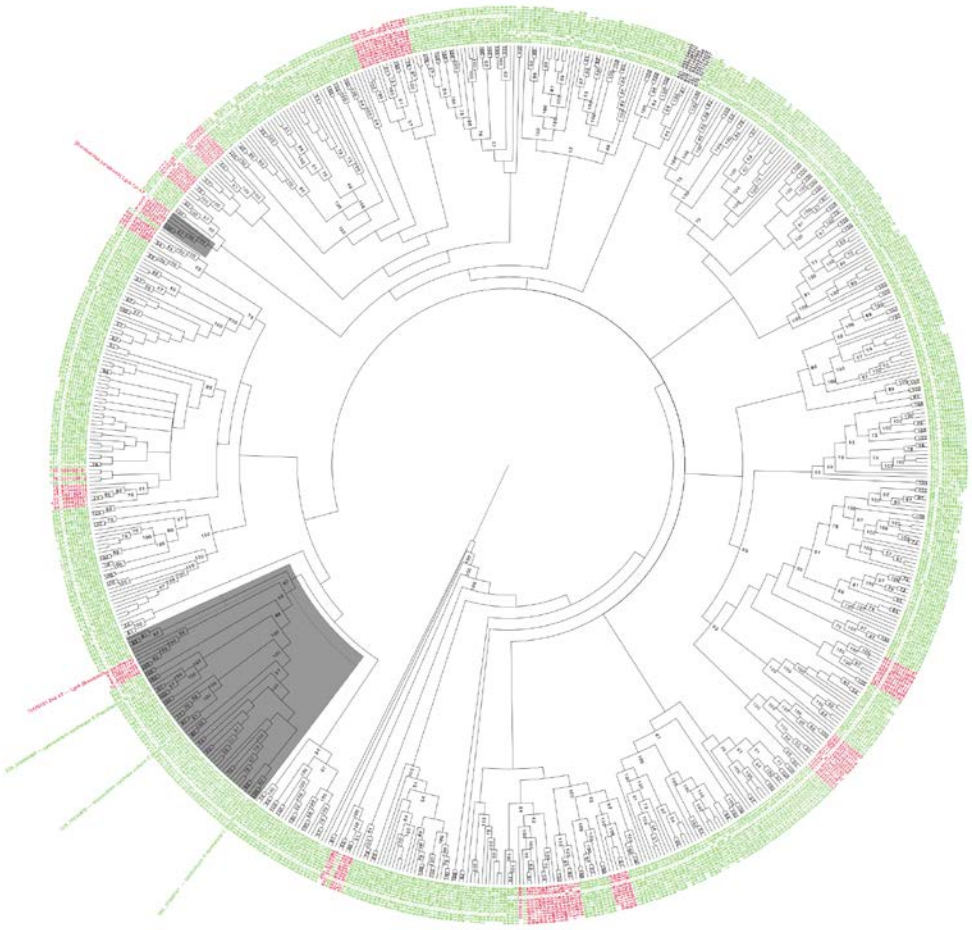
a, The crystal structure in the formylation state is shown superposed on the averaged filtered *ab initio* SAXS model generated with DAMAVER[37], with a NSD value of 0.819±0.052. **b,** The calculated scattering curve for the DAMAVER is overlaid with the experimental scattering with χ^2=3.010, where *I* represents scattering intensity and

q is equivalent to $4\pi\sin(\theta)/\lambda$. **c,** To better understand the flexibility of F-A-PCP, EOM[39] was performed and generated 5 different ensembles. The ensemble resembling the formylation state structure represented over 60% of the optimized models generated, while the remaining <40% resembled the thiolation state structure. **d,** The calculated scattering of the EOM model has a χ^2=1.028 which demonstrates that F-A-PCP has flexibility. The data are consistent with extreme flexibility for A_{sub} and PCP domains, and limited flexibility in F-A_{core}: **e,** All independent molecules from the crystal structures were overlaid to further illustrate the flexibility of the system. **f,** CRYSOL[38] was used to generate predicted scattering curves for the formylation state and thiolation state crystal structures with χ^2= 2.12 and χ^2=5.54, respectively.



Extended Data Fig. 6: Neighbour-joining tree of LgrA F domain and homologs. This neighbour-joining tree of the LgrA F domain and homologs was made using PHYLIP (evolution.genetics.washington.edu) based on an initial Clustal Omega[48]

alignment of the closest 220 homologs of the LgrA F domain (Blast[50] BLAST E-value <1e-14). The most similar FTs to the F domain share ~45% identity, and all of these 220 FTs have only inferred function. The tree was drawn using the program FigTree(tree.bio.ed.ac.uk). The sequences are named with their GI accession codes. Colouring is: red for *Brevibacilli*, green for other Firmicutes, black for other bacteria and blue for Archaea. The clade of the LgrA F domain is highlighted in grey. Only nodes with bootstraps of >50% are shown. Several horizontal transfer events are evident where Firmicute and non-Firmicute proteins cluster together with high bootstrap values (e.g.,>70%). The several horizontal transfer events of FT domains between Firmicutes and other bacterial groups suggest the LgrA F domain likely originated from horizontal transfer.



Extended Data Fig. 7: Neighbour-joining tree of LgrA A-PCP and homologs. This neighbour-joining tree of LgrA A-PCP didomains and homologs was made for the 500 closest homologs (BLAST E-value <1e-14). The sequences are named with

their GI accession codes. Colouring is: red for *Brevibacilli*, green for other Firmicutes, black for other bacteria. The significant clades of the LgrA A-PCP domains are highlighted in grey. Only nodes with bootstraps of >50% are shown. Three functionally characterized homologs of LgrA that are shown to be directly related are labelled. The A-PCP portion of the initiation module is quite divergent, but the second module of LgrA clearly shares a common origin with functionally characterized NRPSs in Bacilli and other Firmicutes.

Extended Data Fig. 8: Conservation and variation of residues involved in the interaction interfaces.

Sequence logos made using the WebLogo server http://weblogo.berkeley.edu [52] show conservation and variation as found in multiple sequence alignments of **a,** F domain residues that interact with the A domain and **b**, A domain residues that interact with the F domain. Below each logo are the corresponding residues in the LgrA proteins from the 5 *Brevibacillus* species, with the crystallized LgrA on the first line. Sequence logos indicate the conservation and variation in F residues involved in binding and interaction with PCP-PPE-Val across **c**, the closest 240 homologs of LgrA and **d**, all of the functionally or structurally characterized formyltransferase proteins (reduced

for redundancy so that no two sequences have >50% sequence identity). **e**, Consensus sequences for the 5 *Brevibacillus* LgrA homologs and for the formyltransferases of known structure for each of three formyltransferase types. Catalytic residues are His73, Asn71, and Asp108.

Extended Data Fig. 9: Interaction surfaces in PCP and A_{sub} domains.

The **d,** A_{sub} and **e,** PCP[2,25,54] domains must maximize the use of their limited surfaces to interact with their many binding partners. Shown are the surfaces observed in this study, and many excellent previous studies have also documented interaction surfaces biochemically or structurally. This includes, for example, the equivalent of PCP domain residues Met249, Phe264, and Ala268, which are required for interaction with the C domain in the acceptor site[55] and form hydrophobic interactions with the C domain[22] in a very similar manner and using an overlapping surface, as the PCP domain does to interact with the F domain. Furthermore, partially overlapping surfaces in PCP domains have been proposed to interact with their (acyl-)PPE arm to protect thioester intermediates[56] or to promote binding to the appropriate partner domain[57]. These interactions might occur during PCP domain transit, but they would have to be broken before productive binding to partner domains: Several of these PPE interactions are incompatible with the productive domain-domain interactions[57], and in catalytic configurations seen here and previously, the PPE arms extend into the partner domain and make little contact with the PCP domain.

Supplemental Video 1: Crystal structures of the initiation module of linear gramicidin synthetase

Supplemental Video 2: Animation of the synthetic cycle of the initiation module of linear gramicidin synthetase

The structures of the initiation module of LgrA interpolated to visualize its entire synthetic cycle. The A domain is open for substrate binding (see also A domain structures[2,20]), and closes to catalyze valine adenylation formation (see also A domain structures[2,18]). Next, the A_{sub} domain rotates 140° to catalyze transfer of the valine onto the thiol of the PPE arm of the PCP domain (see also A-PCP didomain structures[19,21]). Then the PCP domain transports its valine the 50 Å to the F domain to accept a formyl group. The PCP next moves the formyl-valine to the downstream elongation module, where it is passed off in that module's condensation reaction (not shown), which liberates the PCP to participate in the next round of reactions. The experimentally observed structures are indicated by crystallographic statistics (disordered or absent portions for each structure coloured grey). For this animation, details of the adenylate in the A domain and the valine in formylation active site were based on co-complexes of isolated A domains[58-62], A-PCP domains[21,63,64] and FT proteins[5,13,15,24]. The extensive body of NRPS structural biology of protein constructs that contain A and/or PCP domains was also checked[17,21,22,25,58-73].