β-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis

Nicole M. Gaudelli, Darcie H. Long & Craig A. Townsend

Non-ribosomal peptide synthetases are giant enzymes composed of modules that house repeated sets of functional domains, which select, activate and couple amino acids drawn from a pool of nearly 500 potential building blocks. The structurally and stereochemically diverse peptides generated in this manner underlie the biosynthesis of a large sector of natural products. Many of their derived metabolites are bioactive such as the antibiotics vancomycin, bacitracin, daptomycin and the β-lactam-containing penicillins, cephalosporins and nocardicins. Penicillins and cephalosporins are synthesized from a classically derived non-ribosomal peptide synthetase tripeptide (from β-(L-α-aminoacipetyl)-L-cysteinyl-D-valine synthetase). Here we report an unprecedented non-ribosomal peptide synthetase activity that both assembles a serine-containing peptide and mediates its cyclization to the critical β-lactam ring of the nocardicin family of antibiotics. A histidine-rich condensation domain, which typically performs peptide bond formation during product assembly, also synthesizes the embedded four-membered ring. We propose a mechanism, and describe supporting experiments, that is distinct from the pathways that have evolved to the three other β-lactam antibiotic families: penicillin/cephalosporin, clavams and carbapenems. These findings raise the possibility that β-lactam rings can be regio- and stereospecifically integrated into engineered peptides for application as, for example, targeted protease inactivators.

Despite their widespread use for more than half a century, the β-lactam antibiotics, represented most familiarly by the semi-synthetic penicillins and cephalosporins, remain the most frequently prescribed anti-infectives in human medicine. Four structurally distinct clans occur naturally, and the more recently discovered of these and their synthetic analogues such as the antibiotics vancomycin, bacitracin, daptomycin and the β-lactam-containing penicillins, cephalosporins and nocardicins. Penicillins and cephalosporins are synthesized from a classically derived non-ribosomal peptide synthetase tripeptide (from β-(L-α-aminoacipetyl)-L-cysteinyl-D-valine synthetase). Here we report an unprecedented non-ribosomal peptide synthetase activity that both assembles a serine-containing peptide and mediates its cyclization to the critical β-lactam ring of the nocardicin family of antibiotics. A histidine-rich condensation domain, which typically performs peptide bond formation during product assembly, also synthesizes the embedded four-membered ring. We propose a mechanism, and describe supporting experiments, that is distinct from the pathways that have evolved to the three other β-lactam antibiotic families: penicillin/cephalosporin, clavams and carbapenems. These findings raise the possibility that β-lactam rings can be regio- and stereospecifically integrated into engineered peptides for application as, for example, targeted protease inactivators.

Non-ribosomal peptide synthetases are giant enzymes composed of modules that house repeated sets of functional domains, which select, activate and couple amino acids drawn from a pool of nearly 500 potential building blocks. The structurally and stereocchemically diverse peptides generated in this manner underlie the biosynthesis of a large sector of natural products. Many of their derived metabolites are bioactive such as the antibiotics vancomycin, bacitracin, daptomycin and the β-lactam-containing penicillins, cephalosporins and nocardicins. Penicillins and cephalosporins are synthesized from a classically derived non-ribosomal peptide synthetase tripeptide (from β-(L-α-aminoacipetyl)-L-cysteinyl-D-valine synthetase). Here we report an unprecedented non-ribosomal peptide synthetase activity that both assembles a serine-containing peptide and mediates its cyclization to the critical β-lactam ring of the nocardicin family of antibiotics. A histidine-rich condensation domain, which typically performs peptide bond formation during product assembly, also synthesizes the embedded four-membered ring. We propose a mechanism, and describe supporting experiments, that is distinct from the pathways that have evolved to the three other β-lactam antibiotic families: penicillin/cephalosporin, clavams and carbapenems. These findings raise the possibility that β-lactam rings can be regio- and stereospecifically integrated into engineered peptides for application as, for example, targeted protease inactivators.

Despite their widespread use for more than half a century, the β-lactam antibiotics, represented most familiarly by the semi-synthetic penicillins and cephalosporins, remain the most frequently prescribed anti-infectives in human medicine. Four structurally distinct clans occur naturally, and the more recently discovered of these and their synthetic variants are of increasing importance to combat the rising spectre of antibiotic-resistant infectious diseases. Members of this group of antibiotics contain monomycyl and fused bicyclic β-lactams whose high energy, strained-ring skeletons are essential to their antimicrobial activities. Markedly different but chemically efficient biosynthetic pathways have evolved to each of the penicillin and cephalosporin (for example isopenicillin N and cephalosporin C), clavulanic acid and carbapenem (for example thienamycin) groups. Ironically, the fourth and structurally simplest clan of monomycyl β-lactams, exemplified by nocardicin G, has long remained an unsolved problem.

The nocardicin non-ribosomal peptide synthetase (NRPS) encompasses two megaenzymes, NocA and NocB, which together comprise five modules. Each module contains an adenylation (A) domain that binds ATP, selects its cognate building block and performs substrate acyl adenylation. The activated amino acid is then translocated as its aminoacyl thioester to the β-(L-α-aminoacipetyl)-L-cysteinyl-D-valine synthetase (ACVS). Isopenicillin N synthase (IPNS) catalyses oxidative β-lactam formation and bicyclization of ACV to form isopenicillin N with a single molecule of dioxygen and release of two molecules of water. Cephalosporin C is derived after isopenicillin N is epimerized to penicillin N and oxidative ring expansion occurs. The clavams and carbapenems are exemplified by clavulanic acid and thienamycin, respectively. Formation of the β-lactam ring that ultimately appears in clavulanic acid and thienamycin is catalysed by β-lactam synthetase (β-LS) and carbapenam synthetase (CPS), respectively, where transiently formed acyl adenylates are cyclized to β-lactam containing pathway intermediates, AMP and inorganic diphosphate.
tri- and pentapeptide and potential seryl O-activated peptide thioesters all failed to undergo hydrolysis at rates greater than controls. On the other hand, the corresponding tri- and pentapeptide thioesters now bearing a preformed β-lactam ring from cyclization of the seryl residue were not only rapidly hydrolysed but also completely epimerized to the C-terminal d-stereochemistry (Fig. 2c)\(^\text{14}\). NRPS epimerase activity by a TE domain was unprecedented, but this specific instance is due to the anomalously high acidity of a pHPG α-hydrogen relative to other α-amino acids\(^\text{15}\). Competition experiments established that the L,L,L,L-pentapeptide β-lactam thioester is the preferred NocTE substrate\(^\text{16}\), a finding fully in accord with the requirement that all five modules of NocA/B are necessary for nocardicin biosynthesis.

Although NocTE catalyses C-terminal epimerization and hydrolytic product release, it was not observed to mediate β-lactam synthesis\(^\text{14}\). Azetidinone formation, therefore, must logically occur upstream on the NRPS after introduction of the last pHPG unit in module 5 from which the β-lactam ring nitrogen arises. In principle, formation of the embedded β-lactam ring could take place either in cis in this module or occur in trans. The latter alternative invokes the action of auxiliary enzyme(s), which are increasingly precedent in NRPS biochemistry\(^\text{16}\). Among the mechanisms that can be visualized are in trans activation of the seryl hydroxyl group by, for example, phosphorylation or acylation, and intramolecular nucleophilic substitution (S,η) by the adjacent amide to form the critical C4–N bond, a process well supported by chemical precedent\(^\text{17}\) and consistent with the observation of stereochemical inversion at the seryl β-carbon\(^\text{16}\).

Bioinformatic analysis and biochemical experiments, however, did not point to candidate auxiliary enzyme(s) encoded by the nocardicin biosynthetic gene cluster\(^\text{18}\). As a consequence, experiments were first undertaken to probe the in cis strategy with unexpected results.

The termination module of NocB, module 5, is composed of four domains: C\(_5\), A\(_5\), PCP\(_5\) and TE. This 144 kilodalton protein is heterologously expressed in Escherichia coli with a His\(_6\) tag and purified by affinity chromatography. Complete conversion to its corresponding holo form was ensured by Sfp-mediated 4'-phosphopantetheinylation transfer from coenzyme A (CoASH)\(^\text{22,23}\). The final chemical transformations catalysed by the termination module were successfully reconstituted in vitro through incubation of the predicted tetrapeptide-modified PCP domain from module 4 (PCP\(_4\)) with holo-module 5. Bearing in mind that all five modules of NocA/B are required for production of nocardicin A in Nocardia uniformis, and that the β-lactam-containing pentapeptide is preferentially processed by NocTE over the corresponding tripeptide\(^\text{18}\), the tripeptide formed from L-pHPG–L-Arg and ATP showed no product formation. On the other hand, the corresponding dipeptide L-pHPG–L-Ser and ATP were combined, smooth conversion to the pentapeptide L-pHPG–L-Arg–D-pHPG–L-Ser–CoA (Fig. 3b) was observed (Supplementary Information) and linked to apo-PCP\(_4\) in an Sfp-mediated transfer to create L-pHPG–L-Arg–D-pHPG–L-Ser–CoA (Fig. 3b) and Extended Data Fig. 1). It was anticipated that module 5 would activate L-pHPG in the presence of ATP and present this amino acid on PCP\(_5\) for reaction with the tetrapeptide delivered to module 5 by PCP\(_4\). Indeed, when holo-module 5, 10 equivalents of tetrapeptidyl-S-PCP\(_5\) (2), L.pHPG and ATP were combined, smooth conversion to the pentapeptide β-lactam (pro-nocardicin G) was observed (Fig. 3c and Extended Data Fig. 2). Monitoring product formation by high-performance liquid chromatography (HPLC) in a time-course experiment and simultaneous consumption of tetrapeptidyl-S-PCP\(_5\) (2) by electrospray ionization mass spectrometry (ESI–MS) revealed a 1:1 correlation in accord with full catalytic turnover (Fig. 3d). Control experiments lacking L-pHPG or L-pHPG and ATP showed no product formation.

In a negative control experiment, the in vitro reconstitution experiment was repeated with a point mutant of C\(_5\) where the second histidine residue of the conserved active site HHxxxDG sequence, known to be essential for amide bond formation\(^\text{14}\), was replaced by alanine (H792A). No new products were detected (Fig. 3c). To further define acceptable substrates for C\(_5\), the L-pHPG–L-Arg–D-pHPG–L-Ser–S-pantetheine (3, Fig. 3a) substrate mimic was prepared (Supplementary Information) but did not yield pro-nocardicin G when incubated with holo-module 5, ATP and L-pHPG (Extended Data Fig. 3a,b), a result that emphasizes the critical importance PCP\(_5\)C\(_5\) domain–domain interaction plays to β-lactam formation. Next the dipeptide D-pHPG–L-Arg–CoA (4, Fig. 3a)
a mechanism in which His 790 catalyses (water) from the seryl residue of the PCP4-bound tetrapeptidyl-thioester characteristics of C5. The resulting L-pHPG and ATP, nocardicin G was not detected (Extended Data as before to afford D-pHPG–L-Ser– was prepared (Supplementary Information) and loaded onto products obtained after incubation of tetrapeptidyl-S-PCP4 and indicated construct, ATP and l-pHPG. This experiment was reproduced more than five times, and at least in duplicate for all other incubations. Pro-nocardicin G was observed in the wild-type reaction (+M5(WT)) but not in the mutant (+M5*H792A), verified by comparison with synthetic standard (top trace). Right: liquid chromatography–mass spectrometry (LC–MS) traces of products obtained after incubation of tetrapeptidyl-L-PCP4 and holo-module was prepared (Supplementary Information) and loaded onto apo-PCP4 as before to afford l-pHPG–l-Ser–S-PCP4 (5, Extended Data Fig. 4). When this construct was generated in the presence of holo-module 5, l-pHPG and ATP, nocardicin G was not detected (Extended Data Fig. 3c, d). These data suggest that the l-pHPG–l-Arg ‘leader’ present in the tetrapeptidyl-S-PCP4 plays a vital role in the binding and/or recognition of the upstream tetrapeptidyl intermediate in C5, enabling peptide extension and β-lactam formation to occur.

Examination of the primary sequence of C5 showed no unusual insertions or deletions except that, in addition to the conserved HHXXXDG catalytic motif emblematic of condensation domains, a third His residue (H790) lies directly upstream of the His dyad (Extended Data Fig. 5). Sequence analysis also revealed features of a DCL domain despite receiving an l-seryl tetrapeptide from PCP4 (Extended Data Table 1). We propose an L-seryl tetrapeptide from PCP4 (Extended Data Table 1). We propose an L-seryl tetrapeptide from PCP4 (Extended Data Table 1). We propose that the PCP4-bound tetrapeptidyl-thioester and PCP5-tethered l-pHPG achieves β-addition with overall inversion of configuration at the seryl(dehydroalanylated) β-carbon dictated by earlier stereochemical experiments (Fig. 4a)33. The transient loss of the l-seryl stereocentre during the β-addition/elimination may account for the DCL characteristics of C5. The resulting β-aminothioester 6 is then proposed to undergo unconventional amide bond cyclization (allowed 4-exo-trig), thermodynamically driven by amide bond formation from the active PCP5 thioester. The PCP5-bound pentapeptide β-lactam (pro-epi-nocardicin G) is poised for delivery to NocTE for C-terminal epimerization and hydrolytic product release.

To support this mechanistic hypothesis, we prepared a mutant of module 5 in which the tentative catalytic His was replaced by alanine (M5*H790A). Repeating the experiments with PCP5-bound tetrapeptide 2 and mutant M5*H790A, l-pHPG and ATP yielded no product (Extended Data Fig. 6). Similarly site-specific mutation of the His residue typically involved in peptide bond formation (M5*H792A) also gave no reaction, as anticipated. In a further test of the proposed mechanism, the reactive dehydroalanyl tetrapeptide intermediate (6, Fig. 4b) was synthesized (Supplementary Information) and used in an Sfp-catalysed reaction to afford the corresponding l-pHPG–l-Arg–d-pHPG–dehydroalanyl-S-PCP4 substrate (7, Fig. 4c and Extended Data Fig. 7). In the course of preparing this sensitive material, it was discovered that the addition of sulphur, phosphorus and nitrogen nucleophiles occurred preferentially 1,4 rather than 1,2 in keeping with the hypothetical reactivity posed in Fig. 4a. When PCP4-bound dehydroalanyl tetrapeptide 7 was incubated with wild-type holo-module 5, l-pHPG and ATP, β-lactam formation was once again observed (Fig. 4d and Extended Data Figs 8 and 9). Further insight into this process was afforded by the M5*H790A mutant, which did not support complete reaction of the dehydroalanyl substrate to the β-lactam product (Fig. 4d). The proposed catalytic residue H790 must not only act as a base to promote β-elimination but also serve as the acid to consummate amine (l-pHPG–S-PCP4) β-addition. Although interfering with the proper cycling of the protonation state of the enzyme can be partly compensated in the wild-type protein, it cannot in the M5*H790A mutant.

NRPS C domains are pseudodimeric proteins whose N- and C-terminal subdomains are joined to form an extended V-shaped substrate channel that accommodates the donor and acceptor aminoacyl reactants, each delivered by extended pantetheinyl ‘arms’ from proximal PCP domains26. The centrally located HHXXXDG motif promotes peptide bond formation and transfer of the growing peptide chain to the downstream PCP domain. The unprecedented β-lactam formation catalysed by C5 is distinct from the iron-mediated oxidative cyclization to penicillin (Fig. 1a) and the ATP-driven β-amino-acid closures that lead ultimately to clavulanic acid and all of the carbapenems (Fig. 1b)27, and it does not correlate to the heterocyclization of serine residues to oxazolidine rings28. There is no stereoelectronic imperative that β-elimination/β-addition reactions must occur with overall retention of stereochemistry27. It can be readily appreciated that departure of the seryl OH and conjugate addition of the pHPG amine can take place on opposite faces of the
Received 4 May; accepted 19 November 2014.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.
17. Townsend, C. A. & Wilson, B. A. The role of nocardicin G in nocardicin A biosynthesis. *J. Am. Chem. Soc.* **110**, 3320–3321 (1988).
18. Gaudelli, N. M. & Townsend, C. A. Epimerization and substrate gating by a TE domain in β-lactam antibiotic biosynthesis. *Nature Chem. Biol.* **10**, 251–258 (2014).
19. Walsh, C. T. et al. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Struct. Biol.* **5**, 525–534 (2001).
20. Salituro, G. M. & Townsend, C. A. Total syntheses of (−)-nocardicins A–G: a biogenetic approach. *J. Am. Chem. Soc.* **112**, 760–770 (1990).
21. Gunsior, M. et al. The biosynthetic gene cluster for a monocyclic β-lactam antibiotic, nocardicin A. *Chem. Biol.* **11**, 927–938 (2004).
22. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science* **284**, 486–489 (1998).
23. Quadri, L. E. N. et al. Characterization of Sfp, a Bacillus subtilis phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry* **37**, 1585–1595 (1998).
24. Roche, E. D. & Walsh, C. T. Dissection of the EntF condensation domain boundary and active site residues in nonribosomal peptide synthesis. *Biochemistry* **42**, 1334–1344 (2003).
25. Townsend, C. A., Brown, A. M. & Nguyen, L. T. Nocardicin A: stereochemical and biomimetic studies of a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792 (2007).
26. 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).
27. Tahlan, K. & Jensen, S. E. Origins of the β-lactam rings in natural products. *J. Antibiot. (Tokyo)* **66**, 401–410 (2013).
28. Sattely, E. S. & Walsh, C. T. A latent oxazoline electrophile for N-O-C bond formation in pseudomomine biosynthesis. *J. Am. Chem. Soc.* **130**, 12282–12284 (2008).
29. Silverman, R. B. The Organic Chemistry of Enzyme-Catalyzed Reactions revised edn, 399–452 (Academic, 2002).
30. Arnison, P. G. et al. Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat. Prod. Rep.* **30**, 108–160 (2013).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by National Institutes of Health grant AI014937. We are indebted to D. W. Udwary (bioinformatics), K. A. Moshos, J. W. Labonte (chemistry) and R.-f. Li (molecular biology) for advice and discussion. We thank C. T. Walsh for the pET29-Sfp expression plasmid, I. P. Mortimer for high-resolution mass spectrometry (HRMS) data and C. Moore for help with NMR acquisitions.

**Author Contributions** C.A.T. and N.M.G. developed the hypothesis and designed the study. N.M.G. and D.H.L. performed syntheses and biochemical experiments reported. All authors analysed and discussed the results. N.M.G., D.H.L. and C.A.T. prepared the manuscript.

**Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.A.T. (ctownsend@jhu.edu).
METHODS

No statistical methods were used to predetermine sample size.

Synthesis of all compounds used in this study can be found in the associated Supplementary Information.

General methods. Analytical HPLC analyses of enzymatic reactions were performed on an Agilent model 1200 HPLC equipped with a multi-wavelength ultraviolet-visible detector in conjunction with a reverse-phase Phenomenex Luna 5 μm phenyl/hexyl analytical column (250 mm × 4.6 mm internal diameter). Water + ACN + 0.1% TFA: 0–5 min isocratic 93% water + 7% ACN + 0.1% TFA, 5–22 min gradient 7–50% ACN + 0.1% TFA, 22–25 min gradient 50–70% ACN + 0.1% TFA, 25–35 min isocratic 93% water + 7% ACN + 0.1% TFA. Flow rate = 1.0 ml min⁻¹.

Ultra-performance liquid chromatography (UPLC)-HRMS samples were analysed on a Waters Acquity H-Class UPLC system equipped with a multi-wavelength ultraviolet-visible diode array detector in conjunction with a Waters Acquity BEH UPLC column packed with an ethylene bridged hybrid C-18 stationary phase (2.1 mm × 50 mm, 1.7 μm) in tandem with HRMS analysis by a Waters Xevo-G2 Q-ToF ESI mass spectrometer. Mobile phase: 100% water + 0.1% formic acid 0–1 min, 1–7.5 min 80% ACN + 0.1% formic acid, 7.5–8.4 min isocratic 80% ACN + 0.1% formic acid, 8.4–10 min 100% water + 0.1% formic acid. Flow rate = 0.3 ml min⁻¹.

Cloning, expression and purification of His₆-module 5. The module 5 gene containing C-A-PCP₅-TE of the termination module in NocB was PCR amplified from the pMG0531 cosmid containing nocA and nocB genes using the M5-forward and M5-reverse primers (Supplementary Table 1) and Herculase-HF DNA polymerase. The resulting PCR product was incorporated into a pCRBlunt-TOPO subcloning vector and sequence verified (Johns Hopkins University Core Sequencing Facility). The pCRBlunt-M5 construct was digested with Ndel and NotI and ligated with T4 DNA ligase into a similarly digested pET28b vector to create the corresponding N-terminal 6×-His fusion construct. Expression and purification of the PCP₅ monodomain was achieved similarly through procedures described for the module 5 constructs.

In vitro reconstitution of module 5 activity. Loading of peptidyl-S-CoA onto apo-PCP₅. The apo-PCP₅ constructs were converted to their holo forms by an Sfp-mediated transfer of the desired peptidyl-S-CoA substrate with the apo-PCP₅ construct. Apo-PCP₅ (200 μM) was incubated with 250 μM of desired peptidyl-S-CoA substrate in assay buffer supplemented with 10 mM MgCl₂, 4'-Phosphopantetheine transfer reactions were initiated by the addition of 2 μM of Sfp, and the enzymatic mixture was incubated for 45 min at room temperature. Excess peptidyl-S-CoA reagent was removed through serial dilutions of the reaction mixture. This was achieved by adding three volumes of assay buffer to the reaction mixture and concentrating the mixture back down to the initial volume using a 3k MWCO Amicon Ultra centrifugal filter (Millipore). This dilution procedure was repeated three times.

Generation of holo-module 5 construct. To a separate 1.5 ml tube, 20 μM of apo-module 5 construct (either wild type or mutant) was incubated with 40 μM coenzyme A in assay buffer supplemented with 10 mM MgCl₂, 4'-Phosphopantetheine transfer was initiated by the addition of 2 μM of Sfp and the reaction was left to stand for 45 min at room temperature. Excess CoA reagent was removed through serial dilutions of the enzymatic mixture as before.

Module-5-catalysed β-lactam formation. Holo-module 5 constructs were supplemented with 5 mM ATP and 2 mM 5'-pHPG and left to stand for 5 min in assay buffer. Condensation reactions were initiated by adding equal volumes of peptidyl-S-PCP₅ construct with holo-module 5 and left to stand for 2 h. The reaction contained 100 μM peptidyl-S-PCP₅, 10 μM holo-module 5, 2.5 mM ATP and 1 mM 5'-pHPG in assay buffer. Proteins were removed by centrifugation through a 3k MWCO Amicon Ultra centrifugal filter. The filtrate was directly analysed by HPLC and products of interest were collected over multiple injections and concentrated by lyophilization. The concentrated samples were re-suspended in 70 μl of 95:5 water:ACN + 0.1% formic acid and directly analysed by LC–MS.

Incubation of tetrapeptidyl-S-pantetheine 3 with holo-module 5. Reactions in which pantetheinyl substrate 3 was substituted for the holo-PCP₅ construct contained 1 mM 3, 10 μM holo-module 5 construct, 1 mM 5'-pHPG and 2.5 mM ATP and were left to stand for 2 h at room temperature in assay buffer. Reactions were quenched and analysed as described above.

31. Ehrmann, D. E., Trauger, J. W., Stachelhaus, T. & Walsh, C. T. Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. Chem. Biol. 7, 765–772 (2000).

32. Röttig, M. et al. NRPSpredictor2—a web server for predicting NRPS adenylation domain specificity. Nucleic Acids Res. 39, W362–W367 (2011).

33. Rausch, C., Weber, T., Kohlbacher, O., Wohlenberg, W. & Huson, D. H. Specificity prediction of adenylation domains in nonribosomal peptide synthetases (NRPS) using transductive support vector machines (TSVMs). Nucleic Acids Res. 33, 5799–5808 (2005).

34. Clugston, S. L., Sieber, S. A., Marahiel, M. A. & Walsh, C. T. Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a (D)C(L) catalyst. Biochemistry 42, 12095–12104 (2003).

35. Rausch, C., Hoof, L., Weber, T., Wohlenberg, W. & Huson, D. H. Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. BMC Evol. Biol. 7, 78 (2007).
Extended Data Figure 1 | Mass spectrometric verification of apo to holo conversion of PCP₄ with Sfp and L-pHPG-L-Arg-D-pHPG-L-Ser-S-coenzyme A (1), forming L-pHPG-L-Arg-D-pHPG-L-Ser-S-PCP₄ (2). Top: mass spectrum (ESI⁺) of apo-PCP₄. Bottom: mass spectrum (ESI⁺) of L-pHPG-L-Arg-D-pHPG-L-Ser-S-PCP₄ (2) derived from treatment of apo-PCP₄ with Sfp and corresponding synthetic tetrapeptidyl-CoA substrate 1.
Extended Data Figure 2 | Mass spectrum of major product generated from reaction catalysed by holo-module 5 supplemented with ATP, l-pHPG and l-pHPG-l-Arg-l-pHPG-l-Ser-S-PCP$_4$ (2). The major product observed from the reaction containing l-pHPG-l-Arg-l-pHPG-l-Ser-S-PCP$_4$ (2), holo-module 5, ATP and l-pHPG was isolated over multiple injections by HPLC (peak isolated indicated in inset HPLC trace) and then characterized by HRMS (ESI$^+$). The exact mass ion corresponding to the [M + H]$^+$ ion of pro-nocardicin G was observed. Exact mass calculated for pro-nocardicin G: C$_{33}$H$_{39}$N$_8$O$_9$: 691.2835; found: 691.2839 [M + H]$^+$. 

©2015 Macmillan Publishers Limited. All rights reserved
Extended Data Figure 3 | Negative experiments resulting from incubation of holo-module 5 with alternative substrates. 

a, Schematic of incubation of the tetrapeptidyl substrate L-pHPG-L-Arg-D-pHPG-L-Ser-S-pantetheine (3) with holo-module 5 supplemented with ATP and L-pHPG. Pro-nocardicin G was not produced. 

b, HPLC traces of products obtained after incubation of L-pHPG-L-Arg-D-pHPG-L-Ser-S-pantetheine (3) and holo-module 5 (+ M5(WT)) supplemented with ATP and L-pHPG. Pro-nocardicin G was not observed as verified by comparison with authentic standard. Additionally, it was noted that substrate 3 was not consumed over the duration of the experiment. 

(i) HPLC trace of the unbound products resulting from incubation of wild-type holo-module 5 reaction with L-pHPG-L-Arg-D-pHPG-L-Ser-S-pantetheine (3), L-pHPG and ATP. 

(ii) HPLC trace of the unbound products resulting from incubation of holo-M5*H792A reaction with D-pHPG-L-Arg-D-pHPG-L-Ser-S-pantetheine (3), L-pHPG and ATP. 

(iii) HPLC trace of incubation of L-pHPG-L-Arg-D-pHPG-L-Ser-S-pantetheine (3), L-pHPG and ATP without enzyme. The peak corresponding to L-pHPG-L-Arg-D-pHPG-L-Ser-S-pantetheine (3) substrate is indicated. 

(iv) HPLC trace of authentic standard of pro-nocardicin G. 

c, Schematic of incubation of the dipeptidyl substrate D-pHPG-L-Ser-PCP4 (5) with holo-module 5, L-pHPG and ATP. Nocardicin G, the corresponding expected product, was not observed. 

d, LC–MS traces of products obtained after incubation of D-pHPG-L-Ser-PCP4 (5) and holo-module 5 (+ M5(WT)) supplemented with ATP and L-pHPG. The corresponding β-lactam product was not observed, as verified by comparison with authentic standard. 

(i) Total ion chromatogram of the unbound products resulting from wild-type holo-module 5 reaction with D-pHPG-L-Ser-PCP4 (5), L-pHPG and ATP. 

(ii) Extracted ion chromatogram of the wild-type reaction in trace (i), the 386.1 m/z ion, corresponding to [M + H] of nocardicin G was not observed. 

(iii) Total ion chromatogram of unbound products from holo-M5*H792A control reaction with D-pHPG-L-Ser-PCP4 (5), L-pHPG and ATP. 

(iv) Extracted ion chromatogram of the wild-type reaction in trace (iii). 

(v) Extracted ion chromatogram of authentic standard of nocardicin G.
Extended Data Figure 4 | Mass spectrometric comparison of apo with holo conversion of PCP₄ with Sfp and D-pHPG-L-Ser-S-coenzyme A (4) forming D-pHPG-L-Ser-S-PCP₄ (5). Top: mass spectrum (ESI⁺) of apo-PCP₄. Bottom: mass spectrum (ESI⁺) of D-pHPG-L-Ser-S-PCP₄ (5) derived from treatment of apo-PCP₄ with Sfp and corresponding synthetic dipeptidyl CoA substrate 4. Owing to diketopiperazine formation and slow hydrolysis to the unloaded holo-PCP₄, the half-life of 5 is approximately 20 min.
Extended Data Figure 5 | Multiple sequence alignment of homologous NRPS C domains and the NocB C5 domain. The alignment was performed using Clustal2 with gap opening and extending penalties of 10 and 0.1 respectively. The alignment included the N-terminal C fragment of EntF, the bacillibactin synthetase DhbF, the tyrocidine synthetase TycB, the surfactin synthetase SrfAC and the naturally free-standing condensation enzyme VibH. Highlighted in a red box is H790 uniquely present in the NocB C5 domain along with the corresponding residues present in the depicted C domains. The NocB C5 domain is N-terminal to the consensus catalytic motif, HHxxxDG, present in all canonical condensation domains.
Extended Data Figure 6 | HPLC comparative analysis of the reactions catalysed by holo-module 5 and holo-M5*H790A with l-pHPG-l-Arg-d-pHPG-l-Ser-S-PCP₄ (2).

a, Schematic of incubation of the tetrapeptidyl substrate l-pHPG-l-Arg-d-pHPG-l-Ser-S-PCP₄ (2) with holo-M5*H790A supplemented with ATP and l-pHPG. Pro-nocardicin G was not produced.

b, HPLC comparison of unbound products from reaction mixtures containing l-pHPG-l-Arg-d-pHPG-l-Ser-S-PCP₄ (2) and either holo-module 5 or M5*H790A variant supplemented with ATP, l-pHPG, indicating the formation of β-lactam product only from the wild-type reaction. (i) HPLC trace of the unbound products resulting from incubation of wild-type holo-module 5 reaction with l-pHPG-l-Arg-d-pHPG-l-Ser-S-PCP₄ (2), l-pHPG and ATP. (ii) HPLC trace of the unbound products resulting from incubation of holo-M5*H790A reaction with l-pHPG-l-Arg-d-pHPG-l-Ser-S-PCP₄ (2), l-pHPG and ATP. Pro-nocardicin G was not observed. (iii) HPLC trace of the unbound products resulting from incubation of holo-M5*H792A reaction with l-pHPG-l-Arg-d-pHPG-l-Ser-S-PCP₄ (2), l-pHPG and ATP. (iv) HPLC trace of authentic standard of pro-nocardicin G.
Extended Data Figure 7 | Mass spectral comparison of apo with holo conversion of PCP₄ with Sfp and L-pHPG-L-Arg-D-pHPG-dehydroalanyl-S-3'-dephospho coenzyme A (6) forming L-pHPG-L-Arg-D-pHPG-dehydroalanyl-S-PCP₄ (7). Top: mass spectrum (ESI⁺) of apo-PCP₄. Bottom: mass spectrum (ESI⁺) of L-pHPG-L-Arg-D-pHPG-dehydroalanyl-S-PCP₄ (7) derived from treatment of apo-PCP₄ with Sfp and corresponding synthetic eliminated tetrapeptidyl-CoA substrate 6.
Extended Data Figure 8 | Mass spectrum of major product generated from reaction catalysed by holo-module 5 and L-pHPG-L-Arg-D-pHPG-dehydroalanyl-S-PCP₄ (7) supplemented with ATP and L-pHPG. The major product from the reaction containing L-pHPG-L-Arg-D-pHPG-dehydroalanyl-S-PCP₄ (7) was isolated over multiple injections by HPLC (peak indicated in inset HPLC trace) and then determined by HRMS (ESI+). The exact mass ion corresponding to the [M + H]⁺ ion of pro-nocardicin G was observed. Exact mass calculated for C₃₃H₃₉N₈O₉: 691.2835; found: 691.2822 [M + H]⁺.
Extended Data Figure 9 | Tandem mass spectrometric comparison of holo-module 5 synthesized pro-nocardicin G from L-pHPG-L-Arg-D-pHPG-L-Ser-S-PCP₄ (2) and L-pHPG-L-Arg-D-pHPG-dehydroalanyl-S-PCP₄ (7). Fragmentation pattern of holo-module 5 biosynthetic pro-nocardicin G from serine-containing tetrapeptide substrate L-pHPG-L-Arg-D-pHPG-L-Ser-S-PCP₄ (2, bottom) is directly comparable to the tandem mass spectrometric fragmentation pattern of holo-module 5 synthesized pro-nocardicin G from the dehydroalanine-containing substrate L-pHPG-L-Arg-D-pHPG-dehydroalanyl-S-PCP₄ (7, middle). For comparison, the mass spectrum of synthetic pro-nocardicin G is shown at top.
Extended Data Table 1 | Condensation domains with a HHHxxxDG motif and an upstream predicted serine- or threonine-activating A domain

| sequence ID | species | a.a. | predicted natural product |
|-------------|---------|------|---------------------------|
| gil470738102g| Nocardiia uniformis subsp.tsuwaymanensis | ser | nocardicin A |
| gil255918463g| Actinosynnema mirum DSM 43827 | ser | nocardicin A |
| gil45179617g| Streptomyces hygroscopicus subsp. jinggangensis TL01 | ser | nocardicin A |
| gil29979145g| Amycolatopsis mediterranei U32 | ser | N/A |
| gil40730462g| Nocardia brasiliensis ATCC 700358 | thr | N/A |
| gil85108654g| Nocardia brasiliensis strain NBRC 14402 | thr | N/A |
| gil30758261g| Burkholderia sp. CCGE1003 | thr | N/A |
| gil40723395g| Burkholderia phenoliruptix BR3459a | thr | N/A |
| gil32338137g| Burkholderia sp. CCGE1001 | thr | N/A |
| gil30060345g| Candidatus Nitrospira defluvii | ser* | N/A |
| gil33629553g| Paenibacillus mucilaginosus KNP414 | ser* | N/A |
| gil51120478g| Paenibacillus mucilaginosus K02 | ser* | N/A |

Condensation domains with a HHHxxxDG motif and an upstream putative serine- or threonine-activating A domain were identified through a tblastn search on the National Center for Biotechnology Information (NCBI) nucleotide database. An initial search was done using the A4PCP4C5 tri-domain from the nocardicin gene cluster as a query resulting in 25,532 hits from 7,164 sequences under an E-value threshold of $10^{-20}$. These hits were then filtered to exclude sequences with query coverages of less than 75%, yielding 3,834. All sequences were translated and parsed for HHHxxxDG motifs. The subsequent 37 sequences were submitted to NRPSPredictor2 (refs 32, 33) to identify upstream predicted serine or threonine A domains. After removing duplicate sequences, a final set of 12 remained. From the final set of sequences, only those with high confidence predictions of upstream serine- or threonine-specific A domains were aligned with the nocardicin module 5 condensation domain. Analysis revealed that all sequences in the alignment displayed characteristic motifs predominantly observed in the DCL subtype34,35 of condensation domains, despite the fact that nocardicin C5 exhibits LCL binding.