

Cite this: *Chem. Sci.*, 2019, 10, 9466

All publication charges for this article have been paid for by the Royal Society of Chemistry

A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics†

Milda Kaniusaite,^{id}^{ab} Julien Tailhades,^{id}^{ab} Edward A. Marschall,^{ab}
Robert J. A. Goode,^{id}^{ac} Ralf B. Schittenhelm^{ac} and Max J. Cryle^{id}^{*ab}

Non-ribosomal peptide biosynthesis produces highly diverse natural products through a complex cascade of enzymatic reactions that together function with high selectivity to produce bioactive peptides. The modification of non-ribosomal peptide synthetase (NRPS)-bound amino acids can introduce significant structural diversity into these peptides and has exciting potential for biosynthetic redesign. However, the control mechanisms ensuring selective modification of specific residues during NRPS biosynthesis have previously been unclear. Here, we have characterised the incorporation of the non-proteinogenic amino acid 3-chloro- β -hydroxytyrosine during glycopeptide antibiotic (GPA) biosynthesis. Our results demonstrate that the modification of this residue by *trans*-acting enzymes is controlled by the selectivity of the upstream condensation domain responsible for peptide synthesis. A proofreading thioesterase works together with this process to ensure that effective peptide biosynthesis proceeds even when the selectivity of key amino acid activation domains within the NRPS is low. Furthermore, the exchange of condensation domains with altered amino acid specificities allows the modification of such residues within NRPS biosynthesis to be controlled, which will doubtless prove important for reengineering of these assembly lines. Taken together, our results indicate the importance of the complex interplay of NRPS domains and *trans*-acting enzymes to ensure effective GPA biosynthesis, and in doing so reveals a process that is mechanistically comparable to the hydrolytic proofreading function of tRNA synthetases in ribosomal protein synthesis.

Received 25th July 2019
Accepted 29th August 2019

DOI: 10.1039/c9sc03678d

rsc.li/chemical-science

Introduction

Non-ribosomal peptide biosynthesis plays a major role in the formation of peptide-based natural products, of which many have important medicinal properties.¹ The significant diversity of peptide structures produced through non-ribosomal peptide biosynthesis is based on the modular architecture of non-ribosomal peptide synthetase (NRPS) assembly lines. These modules, which are comprised of different catalytic domains, can accept a very wide range of building blocks through the activity of their adenylation (A) domains that perform substrate selection and activation.² At this point in biosynthesis, the amino acid intermediates are transferred to neighbouring peptidyl carrier protein (PCP) domains, where they are tethered

as thioesters *via* the essential PCP phosphopantetheinyl moiety.³ Formation of peptide bonds is then mediated by the activity of condensation (C) domains, in which the upstream (donor) PCP-bound peptide is transferred onto the aminoacyl-PCP acceptor with concomitant peptide bond formation.⁴ A-Domains are typically highly selective for their specific amino acid substrates, and hence largely control the amino acid composition of the final non-ribosomal peptides,⁵ whilst C-domains control the stereochemistry of the upstream (donor) peptide and can also display selectivity for the downstream (acceptor) amino acid.^{4,6–8} Once the complete peptide has been synthesised, it is typically released from the NRPS through the actions of a terminal thioesterase domain in a process that can generate yet further structural changes to the peptide (through cyclisation/dimerisation *etc.*).⁹ Additional diversity in peptide structure can also be installed during biosynthesis by the actions of further catalytic domains found either within the assembly line itself, such as epimerisation performed through (E) domains¹⁰ and methylation performed by methyltransferase (MT) domains,¹¹ or through the activity of enzymes acting in *trans* to the main peptide assembly line such as halogenases or oxidases.^{12,13} Whilst epimerisation typically occurs on PCP-bound peptides and selectivity is ensured by the actions of the

^aThe Monash Biomedicine Discovery Institute, Department of Biochemistry and Molecular Biology, Monash University, Clayton, Victoria 3800, Australia. E-mail: max.cryle@monash.edu

^bEMBL Australia, Monash University, Clayton, Victoria 3800, Australia

^cMonash Proteomics and Metabolomics Facility, Monash University, Clayton, Victoria 3800, Australia

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c9sc03678d



neighbouring C-domain,^{6,7} *trans*-modifications largely target aminoacyl-PCPs, with the mechanism controlling the modification of specific aminoacyl-PCPs largely unclear.

The biosynthesis the glycopeptide antibiotics (GPAs) serves as an important example of a how nature can produce complex, medically-relevant antibiotics through the actions of an equally complex NRPS biosynthesis process (Fig. 1).¹⁴ GPAs are a family of highly crosslinked heptapeptides, and include lipid II binding antibiotics exemplified by vancomycin/balhimycin and teicoplanin, as well as other members such as kistamicin and complestatin that have antiviral activity.¹⁴ Whilst the biosynthesis of GPA peptides initially appears straightforward, closer inspection reveals significant complexity: in particular, this includes the modification of the structure of amino acids (chlorination and/or hydroxylation of tyrosine residues 2 and 6 within vancomycin and teicoplanin type GPAs) and the peptide itself (through extensive crosslinking of aromatic side chains).¹⁵ The crosslinking cascade in GPA biosynthesis has been shown

to depend on recruitment of external cytochrome P450 mono-oxygenases through the conserved X-domain within the final NRPS module, a process unique to GPAs.^{16,17} The process controlling the modification of amino acid residues within the peptide, however, is less clear. Different mechanisms also exist for incorporation of specific amino acid modifications during NRPS-mediated biosynthesis of vancomycin and teicoplanin-type GPAs. Whilst studies support amino acid modifications as largely occurring on aminoacyl-PCP substrates during GPA biosynthesis, these also indicate that selectivity cannot alone be dictated by selectivity of the *trans*-modifying enzymes for the correct PCP domains.^{18–21} Thus, we hypothesised that the selection agent in these modification reactions would be the neighbouring C-domain that would only select the PCP-bound amino acid for peptide elongation once this residue has been appropriately modified. This implies that selectivity in GPA biosynthesis relies on the complex interplay of specificities from

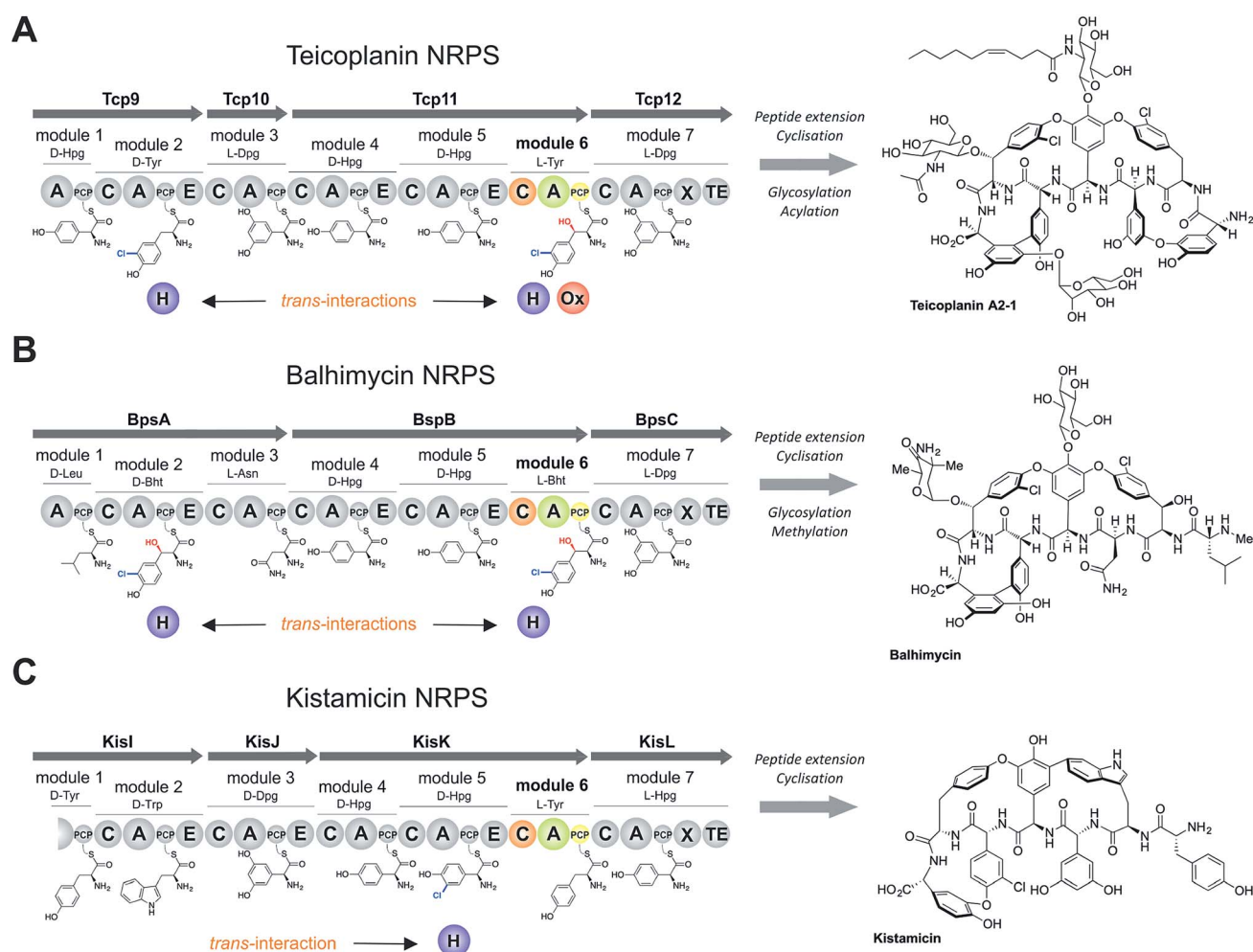


Fig. 1 Non-ribosomal peptide biosynthesis of the glycopeptide antibiotics teicoplanin (A), the vancomycin-type GPA balhimycin (B) and kistamicin (C), concentrating on the NRPS proteins and *trans*-interacting enzymes. Module 6 of the NRPS machineries are shown in colour, the rest of the NRPS is indicated in grey. A – adenylation domain, C – condensation domain, PCP – peptidyl carrier protein domain, E – epimerisation domain, X – cytochrome P450 recruitment domain, TE – type I thioesterase domain H – flavin-dependent halogenase, Ox – non-heme iron oxygenase, Hpg – 4-hydroxyphenylglycine, Dpg – 3,5-dihydroxyphenylglycine, Tyr – tyrosine (1), Bht – β -hydroxytyrosine (3), Trp – tryptophan, Leu – leucine, Asn – asparagine.



multiple enzymatic domains both within and external to the NRPS machinery, which has not yet been investigated.

In this work, we describe a complex interplay between low A-domain selectivity, high C-domain selectivity and type-II TE promiscuity, which together constitutes a robust proof-reading system that ensures incorporation of the correct amino acid during NRPS-biosynthesis in a mechanism comparable to that used in tRNA-dependent biosynthesis. We further show that switching C-domains with different aminoacyl-PCP selectivity results in the formation of different peptide products, indicating how amino acid modification in *trans* can be selectively controlled during non-ribosomal peptide biosynthesis through the exchange of selective C-domains. Our results highlight the importance of a holistic approach to NRPS reconstitution and analysis, and that the selectivity observed in NRPS biosynthesis can stem from the interplay of a number of different mechanisms that need to be understood in concert before effective biosynthetic reengineering can be performed.

Experimental

Construct cloning

Tcp11 modules with typical architecture. Synthetic, codon-optimised genes for *E. coli* expression encoding Tcp11 (UniProt code Q70AZ7) module 4 (C-A-PCP-E architecture), module 5 (C-A-PCP-E architecture) and module 6 (C-A-PCP architecture) from *Actinoplanes teichomyceticus* (DSM 43866) were obtained from Eurofins Genomics MWG. All three module sequences (ESI Table S2†) were designed for direct NcoI/XhoI cloning into a modified pET vector (pET-GB1-1d series), which encodes an N-terminal 6xHis-Tag followed by the IgG-binding B1 domain of *Streptococcus* (GB1) with a TEV cleavage site under the control of a T7 promoter and a Strep-Tag II at the C-terminus.^{22,23} Additionally, synthetic gene sequences at the module 4 C-terminus, module 5 N- and C-termini and module 6 N-terminus were modified with short sequence regions (6–11 amino acid residues) based on other Tcp enzyme linkers in order to improve protein expression and possibly facilitate module interactions. Synthetic genes and the pET-GB1-1d series vector were digested using NcoI and XhoI restriction enzymes (NEB) and then ligated using T4 ligase (NEB). The ligation mixture was used to transform NEB 10-beta competent *E. coli* cells according to standard procedures, before being plated onto LB-agar plates supplemented with kanamycin. The next day, plasmid DNA from individual colonies was isolated and sequenced to confirm that the appropriate module fragment was correctly inserted into the plasmid.

Tcp39. A synthetic, codon optimised gene for *E. coli* expression encoding Tcp39 (UniProt code Q6ZZG3) from *Actinoplanes teichomyceticus* (DSM 43866) was obtained from Eurofins Genomics MWG. The gene sequence (ESI Table S2†) was designed for direct NcoI/XhoI cloning into a modified pET vector (pET-MBP-1c series), which encodes maltose binding protein (MBP) and TEV cleavage site under the control of T7 promoter and a C-terminal 6x-His Tag.²³ The synthetic gene and pET-MBP-1c series vector were digested using NcoI and XhoI restriction enzymes (NEB) and then ligated together with T4

ligase (NEB) and transformed into NEB 5-alpha competent *E. coli* cells according to standard procedures, before being plated onto LB-agar plates supplemented with kanamycin. The next day, plasmid DNA from individual colonies was isolated and sequenced to confirm that the gene was correctly inserted into the plasmid.

Tcp11 modules with alternative architecture. Tcp11 constructs with an alternative module architecture (module 4 C-A-PCP-E-C, module 5 A-PCP-E-C and module 6 A-PCP) were cloned into a pET-GB1-1d vector²² using In-Fusion® HD Cloning kit (Clontech). PCR primers were designed that share 15 bases of homology with adjacent DNA fragments. Then these primers (ESI Table S3†) were used to PCR amplify both the insert/(-s) and plasmid DNA. The plasmid DNA, containing the gene of NRPS module of interest, was used as the template DNA for a PCR reaction. Fragments were amplified using Phusion® Hot Start Flex Master Mix (NEB) and the appropriate forward and reverse primers (ESI Table S3†). The PCR products were analysed on a 0.8% agarose gel in TAE buffer and the DNA subsequently gel-extracted and purified using the Wizard® SV gel and PCR clean-up kit (Promega). The extracted PCR products (insert/(-s) and vector) were combined in the In-Fusion® cloning reaction as per the manufacturer's instructions. In-Fusion® cloning reactions were incubated for 15 min at 50 °C, then placed on ice and 2.5 µL of the reaction mixture was used for transformation of NEB 10-beta competent *E. coli* cells.

Tcp11 di-module 4–5. A Tcp11 di-module construct (C-A-PCP-E-C-A-PCP-E), covering module 4 (C-A-PCP-E) and module 5 (C-A-PCP-E), was cloned into a pET-GB1-1d vector²² using an In-Fusion® HD Cloning kit (Clontech) as described above. The desired PCR fragments were amplified using primers listed in ESI Table S3.†

Full length Tcp11. Full length Tcp11 (C-A-PCP-E-C-A-PCP-E-C-A-PCP) was cloned into a pET-GB1-1d vector²² using an In-Fusion® HD Cloning kit (Clontech) as described above. The desired PCR fragments were amplified using primers listed in ESI Table S3.†

BpsB module 6 (A-PCP) and MbtH-like protein. Synthetic, codon-optimised genes for *E. coli* expression encoding BpsB (UniProt code Q939Z0) module 6 (covering the A-PCP region) and MbtH-like protein (UniProt code Q939Y8) from the balhimycin biosynthesis gene cluster in *Amycolatopsis balhimycina* (DSM 44591) were obtained from Eurofins Genomics MWG. BpsB module 6 A-PCP construct was cloned into a modified pET vector (pET-MBP-1c series), which encodes maltose binding protein (MBP) and TEV cleavage site under the control of T7 promoter and a C-terminal 6x-His Tag²³ and MbtH-like protein into a pCDF vector. DNA fragments were amplified by PCR from a synthetic gene using primers listed in ESI Table S4.† Cloning procedure was performed using an In-Fusion® HD Cloning kit (Clontech) as described above. The desired PCR fragments were amplified using primers listed in ESI Table S4.†

Bhp. A synthetic, codon optimised gene for *E. coli* expression encoding Bhp (UniProt code Q939Y3) from *Amycolatopsis balhimycina* (DSM 44591) was obtained from Geneart. The gene sequence (ESI Table S4†) was designed for direct NdeI/HindIII cloning into a pET-28a(+) vector, which encodes an N-terminal



6x-His Tag and thrombin cleavage site under the control of T7 promoter. The synthetic gene and pET-28a vector were digested using NdeI and HindIII restriction enzymes (NEB) and then ligated together as for Tcp39.

KisK module 6 (A-PCP), A₆ mutant and MbtH-like protein. Synthetic, codon-optimised genes for *E. coli* expression encoding KisK module 6 (A-PCP architecture, UniProt code A0A2P9IC10) and MbtH-like protein KisM (UniProt code A0A2P9IBI0) from kistamicin biosynthesis NRPS gene cluster *Actinomadura parvosata* subsp. *kistnae* (*Nonomuraea* sp. ATCC55076) were obtained from Eurofins Genomics MWG (ESI Table S5†). Both KisK constructs (the wild type A-PCP construct as well as double A-domain mutant) and MbtH protein encoding plasmids were generated using In-Fusion® HD Cloning kit (Clontech) as described in section above. Both KisK constructs were cloned into the pET-GB1-1d series plasmid and KisM into a pCDF vector. DNA fragments were amplified by PCR from a synthetic gene using primers listed in ESI Table S5.† To improve amplification by PCR, the NEB Phusion® Hot Start Flex Master Mix was supplemented with 540 mM betaine, 1.34 mM DTT, 11 µg mL⁻¹ BSA and 1.34% DMSO.²⁴

Tcp11 module 5 A-PCP-E-C hybrid. Tcp11 module 5 (A-PCP-E-C topology) hybrid construct, where the wild type teicoplanin NRPS C-domain was replaced with the analogous KisK M5/6 C-domain was generated using In-Fusion® HD Cloning kit (Clontech) as described in sections above. DNA fragments required for hybrid protein construction were amplified by PCR from previously constructed protein expression plasmid encoding Tcp11 module 5 (A-PCP-E-C topology) and a synthetic gene encoding KisK module 5 (A-PCP-E-C architecture). Primers used for Tcp11 module 5 hybrid cloning are provided in ESI Table S6.†

Protein expression and purification

Tcp11, modules and hybrid modules were co-expressed with the teicoplanin MbtH-like protein Tcp17; KisK module 6 and the A-domain mutant were co-expressed with the kistamicin MbtH-like protein KisM and BpsB module 6 was expressed with the related balhimycin MbtH-like protein. To co-express the proteins, *E. coli* BL21 (DE3) (NEB) competent cells possessing the plasmid encoding appropriate MbtH-like protein gene were co-transformed together with a plasmid encoding the NRPS module of interest. For selection, two antibiotics (kanamycin and streptomycin) were used; expression of all modules were performed in TB media, supplemented with 50 µg mL⁻¹ kanamycin and 50 µg mL⁻¹ streptomycin, with 10 L of media used for each protein expressed. Cells were incubated with shaking at 37 °C until the OD_{600nm} reached 0.4–0.6. The cultures were then cooled on ice for 15 min and protein expression was induced by the addition of IPTG (0.1 mM); cultures were subsequently grown at 18 °C for 16 h. Tcp39 and Bhp were expressed without an MbtH-like protein according the same procedure using *E. coli* BL21 (DE3) (NEB) and TB media supplemented with 50 µg mL⁻¹ kanamycin.

All NRPS proteins were purified according the same procedure, with the exception of Tcp10 that was purified as described

previously.²² Cells were harvested by centrifugation at 4000 rpm for 20 min at 18 °C. Subsequently, the cell pellet was resuspended in Ni-NTA buffer A (50 mM Tris-HCl, pH 8.0; 300 mM NaCl; 10 mM imidazole) supplemented with protease inhibitor cocktail tablets (SIGMAFAST™ Protease Inhibitor Cocktail Tablets, EDTA-Free; Sigma-Aldrich). The cells were lysed by sonication and the lysate cleared by centrifugation at 15 000 rpm for 40 min at 4 °C. The supernatant was incubated at 4 °C for 1 h with 4 mL of equilibrated (50 mM Tris-HCl, pH 8.0; 300 mM NaCl; 10 mM imidazole) Ni-NTA beads (Macherey-Nagel) with gentle shaking. After incubation, beads were washed twice with 10 column volumes of Ni-NTA buffer A (50 mM Tris-HCl, pH 8.0; 300 mM NaCl; 10 mM imidazole) before the protein was eluted with 5 column volumes of Ni-NTA buffer B (50 mM Tris-HCl, pH 8.0; 300 mM NaCl; 300 mM imidazole). Strep-tag® affinity chromatography connected to ÄKTA PURE system (GE Healthcare) was then used as a second protein purification step: prior to purification, a 5 mL StrepTrap™ HD column (GE Healthcare) was equilibrated with 4 column volumes of StrepTrap buffer A (100 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA). The Ni-NTA eluate was then loaded onto the column, the column washed with 4–6 column volumes of StrepTrap buffer A (100 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA) and bound protein then eluted with StrepTrap buffer B (100 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA; 2.5 mM desthiobiotin). The protein was then further purified by gel-filtration chromatography on a Superose™ 12 prep grade XK 26/70 (320 mL) column (GE Healthcare) or a SRT 10 SEC 300 (105 mL) column (Sepax Technologies) connected to ÄKTA PURE system (GE Healthcare). Initially, the column was equilibrated with 1 column volume of gel-filtration buffer (50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA). Subsequently, the protein after Strep-tag® purification was loaded onto the column and the flow rate of gel-filtration buffer (50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 1 mM EDTA) adjusted to 0.7 mL min⁻¹ for Superose™ 12 column or to 3 mL min⁻¹ for the SRT 10 column (see ESI Fig. S1–S3†). Elution fractions containing monomeric protein were analysed by SDS-PAGE and appropriate fractions combined and concentrated using centrifugal filter units (Amicon® Ultra-15 centrifugal filter units (depending on protein size 30 K and/or 100 K was chosen), Merck Millipore). The protein concentration was determined by measuring protein absorbance at 280 nm using a NanoDrop™ One microvolume UV-Vis spectrophotometer (Thermo Scientific). Concentrated protein was then aliquoted (30–50 µL) into chilled 1.5 mL Eppendorf tubes, flash frozen in liquid nitrogen and stored at –80 °C. For the purification of Tcp39 and Bhp, a simplified, two-step purification procedure (using Ni-NTA and gel filtration) was performed according the same portions of the procedure described above.

Enzyme assays

PCP loading. All NRPS multi-domain proteins were expressed and purified in their *apo* form. Therefore, PCP domains were converted from their *apo* to *holo* form by loading the PPant linker in a reaction catalysed by the



phosphopantetheinyl transferase Sfp (R4-4 mutant).²⁵ Two sets of conditions were used depending on the to be substrates loaded, with coenzyme A used to transform PCP domains into their *holo* form and peptidyl-CoA conjugates used to generate *peptidyl*-PCPs. Loading reactions utilised a 10 : 60 (CoA)/20 (peptidyl-CoA) : 1 molar ratio of the PCP domain, *peptidyl*-CoA/CoA and R4-4 Sfp, respectively. The reaction was performed in PCP-loading buffer (50 mM HEPES, pH 7.0; 50 mM NaCl; 10 mM MgCl₂) at 30 °C for 1 h. Depending on the number of reactions to be performed, 50–400 µL final volume was used for the loading reaction. Subsequently the excess of peptidyl-CoA or CoA was removed by repeated concentration and dilution (5×) using reconstitution assay buffer (50 mM HEPES, pH 7.0; 50 mM NaCl) and centrifugal concentrators (Amicon® Ultra-0.5 mL centrifugal filters units (depending on protein size 30 K and/or 100 K was chosen), Merck Millipore). *Holo*-PCP constructs were then immediately used for *in vitro* reconstitution assays.

***In vitro* reconstitution of non-ribosomal peptide biosynthesis.** *Holo*-NRPS proteins were mixed with 1 mM ATP, 10 mM MgCl₂ and 1 mM amino acid substrates for each A-domain in reconstitution assay buffer (50 mM HEPES, pH 7.0; 50 mM NaCl) in a final volume of 200 µL. The reactions were incubated overnight at 30 °C with 300 rpm shaking. At the end of the reaction, peptides attached to any PCP domain were chemically cleaved by the addition of 15 µL of methylamine to liberate the methylamide peptides. The peptide cleavage reaction was incubated for 15 min at room temperature and the mixture was neutralised by addition of 4 mL of reconstitution assay buffer. The peptides formed were then purified from the reaction mixture *via* solid phase extraction using Strata™-X-33 µm Polymeric Reversed Phase Tubes (30 mg mL⁻¹) (Phenomenex). Before sample loading, the cartridges were activated with 1 mL of methanol and subsequently equilibrated with 1 mL of water. Neutralised reconstitution reaction samples were loaded onto an equilibrated cartridge and allowed to pass through the column bed by gravity flow. When all the sample was loaded onto the column, then cartridge was washed with 1 mL of 0.1% formic acid in water. Peptides were then eluted with 500 µL 1% formic acid in methanol. Subsequently, the methanol was evaporated using a centrifugal concentrator (Concentrator plus, Eppendorf) at 45 °C. After concentration, the residue was dissolved in 6 µL of 50% MeCN in water (Optima® LC/MS Grade) and diluted up to 36 µL final volume with 0.1% formic acid in water (Optima® LC/MS Grade). 14 µL of the sample was injected on the XBridge® Peptide BEH C18 column (5 × 300 mm, Waters) and analysed by LCMS (LCMS-2020, ESI, Shimadzu) in positive mode, with a 5–45% acetonitrile in water gradient over 40 min and a 1 mL min⁻¹ flow rate. Peptide yields were determined by the integration of the desired peptide peak divided by the sum of the area of all peptide peaks. Hexapeptide methylamide products were further confirmed by HRMS (Tyr (1), **6T-1** – C₅₁H₅₂N₇O₁₃ [M + H]⁺ expected *m/z* 970.3618, found *m/z* 970.3645; Cl-Tyr (2), **6T-2** – C₅₁H₅₁ClN₇O₁₃ [M + H]⁺ expected *m/z* 1004.3233, found *m/z* 1004.3265; Bht (3), **6T-3** – C₅₁H₅₂N₇O₁₄ [M + H]⁺ expected *m/z* 986.3572, found *m/z* 986.3649; Cl-Bht (4), **6T-4** – C₅₁H₅₁ClN₇O₁₄ [M + H]⁺ expected *m/z* 1020.3177, found *m/z* 1020.3198) and MS² measurements (see ESI Fig. S14–S17†).

A-domain activity assay. The activation rate of A-domains using different amino acid substrates 1–11 was determined using a continuous NADH/pyrophosphate (PPI) coupled assay. PPI formation during amino acid activation is coupled to the consumption of NADH using four enzymatic steps, providing a continuous spectroscopic assay by measuring the absorbance at 340 nm.²² Before testing A-domain activity, all PCP domains were converted into their *holo* form as discussed above using a modified loading buffer (25 mM Tris-HCl, pH 7.4; 5 mM MgCl₂) with 10-fold less Sfp (R4-4 mutant) for 30 min, with excess CoA removed by repeated concentration and dilution as indicated above using A-domain washing buffer (100 mM Tris-HCl; 1 mM MgCl₂, 0.1 mM EDTA). All assays were performed in a 500 µL volume at 30 °C and monitored using a V-750 spectrophotometer (Jasco). Data was fitted using SpectraManager II software and velocity was calculated from the slope of the linear phase using Beer-Lambert law ($v = \text{slope} (\text{Abs/min})/(\epsilon_{340}(\text{NADH}) \cdot l \cdot 2)$). To measure the A-domain activity of each NRPS module, seven different reagents were added to the A-domain assay buffer (100 mM Tris-HCl; 1 mM MgCl₂, 0.1 mM EDTA; 0.2 mM NADH), comprising: 1 µM A-domain, 3 mM D-fructose-6-phosphate, 0.1 U mL⁻¹ fructose-6-phosphate kinase, pyrophosphate-dependent (*Propionibacterium freudenreichii* (*shermanii*)), 1 U mL⁻¹ aldolase, 5 U mL⁻¹ triosephosphate isomerase, 5 U mL⁻¹ glycerophosphate dehydrogenase, 0.5 mM ATP. Reactions mixtures were pre-incubated for 4 min before the substrate amino acid was added and the measurement started by addition of 0.25 mM of amino acid substrate to the desired A-domain, with the decline in absorbance at 340 nm measured over time. Measurements were repeated in triplicate for each amino acid/A-domain combination tested. Before every set of measurements, duplicate control reactions to detect PPI were performed to ensure that the assay was functioning correctly. In these controls, reactions were initiated by addition of 20 µM sodium pyrophosphate into the reaction buffer that was supplemented with assay enzymes and NADH.

Amino acid synthesis

N-Chlorosuccinimide (230 mg, 1.7 mmol), *N*-bromosuccinimide (306 mg, 1.7 mmol) or *N*-iodosuccinimide (38.9 mg, 1.7 mmol) in dioxane (2 mL) was added dropwise to a suspension of tyrosine (2 mL) in dioxane/water/AcOH (8 mL, v/v/v, 1/4) at room temperature. After 15 hours, the reaction was concentrated under reduced pressure to remove the dioxane and the desired products purified using preparative RP-HPLC (gradient of ACN in water of 5–35% over 30 min). After freeze drying, characterisation of the halogenated tyrosine residues 7–11 was performed using LCMS (gradient of 5–35% over 30 min; see ESI Fig. S4–S8†) and by NMR-¹H. Nuclear magnetic resonance spectra were recorded on a Bruker Avance III 400 in CD₃CN/D₂O (v/v; 20/80). Spectra were consistent with literature reports (Br-Tyr (7),²⁶ I-Tyr (8),²⁷ 3,5-dichloro-Tyr (9),²⁸ 3,5-dibromo-Tyr (10)²⁸ and 3,5-diiodo-Tyr (11)²⁸).

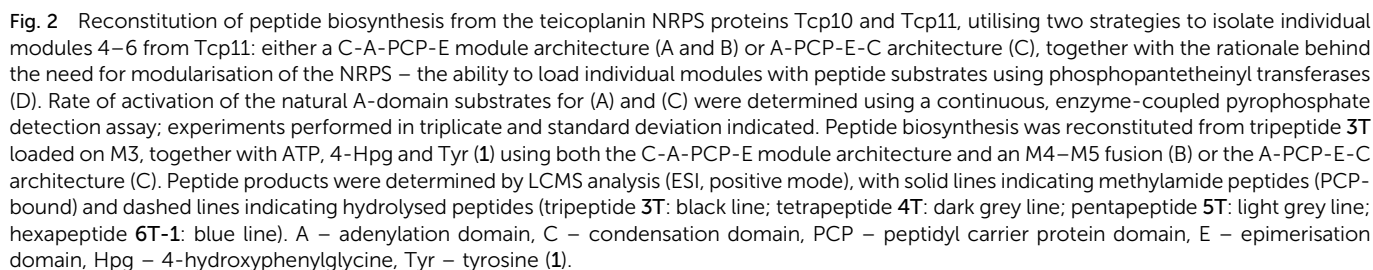
Peptidyl-CoA synthesis

Peptide synthesis for tripeptide **3T-CoA**, tetrapeptide **4T-CoA** and pentapeptide **5T-CoA** CoAs was performed manually by



Chem. Sci., 2019, 10, 9466–9482 | 9471

C-A-PCP-E module architecture (Fig. 2A and B), and determined that whilst the expression of both modules 4 and 6 was satisfactory, module 5 was extremely aggregation-prone and available only in very low yield even when modified with short N/C-terminal sequences taken from other teicoplanin NRPS proteins (ESI Fig. S1†). In contrast, the behaviour of all three modules using the alternate A-PCP-E-C architecture (Fig. 2C) was significantly improved in terms of protein quality and yield (ESI Fig. S2†). This type of modular architecture is reminiscent of the XU building blocks suggested by Bode and co-workers for the



reassembly of NRPS assembly lines *in vivo*, although as far as we are aware this has not been investigated *in vitro*.³⁶

Isolated NRPS modules are catalytically competent

Next, we assessed the activity of the adenylation domains present in these isolated modules from teicoplanin biosynthesis (Fig. 2). Here, we assessed their ability to activate and load amino acids in an ATP-dependent manner using the detection of pyrophosphate to determine the activation rate for each amino acid (Fig. 2A and C).²² For teicoplanin modules 4 and 5, these showed that 4-hydroxyphenylglycine (4-Hpg) was the preferred substrate,³⁷ with very low activity for other structurally related amino acids (such as 3,5-Dpg and phenylglycine). The rate of activation by the C-A-PCP-E modules was faster than the A-PCP-E-C constructs ($\sim 1.5 \text{ min}^{-1}$ vs. $\sim 0.6 \text{ min}^{-1}$), which can be explained by the lack of a C-domain adjacent to the A-domain in case of the A-PCP-E-C constructs. Module 6 from the teicoplanin NRPS demonstrated the activation of Tyr (1) as had been anticipated, and here again the C-A-PCP module was $\sim 3\times$ faster than the A-PCP construct (Fig. 2A and C). However, for this module it was also possible to assay the activity of module 6 in the presence of module 5, which showed that the presence of the adjacent module restored the activity of the M6 A-domain to that observed for other A-domains in this system ($\sim 1.5 \text{ min}^{-1}$). Such interactions between A- and C-domains have been observed previously and can even result in alterations in selectivity of the substrate selected – here, the effect appears restricted to the rate of A-domain substrate activation.^{38,39}

Module 6 A-domains can have broad substrate selectivity

Having determined that the teicoplanin M6 A-domain was able to activate the anticipated Tyr-residue, we next investigated the substrate selectivity of this domain (Fig. 3). Here, we utilised a range of modified Tyr/Phe residues (1–11), and showed that this module showed broad substrate selectivity, accepting Phe (5), 3-chlorotyrosine (Cl-Tyr, 2), β -(R)-OH-Tyr (Bht, 3) and even 3-

chloro- β -(R)-hydroxytyrosine (Cl-Bht, 4). The alteration of the halogen atom did not have a large effect on the selectivity of this A-domain, which could activate F-Tyr (6), Br-Tyr (7) and I-Tyr (8) at similar rates ($\sim 0.7 \text{ min}^{-1}$, Fig. 3). Perhaps even more surprising was the acceptance of 3,5-di-halogenated Tyr residues (3,5-di-Cl-Tyr (9), 3,5-di-Br-Tyr (10) and 3,5-di-I-Tyr (11)). This ability to accept a broad range of different amino acids – in particular the ability to activate Tyr (1) and Phe (5) that would

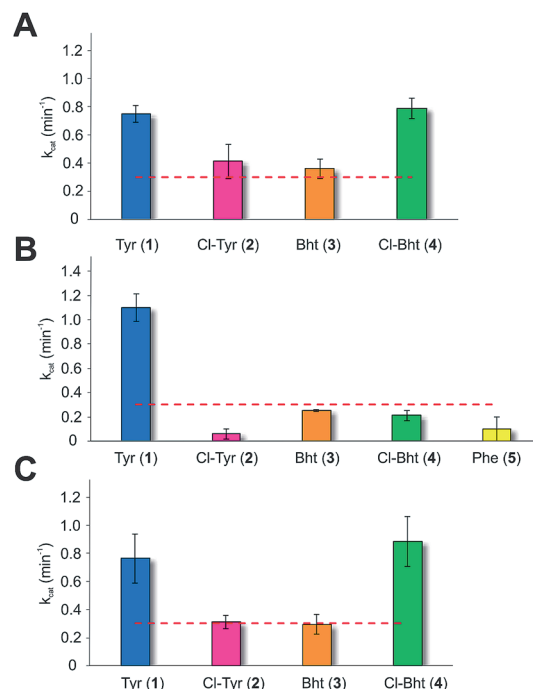


Fig. 4 Module 6 A-domain amino acid activation rates determined for the balhimycin (A) and kistamicin (B) NRPS together with the kistamicin double mutant modelled on the teicoplanin A-domain pocket (C). Aromatic amino acids tested include tyrosine (blue, 1), 3-chlorotyrosine (magenta, 2), Bht (orange, 3), Cl-Bht (green, 4) and phenylalanine (yellow, 5). Triplicate experiments, standard deviations indicated. Dotted line indicates the method detection limit.

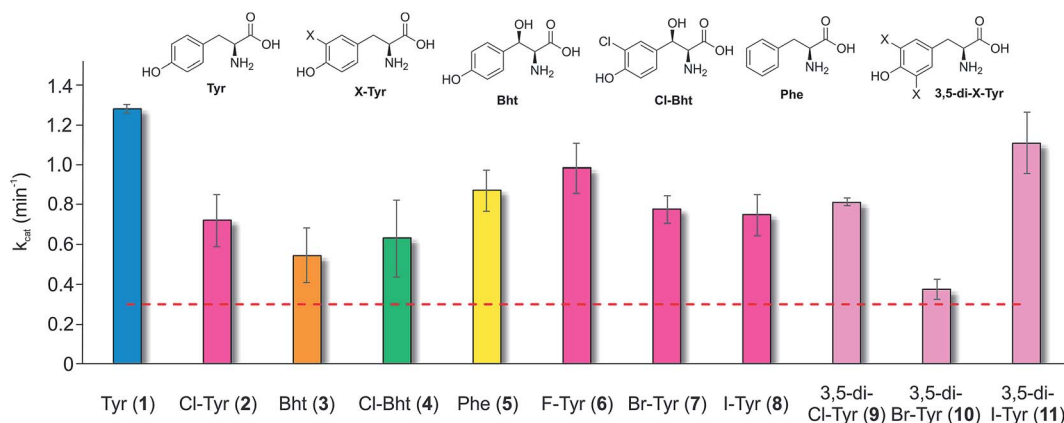
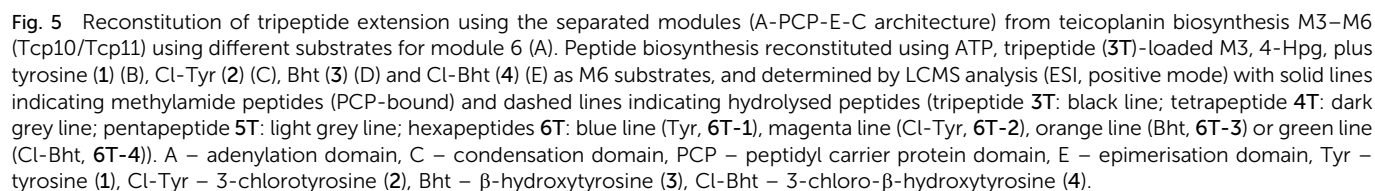


Fig. 3 Module 6 A-domain amino acid activation rates determined for the teicoplanin NRPS. Amino acids tested as substrates include tyrosine (1, blue), halogenated tyrosine residues (2 and 6–8, magenta), di-halogenated tyrosine residues (9–11, pale pink), Bht (orange, 3), Cl-Bht (green, 4) and phenylalanine (yellow, 5). Triplicate experiments, standard deviations indicated. Dotted line indicates the method detection limit.



towards modified Tyr residues (2–4) or Phe (5) (Fig. 4B). The selectivity pocket for the M6 A-domain from kistamicin matches other Tyr specific pockets, whilst the teicoplanin/balhimycin pockets possess two or three different mutations (ESI Table S1[†]). Given that two residues appear to be sufficient to distinguish between permissive and selective Tyr-activating domains (ESI Table S1[†]), we next generated the comparable double mutant (teicoplanin-like pocket) in the kistamicin M6 A-domain. This mutant A-domain was now also able to accept Cl-Bht (4), supporting the hypothesis that these two mutations in the A-domain substrate binding pocket are sufficient to

in the A-domain substrate binding pocket are sufficient to



convert a Tyr-selective domain into a permissive Tyr-activating domain (Fig. 4C).

Hexapeptide forming C-domains are specific for Tyr modification states

Given that the selectivity of the M6 A-domains from teicoplanin and balhimycin biosynthesis was not alone able to explain the specific incorporation of Cl-Bht (4) into the final products of these NRPS assembly lines, we hypothesised that the condensation domain responsible for hexapeptide formation was acting as a selectivity filter to ensure that the M6 bound Tyr (1) residue was properly modified prior to peptide formation. Thus, we next tested the selectivity of the peptide bond forming C-domains responsible for extending the pentapeptide donor substrate onto the M6 PCP-bound Tyr residue acceptor (Fig. 2B and C and 5). As a prequel to these experiments, we first tested the general ability to reconstitute peptide bond formation between the different modules of the teicoplanin NRPS using our purified constructs. To this end, we prepared synthetic peptidyl-CoAs (tripeptide (3T), tetrapeptide (4T) and pentapeptide (5T); ESI Fig. S9–S11†)^{40,41} and loaded these onto the separated modules 4 and 5 as well as the stand-alone module 3 (Tcp10)²² to examine C-domain activity. Following incubation of modules in an equimolar ratio with the exception of M3

(Tcp10), where we used twice as much to boost potential peptide yield, the cleavage of PCP-bound peptides was affected using methylamine, which allowed NRPS-bound peptides to be identified as their methylamide derivatives as opposed to peptides lost from the assembly line through hydrolysis. Results of these single turnover experiments showed that the C-domain activity of the M3/4 and M4/5 C-domains could be effectively reconstituted using 4-Hpg as the A-domain substrate, with activity of the A-PCP-E-C constructs (Fig. 2C) significantly improved over the C-A-PCP-E type (Fig. 2B). The major problem with the C-A-PCP-E constructs was significant levels of hydrolysis observed from module 4; this was only partially overcome by using a fusion of modules 4 and 5 (Fig. 2B), and was still present even in turnovers using full length Tcp11 (ESI Fig. S12†).

Having seen that isolated modules with the A-PCP-E-C construct architecture could be used to reconstitute peptide bond formation, we next turned to the reconstitution of the M5/6 C-domain and different modified Tyr-substrates to examine the effect of different C-domain acceptor substrates on peptide bond formation by this domain (Fig. 5A). In these experiments, we could take advantage of the permissive nature of the teicoplanin M6 A-domain to load various Tyr derivatives (2–4) onto the downstream PCP domain that would normally be generated through the actions of *trans*-modifying enzymes on the PCP-

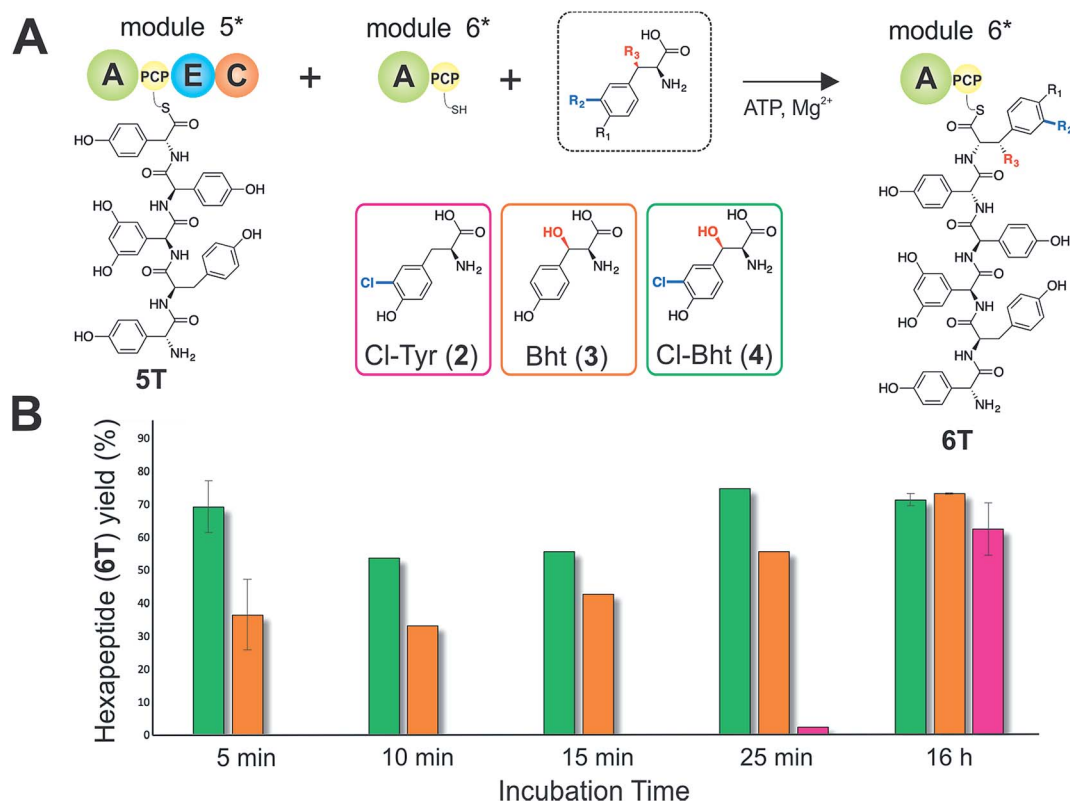


Fig. 6 The rate of pentapeptide extension by teicoplanin module 6 using differently modified tyrosine residues as substrates (A). Hexapeptide biosynthesis reconstituted using ATP, pentapeptide (5T)-loaded M5, plus Cl-Tyr (2), Bht (3) and Cl-Bht (4) as M6 substrates, analysed by LCMS analysis (ESI, positive mode) at various time intervals (B). 5 min and 16 h experiments were performed in triplicate with the standard deviation indicated; other time points are the result of single experiments. A – adenylation domain, C – condensation domain, PCP – peptidyl carrier protein domain, E – epimerisation domain, Cl-Tyr – 3-chlorotyrosine (2), Bht – β -hydroxytyrosine (3), Cl-Bht – 3-chloro- β -hydroxytyrosine (4).

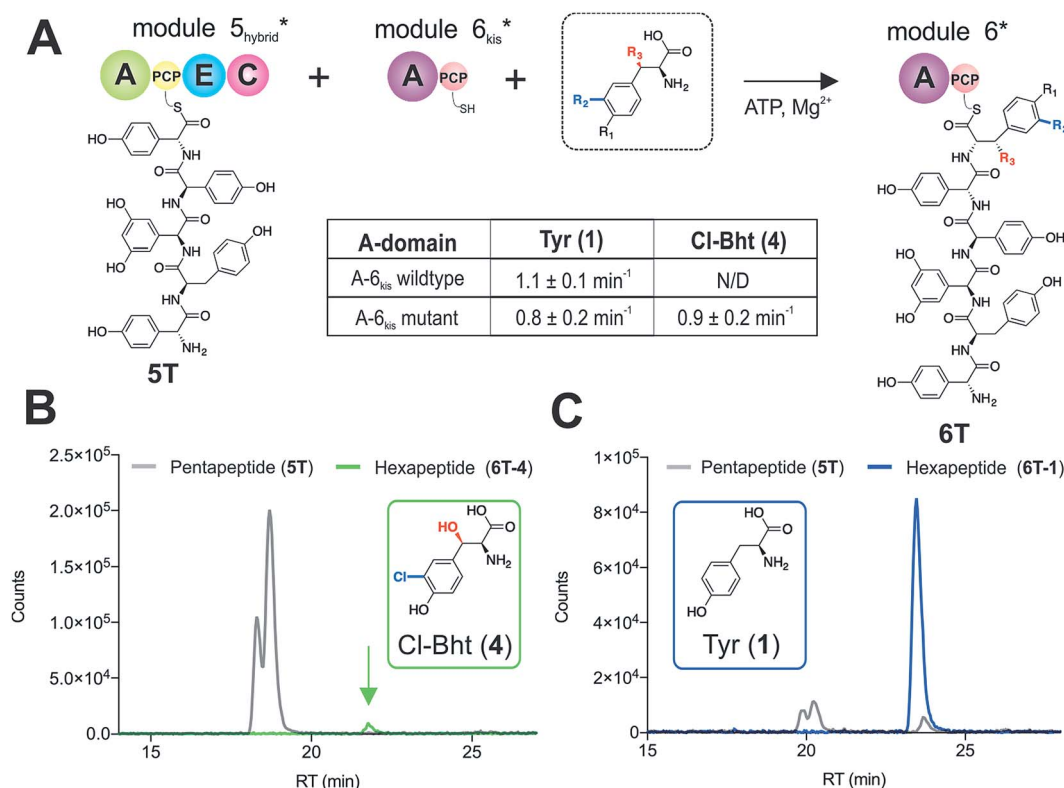


bound Tyr residue, as neither *trans*-interacting enzyme could be isolated in a functional form. To this end, we performed peptide reconstitution assays for the M5/6 C-domain using Tyr (1, Fig. 5B) as well as modified Cl-Tyr (2, Fig. 5C), Bht (3, Fig. 5D) and Cl-Bht (4, Fig. 5E) residues as amino acid substrates and quantified the amount of hexapeptide produced in each case. The results of these experiments clearly showed that the activity of the M5/6 C-domain when using Tyr (1) as an acceptor substrate – whilst detectable – was greatly reduced compared to the other modules after overnight incubation (Fig. 5B). In contrast, all of the modified Tyr residues (2–4) led to hexapeptide formation after overnight incubation, showing the importance of these modifications for the catalytic efficiency of this C-domain. Examining the activity of this C-domain using shorter experiments and only M5 and M6 (Fig. 6A) showed that Cl-Bht (4) was the preferred C-domain acceptor substrate, with Bht (3) also a good acceptor substrate and Cl-Tyr (2) significantly poorer than Cl-Bht (4) or Bht (3) (Fig. 6B). These results are in agreement with our hypothesis that the M5/6 C-domain should be selective for the aminoacyl-PCP acceptor substrate: whilst tyrosine is the substrate anticipated to be loaded by the teicoplanin M6 A-domain *in vivo*,⁴² the biosynthesis of teicoplanin requires two modifications of this residue prior to peptide elongation by this C-domain. This also indicates that the

mechanism of selectivity that lies behind *trans*-enzyme activity in this case is the selectivity of the M5/6 C-domain for the correctly modified state of the aminoacyl-PCP acceptor substrate, with the C-domain selectivity presumably serving to stall peptide bond formation with incompletely modified aminoacyl-PCP intermediates, allowing sufficient time for the *trans*-interacting enzymes to act on these residues.

NRPS assembly lines can be reprogrammed by C-domain exchange

With evidence that the teicoplanin M5/6 C-domain acts as a logic gate to ensure Tyr-6 modification in this system, we now turned our attention to the comparable C-domains from the balhimycin and kistamicin systems.^{32,34} Whilst we were unable to access soluble M5 constructs for either the balhimycin or kistamicin NRPS assembly lines, we were able to express a modified A-PCP-E-C teicoplanin module 5 in which the C-domain was exchanged with the equivalent kistamicin C-domain (Fig. 7A). Whilst it was disappointing to be unable to analyse the activity of the balhimycin M5/6 C-domain, the kistamicin M5/6 C-domain provides a perfect vehicle to examine C-domain selectivity for tyrosine as a substrate, and also the potential viability of altering amino acid modification in *trans* through C-domain exchange within the NRPS machinery.



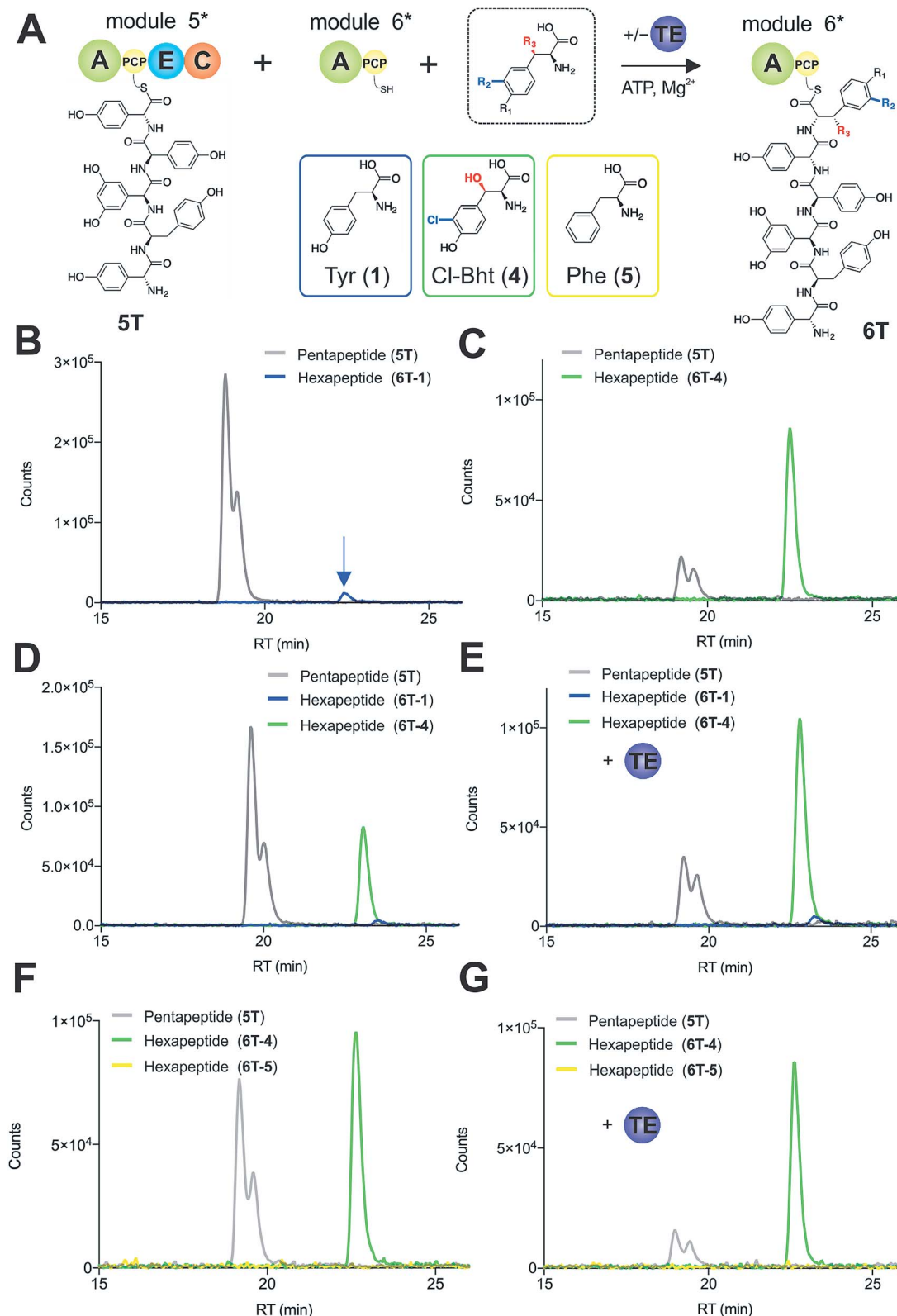
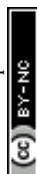


Fig. 8 Reconstitution of pentapeptide extension coupled with the actions of a type-II TE enzyme (A). Peptide biosynthesis reconstituted using ATP, pentapeptide (5T)-loaded M5, plus different combinations of possible amino acids as M6 substrates, as determined by LCMS analysis (ESI, positive mode) with solid lines indicating methylamide peptides (PCP-bound) peptides (pentapeptide 5T: light grey line; hexapeptides 6T: blue line (Tyr, 6T-1), green line (Cl-Bht, 6T-4) or yellow line (Phe, 6T-5)). Results of peptide extension using tyrosine (1) alone (B), Cl-Bht (4) alone (C), an equimolar ratio of Tyr (1) and Cl-Bht (4) (D) as well as an equimolar ratio of Tyr (1) and Cl-Bht (4) together with the incorporation of the type-II TE enzyme Tcp39 (E). Results of peptide extension using an equimolar ratio of Phe (5) and Cl-Bht (4) (F) as well as an equimolar ratio of Phe (5) and Cl-Bht (4) together with the incorporation of the type-II TE enzyme Tcp39 shown (G). A – adenylation domain, C – condensation domain, PCP – peptidyl carrier protein domain, E – epimerisation domain, TE – type-II thioesterase, Tyr – tyrosine (1), Cl-Bht – 3-chloro- β -hydroxytyrosine (4), Phe – phenylalanine (5).



Reconstitution of this hybrid module 5 together with kistamicin module 6 clearly demonstrated the direct incorporation of Tyr (1) into the product hexapeptide, supporting the ability to control peptide extension by the exchange of C-domains exhibiting different aminoacyl-PCP specificities (Fig. 7C). In comparison, the use of teicoplanin module 6 did not lead to efficient reconstitution of hexapeptide formation, which we attribute to the importance of conserving C/A interfaces for effective NRPS reconstitution using our module division strategy. To examine the acceptance of Cl-Bht (4) by the kistamicin M5/6 C-domain, we made use of the double mutant kistamicin M6 construct in which we had altered the amino acid selectivity pocket in the adenylation domain that we had shown to be able to activate Cl-Bht (4). Using this construct, we could show that the kistamicin M5/6 C-domain was now no longer permissive for this larger, modified tyrosine residue – tyrosine was now the preferred substrate (Fig. 7B). These experiments

show that Cl-Bht (4) is not accepted for peptide extension by the kistamicin M5/6 C-domain, and hence that the teicoplanin and kistamicin C-domains show altered specificities for modified Tyr residues as their acceptor substrates during GPA biosynthesis. However, these results did yet not explain how the permissive M6 A-domains present in the NRPS assembly line of both balhimycin and teicoplanin could lead to effective rates of peptide formation in the presence of amino acid substrates that were competitive for the A-domain but that would not be accepted by the C-domain.

Proofreading by the type-II thioesterase

To explore how the NRPS assembly line could function effectively in the presence of competing M6 A-domain substrates, we next turned to competitive assays (Fig. 8A) in which two different amino acid substrates that are both accepted by the

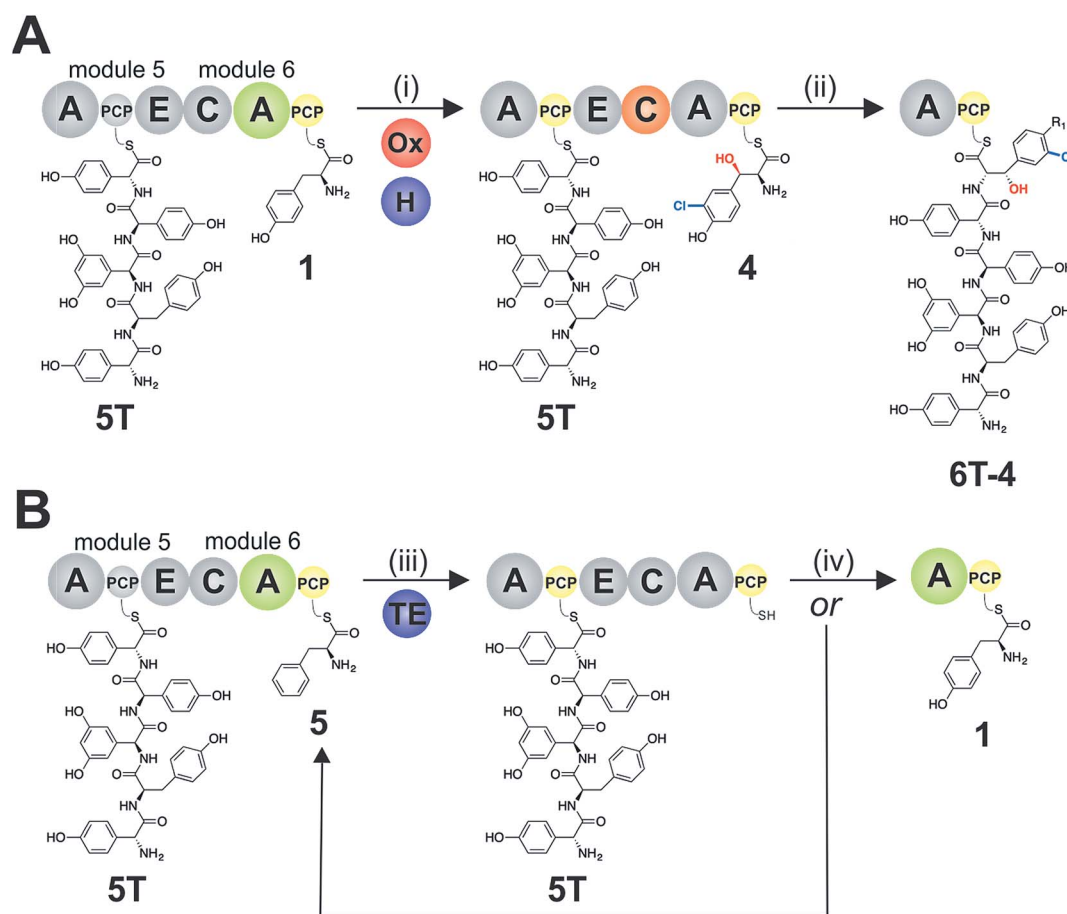


Fig. 9 Schematic representation of A-domain and C-domain selectivity interfacing with the activity of *trans*-modifying enzymes and the type-II TE enzyme during teicoplanin biosynthesis. (A) Pentapeptide 5T extension commences by the activation of Tyr (1) by the M6 A-domain, which is in turn modified by *trans* enzymes (i) to generate PCP-bound Cl-Bht (4), the recognised M5/6 C-domain acceptor substrate; this domain then catalyses peptide extension to generate the desired PCP-bound hexapeptide 6T-4 product (ii). (B) Incorrect amino acid activation (e.g. of Phe (5)) by the M6 A-domain leads to a PCP-bound intermediate that is not accepted by the halogenase or the M5/6 C-domain, which leads to a long-lived aminoacyl-PCP intermediate that is eventually cleaved by the type-II TE enzyme (iii). At this point, another round of amino acid activation catalysed by the M5/6 A-domain can lead to the loading of the correct amino acid (Tyr, iv) and entry into the productive pathway (A, i + ii), or the loading of an incorrect substrate that then will proceed to another round of TE-mediated substrate cleavage (iii). A – adenylation domain, C – condensation domain, PCP – peptidyl carrier protein domain, E – epimerisation domain, H – flavin-dependent halogenase, Ox – non-heme iron oxygenase, TE – type-II thioesterase.



M6 A-domain were included in peptide extension assays (Tyr/Cl-Bht (1/4), Fig. 8B–D; and Phe/Cl-Bht (5/4), Fig. 8F). In these competitive assays, the overall production of hexapeptide **6T-4** was reduced, which is in line with the activation of both substrates by the A-domain but with the product peptide only being formed from PCP₆-bound Cl-Bht (**4**) (Fig. 8D and F) due to the selectivity of the M5/6 C-domain for the acceptor aminoacyl-PCP substrate. To stimulate the removal of the Phe/Tyr (5/1) residues from PCP₆ we then included the Type-II TE enzyme from the teicoplanin system (Tc39)^{31,43} into these assays, which we hypothesised should be active against all PCP-bound amino acid intermediates and that should be effective in removing long-lived PCP-bound amino acids that were not accepted by the M5/6 C-domain. Given that type-II TE enzymes often appear interchangeable amongst different biosynthetic pathways and lack of knowledge concerning the stoichiometry of expression within the teicoplanin system, we tested a range of different TE ratios relative to M6. At relatively low concentrations (2.5 mol%) the inclusion of this TE enzyme enabled the biosynthesis of the hexapeptide **6T-4** to proceed with the same efficiency for the Tyr/Cl-Bht (1/4) and Phe/Cl-Bht (5/4) mixtures as if only Cl-Bht (**4**) was present (Fig. 8E and G). Inclusion of higher ratios of the TE domain beyond 30% of the levels of M6 led to a loss of peptide extension efficiency, presumably due to excessive aminoacyl-PCP hydrolysis in these cases. However, it is important to consider that in the natural system the TE/PCP ratio would be expected to be <15 mol% at a 1 : 1 NRPS/TE stoichiometry, and that the need for low levels of TE in these *in vitro* assays stems from the fact that the majority of the NRPS machinery is absent (ESI Fig. S13†). These experiments show how the selectivity of the M5/6 C-domain for the acceptor substrate works together with a type-II TE domain to ensure that non-ribosomal peptide biosynthesis still produces a highly specific peptide product even when A-domain selectivity is limited. This is particularly relevant with the high activity of the teicoplanin M6 A-domain for Phe (**5**), which is naturally present in cells and would compete with Tyr (**1**) for incorporation into GPAs if no C-domain gating mechanism was present. We also tested whether the same effect could be observed when using the separate TE domain (Bhp) responsible for cleavage of PCP-Bht from the stand alone BpsD (A-PCP) NRPS module that produces Bht (β-hydroxytyrosine, **3**) in balhimycin biosynthesis.^{19,30,44} In these GPA systems, Bht (**3**) is produced by Cytochrome P450 mediated hydroxylation of PCP-Tyr before being hydrolysed specifically through the actions of Bhp.^{29,44} Here, our experiments showed that Bhp was not able to replicate the proofreading function of the Tc39 enzyme (data not shown), which fits with the selectivity of this enzyme for one specific PCP (that found in BpsD) as opposed to the PCP domains from the main peptide producing NRPS.

Discussion

Within non-ribosomal peptide biosynthesis, one of the major issues that slows the reengineering of these systems is the complexity of the peptide biosynthesis process, including the large number of different catalytic domains that can be active in

both *cis* and *trans* on PCP-bound substrates. Whilst the typical functions of the different NRPS domains and ancillary proteins are known, the effects that the specificities of multiple catalytic domains on a single step in NRPS-mediated peptide extension have not been widely investigated. This is particularly the case for the activity of condensation domains, where the classical role of these domains as stereochemical gatekeepers^{4,6–8} has been recently been supplemented with a number of extended catalytic functions in a variety of systems.^{45–48} Beyond extending the activity of C-domains away from peptide bond formation, a major question remains as to what extent these domains act as a selectivity filter for their acceptor aminoacyl-PCP substrates with regards to modification of aminoacyl-PCP domains in *trans*, given that the mechanism leading to the specific targeting of desired PCP-bound substrates has not previously been resolved and can lead to unexpected products after biosynthetic redesign.⁴⁹ In this work, we have concentrated on understanding the pentapeptide extension step during the non-ribosomal biosynthesis of GPAs, given that there is not only significant diversity within the biosynthesis of GPAs for the modification state of the Tyr residue present on module 6 of the NRPS, but also several different mechanisms by which modification of this residue is performed.^{14,19,42} Most GPAs demonstrate both chlorination and hydroxylation of the M6 Tyr residue, and within teicoplanin biosynthesis this is unarguably a complex process, requiring two distinct modifications of this PCP-bound tyrosine residue. Despite the requirement for two *trans*-interacting enzymes to modify this PCP-bound Tyr residue, this process is also very efficient at ensuring the product GPA bears both substituents at this position. Now, through the reconstitution of pentapeptide extension *in vitro*, we have determined by proxy that the actions of these *trans*-modifying enzymes are gated by the acceptor selectivity of the M5/6 C-domain. Whilst C-domains are known to be able to enforce selectivity for their acceptor substrates,⁴ this is the first time to our knowledge that this selectivity has been demonstrated to be present and to control the modification of PCP-bound amino acids *via* the actions of *trans*-acting enzymes. The low acceptance of partially modified Tyr residues as well as very low levels of Tyr incorporation into hexapeptide also provides an explanation why only small amounts of such peptide products can be observed from modified strains in which the *trans*-modifying enzymes have been disrupted.^{20,21,50} Previous work has shown that the halogenation of amino acids during GPA biosynthesis occurs on aminoacyl-PCP substrates, but that the halogenase itself could not distinguish between the PCP domains whose substrates either should (PCP₂, PCP₆) or should not (PCP₁) be chlorinated.¹⁸ The preference of the M5/6 C-domain for Bht (**3**) over Cl-Tyr (**2**) suggests that the halogenase would be expected to act before the hydroxylase to ensure effective incorporation of the Cl-Bht (**4**) residue in teicoplanin biosynthesis. This would further suggest that the halogenase should have higher affinity for the PCP-bound substrate than the hydroxylase, and as the selectivity is unable to be reconstituted using an isolated PCP this further suggests that the interaction interface utilised by the halogenase lies elsewhere in module 6.¹⁸ The need for C-domain scanning of the acceptor-



Beyond the identification of C-domain selectivity as a mechanism to ensure effective *trans*-enzyme modification of aminoacyl-PCPs, in this work we could show that exchange of this M5/6 C-domain for the comparable domain from the kistamicin system (that contains an unmodified Tyr-6)³⁴ allows this selectivity filter in NRPS biosynthesis to be controlled by switching C-domains that display different acceptor specificities. In order to assay the acceptance of Cl-Bht (**4**) by the kistamicin M5/6 C-domain we first had to reduce the selectivity of the kistamicin M6 A-domain for tyrosine by converting this domain into a teicoplanin-like M6 A-domain, as the protein interface between M5/6 C-domain and the teicoplanin M6 A-domain was not conducive to effective peptide biosynthesis. This highlights the importance of maintaining the C/A domain interface in order to allow the modules to reassemble when adopting the A-PCP-E-C module architecture, and is supported by structures of complete NRPS modules indicating a large C/A yet smaller A/C interface.^{51,63,64} Altering the A-domain selectivity of the kistamicin M6 A-domain was successfully accomplished by incorporating two mutations in the A-domain pocket (I5 to V; G7 to A), which supports the role of these mutations in the reduced substrate selectivity observed for some Tyr-activating A-domains found in GPA biosynthesis (see ESI Table S1†). These

results demonstrate that the selectivity of *trans*-acting NRPS enzymes need not be strictly tied to PCP-recognition, but can be controlled through the gating of the modification process by the upstream C-domain. This indicates that the modification of NRPS assembly lines *in vivo* through the alteration of *trans*-interacting enzymes first requires an understanding of the inherent specificities of the C-domains within these NRPS assembly lines. Once characterised, such exchange reactions *via* NRPS hybridisation approaches could be used to generate the specific, desired modification of non-ribosomally produced peptides *in vivo*. Taken together, our results show how A-domains, C-domains and housekeeping type-II TE enzymes can act together to ensure that NRPS-mediated peptide biosynthesis retains high selectivity even when A-domains – typically thought of as highly selective domains responsible for controlling peptide sequence – display reduced specificities for monomer selection. This implies that a combination of A- and C-domain redesign should be included in reengineering paradigms for NRPS biosynthesis pathways, and once more underpins the importance of a holistic approach for studying these fascinating molecular machines that produce many of our most important medicinal agents.

Author contributions

M. J. C. designed the study. M. K. cloned constructs, expressed proteins and performed all activity assays. J. T. synthesised and purified peptidyl-CoA substrates; E. A. M. synthesised and purified halogenated amino acids. R. J. A. G. and R. B. S. performed and analysed HRMS and MS² experiments. M. J. C. wrote the manuscript and prepared the figures with input from all co-authors.

Conflicts of interest

The author(s) declare no competing interests.

Acknowledgements

G. Stier (BZH-Heidelberg) for fusion protein vectors and J. Yin (University of Chicago) for the R4-4 Sfp expression plasmid; M. H. Hansen for assistance with sequence alignments; G. L. Challis (University of Warwick/Monash University), J. J. De Voss (University of Queensland), E. Stegmann and N. Ziemert (University of Tübingen) for helpful discussions. This work was supported by Monash University, EMBL Australia and the National Health and Medical Research Council (APP1140619 to (M. J. C.)) and further supported under the Australian Research Council's Discovery Projects funding scheme (project number DP170102220 and DP190101272 to M. J. C.).

References

- 1 R. D. Süßmuth and A. Mainz, *Angew. Chem., Int. Ed.*, 2017, **56**, 3770–3821.
- 2 C. T. Walsh, R. V. O'Brien and C. Khosla, *Angew. Chem., Int. Ed.*, 2013, **52**, 7098–7124.
- 3 T. Izoré and M. J. Cryle, *Nat. Prod. Rep.*, 2018, **35**, 1120–1139.
- 4 P. J. Belshaw, C. T. Walsh and T. Stachelhaus, *Science*, 1999, **284**, 486–489.
- 5 A. Stanišić and H. Kries, *ChemBioChem*, 2019, **20**, 1347–1356.
- 6 L. Luo, R. M. Kohli, M. Onishi, U. Linne, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2002, **41**, 9184–9196.
- 7 S. L. Clugston, S. A. Sieber, M. A. Marahiel and C. T. Walsh, *Biochemistry*, 2003, **42**, 12095–12104.
- 8 K. Bloudoff, C. D. Fage, M. A. Marahiel and T. M. Schmeing, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, 95–100.
- 9 M. E. Horsman, T. P. A. Hari and C. N. Boddy, *Nat. Prod. Rep.*, 2016, **33**, 183–202.
- 10 D. B. Stein, U. Linne and M. A. Marahiel, *FEBS J.*, 2005, **272**, 4506–4520.
- 11 S. Mori, A. H. Pang, T. A. Lundy, A. Garzan, O. V. Tsodikov and S. Garneau-Tsodikova, *Nat. Chem. Biol.*, 2018, **14**, 428–430.
- 12 A. Greule, J. E. Stok, J. J. De Voss and M. J. Cryle, *Nat. Prod. Rep.*, 2018, **35**, 757–791.
- 13 V. Agarwal, Z. D. Miles, J. M. Winter, A. S. Eustáquio, A. A. El Gamal and B. S. Moore, *Chem. Rev.*, 2017, **117**, 5619–5674.
- 14 G. Yim, M. N. Thaker, K. Koteva and G. Wright, *J. Antibiot.*, 2014, **67**, 31–41.
- 15 E. Stegmann, H.-J. Frasch and W. Wohlleben, *Curr. Opin. Microbiol.*, 2010, **13**, 595–602.
- 16 M. Peschke, M. Gonsior, R. D. Süßmuth and M. J. Cryle, *Curr. Opin. Struct. Biol.*, 2016, **41**, 46–53.
- 17 K. Haslinger, M. Peschke, C. Brieke, E. Maximowitsch and M. J. Cryle, *Nature*, 2015, **521**, 105–109.
- 18 T. Kittilä, C. Kittel, J. Tailhades, D. Butz, M. Schoppet, A. Büttner, R. J. A. Goode, R. B. Schittenhelm, K.-H. van Pee, R. D. Süßmuth, W. Wohlleben, M. J. Cryle and E. Stegmann, *Chem. Sci.*, 2017, **8**, 5992–6004.
- 19 O. Puk, D. Bischoff, C. Kittel, S. Pelzer, S. Weist, E. Stegmann, R. D. Süßmuth and W. Wohlleben, *J. Bacteriol.*, 2004, **186**, 6093–6100.
- 20 S. Stinchi, L. Carrano, A. Lazzarini, M. Feroggio, A. Grigoletto, M. Sosio and S. Donadio, *FEMS Microbiol. Lett.*, 2006, **256**, 229–235.
- 21 F. Beltrametti, A. Lazzarini, C. Brunati, A. Marazzi, S. Jovetic, E. Selva and F. Marinelli, *J. Antibiot.*, 2003, **56**, 773–782.
- 22 T. Kittilä, M. Schoppet and M. J. Cryle, *ChemBioChem*, 2016, **17**, 576–584.
- 23 J. Bogomolovas, B. Simon, M. Sattler and G. Stier, *Protein Expression Purif.*, 2009, **64**, 16–23.
- 24 M. Ralser, R. Querfurth, H.-J. Warnatz, H. Lehrach, M.-L. Yaspo and S. Krobitsch, *Biochem. Biophys. Res. Commun.*, 2006, **347**, 747–751.
- 25 M. Sunbul, N. J. Marshall, Y. Zou, K. Zhang and J. Yin, *J. Mol. Biol.*, 2009, **387**, 883–898.
- 26 Y. Yokoyama, T. Yamaguchi, M. Sato, E. Kobayashi, Y. Murakami and H. Okuno, *Chem. Pharm. Bull.*, 2006, **54**, 1715–1719.
- 27 J. R. Cochrane, J. M. White, U. Wille and C. A. Hutton, *Org. Lett.*, 2012, **14**, 2402–2405.
- 28 X. Xu, R. Ge, L. Li, J. Wang, X. Lu, S. Xue, X. Chen, Z. Li and J. Bian, *Eur. J. Med. Chem.*, 2018, **143**, 1325–1344.



- 29 M. J. Cryle, A. Meinhart and I. Schlichting, *J. Biol. Chem.*, 2010, **285**, 24562–24574.
- 30 O. Puk, P. Huber, D. Bischoff, J. Recktenwald, G. Jung, R. D. Süssmuth, K.-H. van Pée, W. Wohlleben and S. Pelzer, *Chem. Biol.*, 2002, **9**, 225–235.
- 31 M. Sosio, H. Kloosterman, A. Bianchi, P. de Vreugd, L. Dijkhuizen and S. Donadio, *Microbiology*, 2004, **150**, 95–102.
- 32 S. Pelzer, R. Süssmuth, D. Heckmann, J. Recktenwald, P. Huber, G. Jung and W. Wohlleben, *Antimicrob. Agents Chemother.*, 1999, **43**, 1565–1573.
- 33 B. Nazari, C. C. Forneris, M. I. Gibson, K. Moon, K. R. Schramma and M. R. Seyedsayamdost, *MedChemComm*, 2017, **8**, 780–788.
- 34 A. Greule, T. Izoré, D. Iftime, J. Tailhades, M. Schoppet, Y. Zhao, M. Peschke, I. Ahmed, A. Kulik, M. Adamek, R. J. A. Goode, R. B. Schittenhelm, J. A. Kaczmariski, C. J. Jackson, N. Ziemert, E. H. Krenske, J. J. De Voss, E. Stegmann and M. J. Cryle, *Nat. Commun.*, 2019, **10**, 2613.
- 35 J. Yin, A. J. Lin, D. E. Golan and C. T. Walsh, *Nat. Protoc.*, 2006, **1**, 280.
- 36 K. A. J. Bozhüyük, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C.-P. Niesert and H. B. Bode, *Nat. Chem.*, 2017, **10**, 275.
- 37 R. S. Al Toma, C. Brieke, M. J. Cryle and R. D. Süssmuth, *Nat. Prod. Rep.*, 2015, **32**, 1207–1235.
- 38 S. Meyer, J.-C. Kehr, A. Mainz, D. Dehm, D. Petras, R. D. Süssmuth and E. Dittmann, *Cell Chem. Biol.*, 2016, **23**, 462–471.
- 39 R. Li, R. A. Oliver and C. A. Townsend, *Cell Chem. Biol.*, 2017, **24**, 24–34.
- 40 J. Tailhades, M. Schoppet, A. Greule, M. Peschke, C. Brieke and M. J. Cryle, *Chem. Commun.*, 2018, **54**, 2146–2149.
- 41 C. Brieke, V. Kratzig, K. Haslinger, A. Winkler and M. J. Cryle, *Org. Biomol. Chem.*, 2015, **13**, 2012–2021.
- 42 J. Recktenwald, R. Shawky, O. Puk, F. Pfennig, U. Keller, W. Wohlleben and S. Pelzer, *Microbiology*, 2002, **148**, 1105–1118.
- 43 O. Yushchuk, B. Ostash, T. H. Pham, A. Luzhetskyy, V. Fedorenko, A. W. Truman and L. Horbal, *ACS Chem. Biol.*, 2016, **11**, 2254–2264.
- 44 S. Mulyani, E. Egel, C. Kittel, S. Turkanovic, W. Wohlleben, R. D. Süssmuth and K.-H. van Pée, *ChemBioChem*, 2010, **11**, 266–271.
- 45 N. M. Gaudelli, D. H. Long and C. A. Townsend, *Nature*, 2015, **520**, 383–387.
- 46 J. B. Patteson, Z. D. Dunn and B. Li, *Angew. Chem., Int. Ed.*, 2018, **57**, 6780–6785.
- 47 T. Izoré, J. Tailhades, M. H. Hansen, J. A. Kaczmariski, C. J. Jackson and M. J. Cryle, *Proc. Natl. Acad. Sci.*, 2019, **116**, 2913–2918.
- 48 B. J. C. Law, Y. Zhuo, M. Winn, D. Francis, Y. Zhang, M. Samborsky, A. Murphy, L. Ren, P. F. Leadlay and J. Micklefield, *Nat. Catal.*, 2018, **1**, 977–984.
- 49 X. Yin, Y. Chen, L. Zhang, Y. Wang and T. M. Zabriskie, *J. Nat. Prod.*, 2010, **73**, 583–589.
- 50 M. Schoppet, M. Peschke, A. Kirchberg, V. Wiebach, R. D. Süssmuth, E. Stegmann and M. J. Cryle, *Chem. Sci.*, 2019, **10**, 118–133.
- 51 M. J. Tarry, A. S. Haque, K. H. Bui and T. M. Schmeing, *Structure*, 2017, **25**, 783–793.
- 52 F. Ishikawa, A. Miyanaga, H. Kitayama, S. Nakamura, I. Nakanishi, F. Kudo, T. Eguchi and G. Tanabe, *Angew. Chem., Int. Ed.*, 2019, **58**, 6906–6910.
- 53 N. Kessler, H. Schuhmann, S. Morneweg, U. Linne and M. A. Marahiel, *J. Biol. Chem.*, 2004, **279**, 7413–7419.
- 54 E. Yeh, R. M. Kohli, S. D. Bruner and C. T. Walsh, *ChemBioChem*, 2004, **5**, 1290–1293.
- 55 D. Schwarzer, H. D. Mootz, U. Linne and M. A. Marahiel, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 14083–14088.
- 56 O. Nureki, D. G. Vassilyev, M. Tateno, A. Shimada, T. Nakama, S. Fukai, M. Konno, T. L. Hendrickson, P. Schimmel and S. Yokoyama, *Science*, 1998, **280**, 578–582.
- 57 E. Schmidt and P. Schimmel, *Science*, 1994, **264**, 265–267.
- 58 T. Hussain, V. Kamarthapu, S. P. Kruparani, M. V. Deshmukh and R. Sankaranarayanan, *Proc. Natl. Acad. Sci.*, 2010, **107**, 22117–22121.
- 59 J. Ling, N. Reynolds and M. Ibba, *Annu. Rev. Microbiol.*, 2009, **63**, 61–78.
- 60 M. Kotowska and K. Pawlik, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 7735–7746.
- 61 M. L. Heathcote, J. Staunton and P. F. Leadlay, *Chem. Biol.*, 2001, **8**, 207–220.
- 62 P. A. Storm and C. A. Townsend, *Chem. Commun.*, 2018, **54**, 50–53.
- 63 E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlov, C. Leigh Allen, G. Skiniotis, C. C. Aldrich and A. M. Gulick, *Nature*, 2016, **529**, 235–238.
- 64 A. Tanovic, S. A. Samel, L.-O. Essen and M. A. Marahiel, *Science*, 2008, **321**, 659–663.

