Understanding condensation domain selectivity in non-ribosomal peptide biosynthesis: structural characterization of the acceptor bound state

Max Cryle (max.cryle@monash.edu)  
Monash University  https://orcid.org/0000-0002-9739-6157

Thierry Izore  
Monash University

Y. T. Ho  
Monash University

Joe Kaczmarski  
Australian National University  https://orcid.org/0000-0001-8549-0122

Athina Gavrilidou  
University of Tuebingen

Ka Chow  
University of Queensland

David Steer  
Monash University

Robert Goode  
Monash University  https://orcid.org/0000-0002-3028-8928

Ralf Schittenhelm  
Monash University  https://orcid.org/0000-0001-8738-1878

Julien Tailhades  
Monash University

Manuela Tosin  
University of Warwick

Gregory Challis  
University of Warwick  https://orcid.org/0000-0001-5976-3545

Elizabeth Krenske  
The University of Queensland  https://orcid.org/0000-0003-1911-0501

Nadine Ziemert  
University of Tübingen  https://orcid.org/0000-0002-7264-1857

Colin Jackson  
Australian National University  https://orcid.org/0000-0001-6150-3822
Article

Keywords: Complex Peptide Natural Products, Multi-modular Assembly Lines, Chain Assembly, Acceptor Substrates

DOI: https://doi.org/10.21203/rs.3.rs-125509/v1

License: © This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Non-ribosomal peptide synthetases are important enzymes for the assembly of complex peptide natural products. Within these multi-modular assembly lines, condensation domains perform the central function of chain assembly, typically by forming a peptide bond between two peptidyl carrier protein (PCP)-bound substrates. In this work, we report the first structural snapshots of a condensation domain in complex with an aminoacyl-PCP acceptor substrate. These structures allow the identification of a mechanism that controls access of acceptor substrates to the active site in condensation domains. The structures of this previously uncharacterized complex also allow us to demonstrate that condensation domain active sites do not contain a distinct pocket to select the side chain of the acceptor substrate during peptide assembly but that residues within the active site motif can instead serve to tune the selectivity of these central biosynthetic domains.

Introduction

Non-ribosomal peptide synthetases (NRPSs) are important biosynthetic enzymes for the production of highly diverse and extensively modified peptides.\textsuperscript{1} The diversity of non-ribosomal peptides is due to the combination of an ability to incorporate an expanded range of monomers compared to ribosomal peptide biosynthesis together with extensive modifications of the peptide both during and after chain assembly.\textsuperscript{2} This is enabled by the modular architecture of NRPSs, which use repeating groups of catalytic domains to install one monomer into the growing peptide (Fig. 1A). Within a minimal chain extension module, an adenylation (A) domain performs the selection and activation of amino acid building blocks at the expense of ATP, prior to the loading of the monomer onto the phosphopantetheinyI (PPant) moiety of an adjacent peptidyl carrier protein (PCP) domain.\textsuperscript{1} Chain assembly is then performed by condensation (C) domains, which typically accept two PCP-bound substrates and catalyze peptide bond formation through the attack of the downstream acceptor substrate upon the thioester of the upstream donor substrate (Fig. 1B).\textsuperscript{3} The first X-ray crystal structure of an NRPS C domain (VibH from the vibriobactin NRPS, Fig. 1C)\textsuperscript{4} showed that they comprise a pseudo-dimer of the chloramphenicol acetyl transferase (CAT) enzyme fold, with key catalytic residues forming a conserved HHxxxDG motif located at the interface between the two subdomains. In addition, it was shown that C domains harbor two catalytic tunnels that lead from the donor-PCP and acceptor-PCP domain docking sites to the active site and represent the access route for the donor and the acceptor substrates, respectively. This architecture has since been confirmed by other structures.\textsuperscript{4–16} While the conserved central histidine (HHxxxDG) is generally thought to act as the primary catalytic residue that promotes deprotonation of the \textalpha-amino group in the acceptor aminoacyl-PCP as it attacks the thioester, this remains a matter of debate.\textsuperscript{3} Perhaps more importantly, the role C domains play in determining NRPS specificity is unclear, in part due to the lack of structural characterization of relevant PCP-bound C domain complexes.

Whilst the modular architecture of NRPSs has attracted great interest from the perspective of biosynthetic engineering,\textsuperscript{17–19} such efforts have not always been successful. This can be attributed to the complexity
of the machinery combined with the necessity for non-native substrates to pass through multiple catalytic domains, each of which imparts a degree of specificity. A pertinent example of this is the recent recognition of the diverse functions of C domains in peptide biosynthesis, extending their well-established role in controlling peptide stereochemistry (working in concert with epimerization (E) domains) to gating in trans modifications, recruiting trans-acting enzymes and performing additional chemical transformations of their substrates during peptide bond formation (Fig. 1C).\textsuperscript{20–24} Whilst A domains are the main origin of structural diversity in non-ribosomal peptides,\textsuperscript{25} C domains play a key role in peptide bond formation and make important contributions to structural diversification in many valuable compound classes. Thus, gaining a deeper understanding of their function a high priority.

The structural analysis of key domains, complexes and complete modules has made major contributions to our understanding of how selectivity is achieved by NRPS assembly lines.\textsuperscript{26} Whilst the structures of isolated modules are incredibly informative, NRPSs are highly flexible\textsuperscript{7,13,27,28} and the interactions between individual domains change during the process of chain assembly. Structural characterization of key domain–domain complexes is thus an important goal to improve our understanding of NRPS selectivity. For example, structures of A domains in complex with PCP domains in distinct states, corresponding to substrate binding, substrate activation and PPant loading have provided insight into the mechanisms underlying A domain selectivity \textsuperscript{10,29} However, C domains and C domain-containing complexes have proved more challenging to structurally characterize, with fewer examples reported to date (Fig. 1C).\textsuperscript{3,26} Furthermore, no structures of a C domain in complex with an acceptor PCP-domain bearing a substrate have been reported, which makes understanding the origins of C domain specificity for their acceptor substrates unclear, and also limits our understanding of the role of active site residues in C domain catalysis.\textsuperscript{3}

To address this, we report the structure and biochemical characterization of complexes of a PCP domain bearing a stable analogue of the acyl acceptor complexed to the acceptor site of a C domain from the NRPS that biosynthesizes fuscachelin in the thermophile \textit{Thermobifida fusca} (Fig. 1A).\textsuperscript{30} This structure reveals that the interface between the PCP and C domains is dominated by hydrophobic interactions and that access to the C domain active site is gated by an arginine residue that prevents unloaded PCP-substrates from accessing the active site of the C domain. The C domain is shown to be tolerant of a range of aliphatic acceptor amino acid acceptor substrates, with the limited acceptance of other substrates rationalized through interactions with key residues within the C domain active site. We demonstrate that C domains do not appear to contain an “A domain-like” side chain selectivity pocket to control their acceptor substrates and resolve how substrates engage with central catalytic residues in C domains, both of which are key unanswered questions central to NRPS-mediated peptide biosynthesis.

\textbf{Results}

\textbf{Structure of the PCP\textsubscript{2}-C\textsubscript{3} didomain.} To elucidate the structure of C domain with a PCP domain bound in the acceptor site, we screened several systems including a thermophilic example of a PCP\textsubscript{2}-C\textsubscript{3} didomain
(containing the PCP domain from module 2 and the C domain from module 3) of the fuscachelin NRPS from the thermophilic organism *Thermobifida fusca* (Fig. 1A [red rectangle]). Expression of the fuscachelin PCP$_2$-C$_3$ didomain in *E. coli* yielded 0.8 mg/L of culture of stable protein and afforded crystals that grew rapidly in 18–22% w/v PEG 3350 and 0.17–0.3 M magnesium formate at room temperature. Crystals were harvested, cryoprotected in 20–30% glycerol and diffraction data collected at the Australian Synchrotron, with initial phases obtained from a single anomalous diffraction experiment (SAD) using xenon-derivatised crystals (see Methods). The crystals belonged to the P2$_1$2$_1$2$_1$ space group, with the unit cell comprising two highly similar copies of the PCP$_2$-C$_3$ construct (RMSD (all atoms) 0.74 Å).

The PCP$_2$-C$_3$ didomain structure we obtained from these experiments was solved at a resolution of 2.2 Å (Fig. 2A, SI Table S1). When considered separately, the overall folds of both the PCP$_2$ domain and C$_3$ domain were consistent with previously reported structures. The PCP$_2$ domain comprises a 4-helix bundle with a small α-turn between helices 1 and 2 (seen in most crystal structures but absent from NMR structures); the serine residue that is the site of 4'-phosphopantetheine (PPant) attachment is located at the start of helix 2 (Fig. 2B). Of the published crystal structures of PCP domains, this structure is most similar to the PCP domain found in the PCP-Te/R didomain NRPS construct from the archaeon *Methanobrevibacter ruminantium* M1 (PDB ID 6VTJ; RMSD (all atoms) 1.3 Å, 32% sequence similarity, see SI Table S2). The C$_3$ domain of the didomain resembles other members of its class (see SI Table S3), comprising a pseudo-dimer of CAT domains with bridge (R2923 to T2944) and floor loop (A2843 to L2858) regions (Figs. 1C & 2C). The catalytic residues sit at the core of the C$_3$ domain and can be accessed from the bulk solvent via tunnels formed along the interface of the two pseudo-domains (Fig. 2D). Differences in the relative position of these two halves are observed in structures of C domain homologs and can alter the size and character of the acceptor and donor catalytic tunnels. A superimposition of the fuscachelin C$_3$ domain with two well-characterized C domains (from surfactin and linear gramicidin NRPSs) highlights this, with a pronounced difference in displacements observed when comparing the fuscachelin C$_3$ domain and Srf-A domain (SI Figure S1). This aspect of C-domain conformational flexibility and diversity is currently not broadly understood, although recent efforts have been made to understand these conformational differences in terms of the accessibility of the substrates to the active site the C-domain.

In the PCP$_2$-C$_3$ didomain structure, the PCP$_2$ domain sits at the acceptor-PCP binding site (near the opening of the acceptor substrate channel) on the C$_3$ domain from the second chain in the asymmetric unit. The interface between the PCP$_2$ domain and C$_3$ domain is mostly hydrophobic in nature (537/510 Å$^2$ buried surface area (chain A/B) excluding PPant), with the side chains of V2534, L2515, L2518, F2508 and F2538 of the PCP domain playing a major role in the interaction along with residues A2907, V2908, V2584, L2580 and W2579 in the C domain (Fig. 2E, SI Tables S4-5). This interface is reminiscent of the hydrophobic interaction pattern described in other structures of PCP domains docked at the acceptor site of C domains (SrfA-C (PDB ID 2VSQ), AB3403 (PDB ID 4ZXH), see also 26). These interfaces center
around a hydrophobic residue (L2515) immediately following the serine to which the PPant is attached (S2514) and at least one hydrophobic residue ~ 20 amino-acids after the serine residue. R2906 also plays an important role in positioning the docked PCP domain via interactions with the phosphate moiety of the PPant arm. In the PCP<sub>2</sub>-C<sub>3</sub> structure, these residues are V2534 and the aliphatic moiety of R2535 that interacts with V2908 of the C domain. The overall orientation of the PCP domain relative to the C domain is similar to what has been observed in the structures of SrfA-C<sup>12</sup> and ObiF1 (PDB ID 6N8E)<sup>8</sup> (SI Figure S2A-B), whilst other structures contain a PCP domain that is rotated by several degrees around the conserved serine (AB3403,<sup>10</sup> LgrA (PDB ID 6MFZ)<sup>7</sup>) (SI Fig. 2C-D). Although the overall orientation of these PCP domains in relation to the C domain are different, it is important to note that the position of the PPant-modified serine (located at the beginning of the second helix) is always maintained at the entrance of the acceptor substrate channel of the C domain.

Since the PCP<sub>2</sub> domain precedes the C<sub>3</sub> domain in the fusachelin NRPS, we had expected that the PCP<sub>2</sub> domain would dock at the donor-PCP binding site of the C<sub>3</sub> domain. We were surprised, therefore, to find that this construct crystallized with the PCP<sub>2</sub> domain docked into the acceptor-PCP binding site of a symmetry-related C<sub>3</sub> domain (Fig. 2A). Given that the PCP<sub>2</sub> and PCP<sub>3</sub> domains of the fusachelin NRPS are highly similar (65% sequence identity, Fig. 3), and that PCP domains can act as both aminoacyl donors and acceptors for C domains, we rationalized that the arrangement observed in our structure is a valid model of an acceptor-PCP-bound C domain. Indeed, when we determined the structure of the isolated PCP<sub>3</sub> domain, we found its structure to be highly similar to the PCP<sub>2</sub> domain (RMSD (all atoms) 2 Å; Fig. 3A-C). Importantly, the residues at the interface with the C domain are conserved or highly similar (Fig. 3D). Furthermore, computational docking of the PCP<sub>3</sub> domain onto the acceptor-PCP binding site of the C<sub>3</sub> domain showed that it binds in an almost identical orientation to the PCP<sub>2</sub> domain in the structure of the PCP<sub>2</sub>-C<sub>3</sub> didomain (SI Figure S4). This supports the notion that the PCP<sub>2</sub>-C<sub>3</sub> didomain structure is a valid representation of an acceptor-PCP-bound C-domain.

Analysis of the PCP<sub>2</sub>-C<sub>3</sub> didomain structure revealed extra density extending from the conserved Ser (S2514) at the beginning of helix 2 of the PCP domain. This serine residue is the target of phosphopantetheinyl transferases, a class of enzymes that attach the essential PPant moiety to PCP domains. Mass spectrometric analysis of the PCP<sub>2</sub>-C<sub>3</sub> didomain construct revealed a 340 Dalton mass increase, consistent with attachment of PPant to S2514, likely installed by the phosphopantetheinyl transferase EntD that phosphopantetheinylates some PCP domains when they are expressed in <i>E. coli</i>. Indeed, expression of the PCP<sub>2</sub>-C<sub>3</sub> didomain construct in an <i>entD</i> mutant showed no increase in mass, supporting this hypothesis. Having confirmed the presence of a PPant arm, we modeled this into the electron density observed in our structure. Interestingly, we found that this did not extend into the active site of the C domain, but instead curled back towards the outer surface of the C domain (Fig. 4A). The side chain of R2577 appears to block the channel that leads to the active site of the C domain (Fig. 4A). Molecular dynamics simulations initiated from structures of the C<sub>3</sub> domain (with the PCP-PPant removed) highlight the intrinsically dynamic nature of the acceptor substrate channel and the important
role that R2577 has in modulating its shape and size ([SI Figure S5]). This residue forms the bottleneck of the channel and samples alternate rotamers (primarily rotation around chi-3) that, in concert with a displacement of alpha-helix 1, largely determines its size. When we compared our PCP$_2$-C$_3$ didomain structure with published structures of other C domains in complex with a PPant-modified PCP domain, we found residues with shorter side chains at this position (G21 in AB3403$^{10}$ and A18 in ObiF1$^8$), resulting in channels that do not block PPant access. Interestingly, this Arg residue appears largely conserved in $^{\text{L}}$C$_{\text{L}}[1]$ domains (73% harbor an Arg at this position), but is not seen in $^{\text{D}}$C$_{\text{L}}$ domains (Gly (80%) or Ala (4%) are found instead ([SI Figure S6])). Whilst it was unclear what role this residue plays in NRPS function, we hypothesized that it could influence access to acceptor channel of the C domain.

**Effect of R2577G mutation on substrate position.** To verify the role of the R2577 in controlling access to the catalytic channel, we generated the Arg to Gly mutant (R2577G) of the C$_3$ domain. To control the modification state of the PCP$_2$ domain, the mutant PCP$_2$-C$_3$ didomain construct was expressed in the entD mutant of *E. coli.*$^{31}$ After purification, the protein was modified using the promiscuous PPant transferase Sfp R4-4 mutant$^{32}$ and coenzyme A (CoA) (See Methods) to ensure homogeneous PPant loading. Similar to the wild type construct, the protein expressed well and crystallized in the same conditions. Crystals diffraction to 2 Å and the structure was phased using molecular replacement with the previous model ([SI Table S1]). The structure of the R2577G mutant is very similar to that of the wild type protein, with the PCP$_2$ domain sitting at the acceptor site of the C$_3$ domain (RMSD (all atoms) 1.2 Å compared to wild type). The first noticeable difference is a small rotation of the PCP domain in relation to the C domain and slight alterations in the PCP interacting regions of the C domain, likely attributable to the R2577G mutation allowing the first helix of the C domain to sit deeper in the acceptor channel ([SI Figure S7]).$^5$ The major difference, however, is the positioning of the PPant moiety, which now fully extends thought the acceptor channel into the active site (Fig. 4B, [SI Figure S8]) in a similar way to that seen in the ObiF1, SrfA-C and AB3403 structures.$^{8,10,12}$ This observation supports the hypothesis that R2577 acts to control substrate access to the active site of the C domain. One possibility is that this process operates by charge repulsion: when an aminoacyl-PPant approaches the acceptor channel, the ammonium group of the substrate triggers the rotation of the Arg side chain due to charge repulsion, which opens the channel, allowing the aminoacyl-PPant to enter it. This would explain our inability to crystallize the wild type PCP$_2$-C$_3$ construct loaded with PPant derivatives lacking an amino group (such as propionyl and propan-1,3-dioyl,$^{33}$ data not shown), due to interactions that interfere with crystallization when the substrate is not bound in the acceptor channel of the C domain. To further explore this mechanism, we next turned to the characterization of the PCP$_2$-C$_3$ construct with an aminoacyl group appended to the PPant thiol group.

**Structure of the amino acid acceptor bound substrate.** To append the glycyl substrate of module 3 to the PCP$_2$ domain, we attempted to load the apo-PCP$_2$C$_3$ didomain using Sfp and the CoA thioester of glycine. Crystals in the same space group were readily obtained using the same method as for the two previously described structures. Somewhat surprisingly, in this structure it was clear that the electron density
exploring C domain activity and specificity of the PCP2-C3 construct. To test the activity and selectivity of the C domain, as well as the effect of mutating key residues, we first needed to generate an activity
assay for the C domain using the PCP$_2$-C$_3$ construct and downstream PCP$_3$ domain. Given that the interaction between PCP and C domains is weak and transient in nature, we first validated the importance of this restraint in an assay using separately isolated PCP$_2$-C$_3$ (loaded with a synthetic dihydroxybenzoic acid (DHB)-D-Arg-Gly donor substrate) and PCP$_3$-Gly constructs. This experiment revealed no elongation when these constructs were incubated together. Thus, we turned to the use of a fused PCP$_2$-C$_3$-PCP$_3$ construct, albeit one in which the PCP-constructs could be separately loaded with substrates prior to generation of the fused complex (Fig. 6A). To accomplish this, we cloned the donor PCP$_2$-C$_3$ construct with a C-terminal SpyCatcher domain and the acceptor PCP$_3$ with an N-terminal SpyTag peptide. This system allows for the separate loading of the substrates on the PCP domain of each construct using Sfp and synthetic CoA substrates whilst also allowing the reconstitution of the NRPS assembly line.

Using this experimental setup, we confirmed that the condensation reaction was performed as expected, with high levels of conversion of the canonical donor DHB-D-Arg-Gly tripeptide into the Gly-extended tetrapeptide, as determined by high-resolution LC-MS/MS experiments (Fig. 6B). Next, we tested a simplified benzoic acid (BA)-D-Arg-Gly donor substrate in these assays, which showed acceptable levels of conversion (61%) and hence we retained this simplified substrate for all subsequent assays. With a functional condensation assay in hand, we first could verify that the stabilized Gly$_{stab}$ acceptor substrate was a functional mimic of Gly in this C domain (SI Figure S11) using intact protein MS together with PPant ejection (see Methods). With confidence that the Gly$_{stab}$ structure represents a functional acceptor substrate-bound C domain state, we then set out to investigate the effect that mutating key residues had on the condensation activity. Firstly, we confirmed that the R2577G mutant C domain retained activity (with Gly), although this was reduced compared to the wildtype C domain (32%), possibly due to the loss of stabilizing interactions with the PPant arm (Fig. 6C, SI Figure S10). We next generated an active site H2697Q mutant and determined that H2697 is indeed essential for activity with this C domain, as the mutant only retains ~1% of the WT activity with Gly as the acceptor substrate (Fig. 6C).

In addition to Gly and Gly$_{stab}$, we found that C$_3$ could also accept PPant-linked L-Ala and L-Leu as substrates, with 99% and 75% conversion levels, respectively (Fig. 6C). In contrast, PPant-linked L-Phe was a poor substrate, with minimal (6%) levels of conversion. In order to rationalize these differences, we analyzed the structures and performed molecular docking. First, assuming that the position of Gly$_{stab}$ in our Gly$_{stab}$-PCP$_2$-C$_3$ complex represents that catalytically-competent conformation, and that alternate amino acid acceptor substrates must bind in a way that positions the terminal amine group in a similar position, we identified several residues in the central cavity that would likely interact with the side chain of an alternate acceptor substrate. In particular, the side chains of M2917, S2919, Q2921, P2941 and E2950 could contribute a putative side chain binding pocket for this C-domain, in a manner reminiscent of A-domains (Fig. 5B). Molecular docking revealed that sidechains of L-Ala and L-Leu could be accommodated by the active site cavity's side-chain binding pocket and had top scoring poses that positioned the terminal amine towards the catalytic residues (although the L-Leu pose was slightly strained, SI Figure S12). In contrast, the bulky side chain of L-Phe could only be accommodated within
the central cavity in poses that positioned the terminal amino acid amine away from the catalytic histidine and that would not be compatible with catalysis (SI Figure S12). Analysis of the putative pocket residues (M2917, S2919, Q2921, P2941 and E2950) compared to the reported activity of the downstream A-domain did not reveal any correlation between acceptor substrate and these possible “pocket” residues (Spearman’s rho: -0.05), indicating the lack of a C-domain side chain binding pocket and hence “C domain code” comparable to those found with A domains (SI Figure S13). Our results do however indicate that alterations in the C domain active site can lead to changes in selectivity, and hence we turned to further analysis of the residues within the active site motif.

Although most C domains contain a canonical HHxxxDG motif, the C3 domain from fuscachelin NRPS features an unusual HHxxxD_E variant. We hypothesized that, in absence of a side chain in the acceptor substrate (Gly) to position the acceptor substrate, the role of this glutamate (E2702) could be to stabilize and orient the acceptor substrate amine group to ensure an efficient nucleophilic attack of the donor substrate thioester. Indeed, an analysis of C domains where their acceptor substrates are known demonstrated that there is a higher proportion of modified motifs where the acceptor substrate is small as opposed to traditional HHxxxDG containing C domains (SI Figure S14). To test this hypothesis, we mutated this glutamate to its canonical glycine residue (E2702G) and performed condensation reactions with Gly as the acceptor substrate. As expected, the condensation level with Gly as the acceptor substrate was reduced by almost half (61–35%) when compared to the WT, demonstrating the non-essential, although beneficial role of this glutamate residue. Interestingly, while the E2702G mutation had reduced activity with the Gly acceptor substrate, this substitution improved the activity for PPant-linked L-Leu from 75–92% (Fig. 6C). This result indicates that the E2702 residue can play a particularly important role in supporting condensation reactions involving Gly as an acceptor substrate, but may be detrimental for other acceptor substrates. Molecular docking of the Gly-PPant and Gly_stab-PPant into a model of the E2702G C3 mutant reveals how the removal of the glutamic acid results in substrate poses that are unlikely to be compatible with catalysis, with the terminal amine of the substrates instead interacting with Glu2950 (SI Figure S15).

[1] Superscript indicates the stereochemistry of the C-terminal residue of the donor substrate, subscript indicates the stereochemistry of the acceptor substrate

Discussion

Non-ribosomal peptide synthetases are widely recognized for their impressive selectivity in assembling specific peptide products. While the role of the A domain in substrate selection is clear, the possible role of C domains as a second selectivity filter during peptide assembly has been less well defined. Early studies suggested C domains may show selectivity towards their acceptor substrates, but more recent work has questioned this.25

The structural characterization and bioinformatics analysis we have performed of PCP-bound acceptor complexes in this work shows no general correlation between the size or chemical nature of the acceptor
amino acid side chain and potential side chain binding residues in the C domain. Whilst C domain selectivity has recently been characterized in glycopeptide antibiotic biosynthesis,\textsuperscript{21} there the mechanism rather acts to ensure that important modifications of the PCP-bound aminoacyl thioester are performed prior to condensation. Whilst some selectivity for the amino acid substrate is seen here, for example in the low conversion (albeit still present) of L-Phe, this appears likely to be due to the significant difference between the small, flexible Gly substrate and L-Phe, with its large, rigid side chain. The influence of the atypical HHxxxDE motif of this C-domain is also seen on lower levels of acceptance of larger amino acids (such as Leu), which can be released upon conversion of the motif into the typical HHxxxDG sequence. This demonstrates the versatile nature of C domains for tolerating active site modifications, some of which can play important additional roles in supporting catalysis.\textsuperscript{24}

Within the active site, the amino group of the aminoacyl acceptor lies close to the central histidine residue, with calculations suggesting that this residue could indeed act as a base to deprotonate the zwitterionic intermediate. Further characterization of the PCP-C complex shows that the PCP binding site of the C domain is, as anticipated, dominated by hydrophobic interactions and is one that is relatively flexible with regards to the PCP domain.\textsuperscript{26} Access of the PPant arm to the C domain active site appears to be gated by R2577, which repels the unmodified PPant arm (or neutral/negatively charged substrates) in favor of the aminoacyl-PPant. Whilst this residue is largely conserved in \textsuperscript{1}L\textsubscript{C}L domains, it is typically Gly or other small residues in \textsuperscript{0}C\textsubscript{L} domains, which we have confirmed allows the unmodified PPant into the C-domain active site. One hypothesis for the role of this residue would be to prevent the unwanted “pass-through” of donor substrates without elongation (e.g. from PCP\textsubscript{2} to PCP\textsubscript{3}). Examples of NRPS-dependent pathways in which CP-bound substrate transfer could occur reveals that the C domains implicated bear the Arg to (Gly/small) amino acid mutation (e.g. burkholdac biosynthesis),\textsuperscript{38} which provides some support for this hypothesis. For \textsuperscript{0}C\textsubscript{L} domains, mutation of this Arg residue could be a requirement due to the need for E domain-catalyzed inversion of stereochemistry prior to chain elongation, as we note that the Arg to (Gly/small) mutation generally appears to be somewhat deleterious to peptide conversion levels, possibly due to a lack of interactions between the Arg and PPant arm in these C-domains. We anticipate that the structural snapshots presented here will pave the way for studies to probe the roles of this Arg residue as well as other active site residues in C domain catalysis, which is important due to the ever-increasing roles of C-type domains in non-ribosomal peptide biosynthesis.

**Declarations**

**Competing Interests**

G.L.C. is a co-director of Erebagen Ltd. The other authors declare no competing interests.

**Author Contributions**
The study was designed by T.I and M.J.C. All cloning and protein purification was performed by T.I. and Y.T.C.H. Structural analysis was performed by T.I., Y.T.C.H. and M.J.C. with insightful contributions from G.L.C. and J.A.K. Chemical synthesis was performed by Y.T.C.H. and turnovers results were analyzed by Y.T.C.H. and D.L.S., with M.T. assisting with analysis of HRMS experiments. HRMS and protein MS measurements were performed by D.L.S., R.J.A.G. and R.B.S. Molecular docking and molecular dynamics simulations were performed and analyzed by J.A.K. and C.J.J. Bioinformatics and correlation analyses were performed by A.G. and N.Z. Computational analysis was performed by K.H.C. and E.H.K. The manuscript was written by T.I., Y.T.C.H and M.J.C., with input from the other authors.

Acknowledgements

J. Yin (University of Chicago) for the R4-4 Sfp expression plasmid; M. Kaniusaite (Monash) for assistance with cloning; T. Harshegyi, L. Scully and S. Stamatis (Monash) for assistance with protein purification; D. Maksel & G. Kong (MMCF, Monash) for assistance with crystal screening experiments. This research was undertaken on the MX1 and MX2 beamlines at the Australian Synchrotron, part of ANSTO, and made use of the Australian Cancer Research Foundation (ACRF) detector. We would like to thank the beamline scientists at the Australian Synchrotron for their support during data collection. Computational resources were provided by the National Facility of the Australian National Computational Infrastructure through the National Computational Merit Allocation Scheme and by the University of Queensland Research Computing Centre. This work was supported by Monash University, EMBL Australia, the Australian Research Council (Discovery Project DP190101272 and DP180103047) and the National Health and Medical Research Council (APP1140619 to M.J.C.). T.I. is grateful for the support of the CASS foundation (grant #8583). A.G. is grateful for the support of the Deutsche Forschungsgemeinschaft (DFG; Project ID # 398967434-TRR 261). This research was conducted by the Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and funded by the Australian Government.

References

1. Süßmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. Angew Chem Int Ed 56, 3770–3821, (2017).

2. Walsh, C. T., O’Brien, R. V. & Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. Angew. Chem. Int. Ed. 52, 7098–7124, (2013).

3. Bloudoff, K. & Schmeing, T. M. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. Biochim. Biophys. Acta 1865, 1587–1604, (2017).

4. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. Nat. Struct.
5. Tan, K. et al. Structures of teixobactin-producing nonribosomal peptide synthetase condensation and adenylation domains. *Curr. Res. Struct. Biol.* 2, 14–24, (2020).

6. Wang, L., Yuan, M. & Zheng, J. Crystal structure of the condensation domain from lovastatin polyketide synthase. *Synthetic and Systems Biotechnology* 4, 10–15, (2019).

7. Reimer, J. M. et al. Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* 366, eaaw4388, (2019).

8. Kreitler, D. F., Gemmell, E. M., Schaffer, J. E., Wencewicz, T. A. & Gulick, A. M. The structural basis of N-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nature Commun.* 10, 3432, (2019).

9. Kosol, S. et al. Structural basis for chain release from the enacyloxin polyketide synthase. *Nature Chem.* 11, 913–923, (2019).

10. Drake, E. J. et al. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* 529, 235–238, (2016).

11. Bloudoff, K., Alonzo, Diego A. & Schmeing, T. M. Chemical Probes Allow Structural Insight into the Condensation Reaction of Nonribosomal Peptide Synthetases. *Cell Chem. Biol.* 23, 331–339, (2016).

12. Tanovic, A., Samel, S. A., Essen, L.-O. & Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science* 321, 659–663, (2008).

13. Tarry, M. J., Haque, A. S., Bui, K. H. & Schmeing, T. M. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* 25, 783–793.e784, (2017).

14. Zhang, J. et al. Structural basis of nonribosomal peptide macrocyclization in fungi. *Nature Chem. Biol.* 12, 1001–1003, (2016).

15. Bloudoff, K., Rodionov, D. & Schmeing, T. M. Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases. *J. Mol. Biol.* 425, 3137–3150, (2013).

16. 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).

17. Bozhüyük, K. A. J. et al. Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nature Chem.* 11, 653–661, (2019).

18. Niquille, D. L. et al. Nonribosomal biosynthesis of backbone-modified peptides. *Nature Chem.* 10, 282, (2017).

19. Kaniusaite, M., Goode, R. J. A., Tailhades, J., Schittenhelm, R. B. & Cryle, M. J. Exploring modular reengineering strategies to redesign the teicoplanin non-ribosomal peptide synthetase. *Chem. Sci.* 11, 9443–9458, (2020).
20. Reitz, Z. L., Hardy, C. D., Suk, J., Bouvet, J. & Butler, A. Genomic analysis of siderophore β-hydroxylases reveals divergent stereocontrol and expands the condensation domain family. *Proc. Natl. Acad. Sci. USA*, 201903161, (2019).

21. Kaniusaitė, M. *et al.* A proof-reading mechanism for non-proteinogenic amino acid incorporation into glycopeptide antibiotics. *Chem. Sci.* **10**, 9466–9482, (2019).

22. Patteson, J. B., Dunn, Z. D. & Li, B. In Vitro Biosynthesis of the Nonproteinogenic Amino Acid Methoxyvinylglycine. *Angew. Chem. Int. Ed.* **57**, 6780–6785, (2018).

23. Haslinger, K., Peschke, M., Brieke, C., Maximowitsch, E. & Cryle, M. J. X-domain of peptide synthetases recruits oxygenases crucial for glycopeptide biosynthesis. *Nature* **521**, 105–109, (2015).

24. Gaudelli, N. M., Long, D. H. & Townsend, C. A. beta-Lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* **520**, 383–387, (2015).

25. Calcott, M. J., Owen, J. G. & Ackerley, D. F. Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nature Commun.* **11**, 4554, (2020).

26. Izoré, T. & Cryle, M. J. The many faces and important roles of protein–protein interactions during non-ribosomal peptide synthesis. *Nat. Prod. Rep.* **35**, 1120–1139, (2018).

27. Dehling, E., Rüschenbaum, J., Diecker, J., Dörner, W. & Mootz, H. D. Photo-crosslink analysis in nonribosomal peptide synthetases reveals aberrant gel migration of branched crosslink isomers and spatial proximity between non-neighboring domains. *Chem. Sci.*, (2020).

28. Alfermann, J. *et al.* FRET monitoring of a nonribosomal peptide synthetase. *Nature Chem. Biol.* **13**, 1009–1015, (2017).

29. Reimer, J. M., Aloise, M. N., Harrison, P. M. & Schmeing, T. M. Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239–242, (2016).

30. Dimise, E. J., Widbooom, P. F. & Bruner, S. D. Structure elucidation and biosynthesis of fuscachelins, peptide siderophores from the moderate thermophile *Thermobifida fusca*. *Proc. Natl. Acad. Sci. USA* **105**, 15311–15316, (2008).

31. Owen, J. G., Robins, K. J., Parachin, N. S. & Ackerley, D. F. A functional screen for recovery of 4′-phosphopantetheinyl transferase and associated natural product biosynthesis genes from metagenome libraries. *Environ. Microbiol.* **14**, 1198–1209, (2012).

32. Sunbul, M., Marshall, N. J., Zou, Y., Zhang, K. & Yin, J. Catalytic Turnover-Based Phage Selection for Engineering the Substrate Specificity of Sfp Phosphopantetheinyl Transferase. *J. Mol. Biol.* **387**, 883–898, (2009).

33. Izoré, T. *et al.* *Drosophila melanogaster* nonribosomal peptide synthetase Ebony encodes an atypical condensation domain. *Proc. Natl. Acad. Sci. USA* **116**, 2913–2918, (2019).

34. Sztain, T. *et al.* Modifying the Thioester Linkage Affects the Structure of the Acyl Carrier Protein. *Angew. Chem. Int. Ed.* **58**, 10888–10892, (2019).

35. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci. USA* **109**, E690-E697, (2012).
36. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. Aminoacyl-CoAs as Probes of Condensation Domain Selectivity in Nonribosomal Peptide Synthesis. *Science* **284**, 486–489, (1999).

37. Bergendahl, V., Linne, U. & Marahiel, M. A. Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.* **269**, 620–629, (2002).

38. Ho, Y. T. C. *et al.* Novel chemical probes for the investigation of nonribosomal peptide assembly. *Chem. Commun.* **53**, 7088–7091, (2017).

39. Biggins, J. B., Gleber, C. D. & Brady, S. F. Acyldepsipeptide HDAC Inhibitor Production Induced in Burkholderia thailandensis. *Org. Lett.* **13**, 1536–1539, (2011).

**Figures**

**Figure 4**

PCP2-C3 interaction surfaces showing the differences in substrate acceptance. a) Structure of WT C3 domain with unloaded PPant, showing the PPant not extending into the C-domain as the side chain of R2577 prevents the PPant accessing the C-domain active site. b) Structure of R2577G C3 domain with an unloaded PPant, showing the PPant fully extended into the C-domain catalytic channel. c) Structure of WT C3 domain where the PPant is loaded with a Glystab substrate, rotated 90 degree anticlockwise compared to panels (a) and (b). Here, PPant-Glystab extends fully into the catalytic channel. d) Cutaway representation of the C domain indicating the path of the Glystab substrate from the PCP domain (shown in orange). All densities shown as 2Fo-Fc maps.
Figure 5

The C3 domain catalytic site showing the docked PPant-Glystab. a) PPant-Glystab substrate extends fully into the active site, approaching the active site HHxoxDE motif (H2696 to E2707); electron density shown as a 2Fo-Fc map. b) The Glystab substrates is stabilized by a network of hydrophilic interactions. Note that residues M2917, S2919, Q2921, P2941 and E2950 are in a position that could potentially interact with the side chain of alternate acceptor substrates. c) Mechanism of peptide bond formation via concerted N–C bond formation and N-deprotonation (upper) or sequential N–C bond formation and N-deprotonation (lower). d) Zwitterionic intermediate in the sequential N–C bond formation/N-deprotonation pathway, in which the oxyanion is stabilized by two water molecules and the ammonium ion forms a hydrogen bond to histidine.

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.

- 7KVWvalidationreport.pdf
- 7KW2validationreport.pdf
- 7KW3validationreport.pdf
- SupplementaryInformationCryle2020.pdf
- ReportingSummaryCryle.pdf