JBC RESEARCH ARTICLE

Structural characterization of a PCP–R didomain from an archaeal nonribosomal peptide synthetase reveals novel interdomain interactions

Received for publication, October 6, 2020, and in revised form, February 13, 2021 Published, Papers in Press, February 18, 2021,
https://doi.org/10.1016/j.jbc.2021.100432

Sandesh Deshpande[1], **Eric Altermann**[2,3], **Vijayalekshmi Sarojini**[4], **J. Shaun Lott**[1], **and T. Verne Lee**[1,*]

From the [1]*School of Biological Sciences, University of Auckland, Auckland, New Zealand;* [2]*AgResearch Limited, Food System Integrity, Palmerston North, New Zealand;* [3]*Riddet Institute, Massey University, Palmerston North, New Zealand; and* [4]*School of Chemical Sciences, University of Auckland, Auckland, New Zealand*

Edited by Joseph Jez

Nonribosomal peptide synthetases (NRPSs) are multimodular enzymes that produce a wide range of bioactive peptides, such as siderophores, toxins, and antibacterial and insecticidal agents. NRPSs are dynamic proteins characterized by extensive interdomain communications as a consequence of their assembly-line mode of synthesis. Hence, crystal structures of multidomain fragments of NRPSs have aided in elucidating crucial interdomain interactions that occur during different steps of the NRPS catalytic cycle. One crucial yet unexplored interaction is that between the reductase (R) domain and the peptide carrier protein (PCP) domain. R domains are members of the short-chain dehydrogenase/reductase family and function as termination domains that catalyze the reductive release of the final peptide product from the terminal PCP domain of the NRPS. Here, we report the crystal structure of an archaeal NRPS PCP–R didomain construct. This is the first NRPS R domain structure to be determined together with the upstream PCP domain and is also the first structure of an archaeal NRPS to be reported. The structure reveals that a novel helix–turn–helix motif, found in NRPS R domains but not in other short-chain dehydrogenase/reductase family members, plays a major role in the interface between the PCP and R domains. The information derived from the described PCP–R interface will aid in gaining further mechanistic insights into the peptide termination reaction catalyzed by the R domain and may have implications in engineering NRPSs to synthesize novel peptide products.

Nonribosomal peptide synthetases (NRPSs) are multimodular enzymes that function as molecular assembly lines to synthesize a wide range of structurally and functionally diverse peptides (1). Nonribosomal peptides include a large array of natural products, including siderophores, toxins, and antibacterial and insecticidal agents (2). A large number of nonribosomal peptides have been used as antimicrobial, antifungal, antitumor, or immunosuppressant drugs (2). Nonribosomal peptides are produced predominantly by bacteria and fungi and are rare in archaea (2, 3); only three NRPS gene clusters belonging to classes *Methanomicrobia* and *Methanobacteria* were detected in the 128 archaeal genomes analyzed by Wang and co-workers (3).

A minimal NRPS module consists of an adenylation (A) domain, a peptide carrier protein (PCP) domain, and a condensation (C) domain (1). In addition, the final module possesses a termination domain at its C-terminal end that catalyzes release of the peptide product. Reductase (R) domains are one of the two main types of termination domains employed by NRPSs, the other one being thioesterase (TE) domains. The growing peptide chain is shuttled across the NRPS while covalently tethered to a 4′-phosphopantetheine (Ppant) arm that is post-translationally added to a conserved serine residue in the PCP domain. The termination domain releases the completed peptide product from its covalent tether to the terminal PCP domain (2, 4). Termination reactions catalyzed by TE domains result in offloading the peptide product by hydrolysis of the thioester bond between the peptide and Ppant. R domains, on the other hand, catalyze NAD(P)H-dependent $2e^-$ or $4e^-$ reductive release of the peptides (5) (Fig. S1*A*). The products released by R domains can be alcohols (6), linear aldehydes (7), or macrocyclic products that occur when the reactive aldehyde group is attacked by internal amines in the peptide chain to give rise to cyclic structures, as seen in case of pyrazinone products (8) (Fig. S1*B*).

Crystal structures of stand-alone NRPS A domains (9, 10), C domains (11), PCP domains (12), and TE domains (13) have aided in understanding the individual reactions catalyzed by these domains. Crystal structures of multidomains (14, 15), modules (16–18), and crossmodules (17, 19) have been crucial in elucidating the functionally significant domain movements resulting from several intradomain and interdomain interactions during different steps of the NRPS synthetic cycle. These structures have highlighted the importance of interdomain interfaces in NRPS enzymology. Mutation of conserved residues in the interfaces has resulted in the loss or reduction of catalytic activity (14, 18, 20). The A–PCP interface described by Sundlov and co-workers (21, 22) and the PCP–TE interface described by Liu and co-workers (14)

* For correspondence: T. Verne Lee, t.lee@auckland.ac.nz.

© 2021 THE AUTHORS. Published by Elsevier Inc on behalf of American Society for Biochemistry and Molecular Biology. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

revealed a common region of the PCP domain participating in all interactions with its upstream and downstream partner domains. Furthermore, understanding NRPS domain interactions has led to advancements in NRPS enzyme engineering to synthesize novel peptides *in vitro* (23, 24).

R domains are members of the NAD(P)H-dependent short-chain dehydrogenase/reductase (SDR) superfamily (25, 26). Members of this superfamily are widespread in all kingdoms of life, with archaea having the lowest number of characterized SDRs (25). Although the overall sequence identity of proteins in this superfamily is low (20–30%), all their tertiary structures possess the conserved and cofactor-binding Rossmann fold (26). NRPS R domains are further subclassified as "extended" SDRs, in contrast to "classical" SDRs, as they possess an additional C-terminal subdomain of about 100 residues (26). In the last decade, several R domains have been structurally and functionally characterized (6, 8, 27). Structural studies on the R domain (Mtb-R) of a *Mycobacterium tuberculosis* NRPS that produces an unknown peptide (Protein Data Bank [PDB] ID: 4DQV and 4U5Q) have provided insights into the intra-domain interactions that regulate NADPH binding and the subsequent two-step ($4e^-$) reduction of the product to an alcohol *via* an aldehyde intermediate (27). The terminal R domain (MxaA-R) of the NRPS module of the myxalamid biosynthetic pathway (PDB ID: 4U7W) from *Stigmatella aurantiaca* Sga15 is the only "NADPH-bound" NRPS R domain structure to be reported to date (6). The structure of the R domain (AusA-R) of a dimodular NRPS (PDB ID: 4F6C) from the aureusimine biosynthetic cluster in *Staphylococcus aureus* strain Mu50 is the only aldehyde-producing NRPS R domain structure to be reported to date (8). In the same article, Wyatt and co-workers (8) also reported the structure of a protein construct (AusA-PCP–R) containing both the upstream PCP and the R domain (PDB ID: 4F6L). However, electron density was visible only for the R domain, likely because of the positional disorder of the PCP domain.

One crucial yet underexplored mechanism during peptide termination by NRPS R domains is the interaction of the R domain with the upstream PCP domain. Some insights have been provided into the interdomain interactions between the PCP and R domains in the recently reported didomain structures (CAR–PCP–R) of carboxylic acid reductases (CARs) from *Nocardia iowensis* and *Segniliparus rugosus* (PDB ID: 5MSV, 5MSP, and 5MSR) (20). CARs are made up of an A domain, a PCP domain, and an R domain, which is analogous to the architecture of an NRPS termination module. These enzymes accept carboxylic acids as their substrates and catalyze their NADPH-assisted reduction to aldehydes, which is a similar activity to the R domain–containing NRPS enzymes.

In this study, we report the crystal structures of the Ppant-modified and unmodified PCP–R didomain (351-PCP–R) of the terminal module of the NRPS Mru_0351 from *Methanobrevibacter ruminantium* M1, one of the prevalent species of methanogenic archaea found in ruminants (28). Mru_0351 is one of the two NRPSs discovered in *M. ruminantium* (29), the products of which are still unknown. Therefore, identification and functional characterization of the peptide product of Mru_0351 will further our understanding of methanogen biology and will have implications in developing novel methane mitigation strategies. As a step toward understanding the peptide production mechanism, we have determined the structure of the Mru_0351 PCP–R didomain that is involved in release of the mature peptide product. This is the first structure of an NRPS R domain to be determined in association with its preceding PCP domain and also the first structure of an archaeal NRPS to be reported. In contrast to the CAR–PCP–R structure, the Ppant-modified 351-PCP–R adopts a previously unobserved conformation, thereby providing further insights into the PCP–R interdomain interactions. Furthermore, we have identified a helix–turn–helix (HTH) motif, found in NRPS R domains but not in other SDR family members, that mediates the interface between the PCP and R domains.

Results

The overall structure of 351-PCP–R

The unmodified and Ppant-modified structures of 351-PCP–R were determined to resolutions of 2.65 and 1.95 Å, respectively. The atomic coordinates and structure factors have been deposited in the PDB with codes 6VTZ (unmodified) and 6VTJ (Ppant modified). Structural alignment of the unmodified and Ppant-modified structures using secondary structure matching (30) resulted in an overall RMSD value of 0.45 Å, showing that the modification did not induce any significant change in the overall protein structure. The 351-PCP–R domain can be structurally divided into a cofactor-binding N-terminal subdomain and a substrate-binding C-terminal subdomain (Fig. 1), as is the case for other previously determined NRPS R domain structures. No electron density was visible for the PCP domain (residues V3701–S3778) in the unmodified structure, likely because of positional disorder, similar to the disordered PCP domain that has been reported in the crystal structure of the AusA-PCP–R (8). However, electron density was visible for the PCP domain in our Ppant-modified structure, showing that the PCP domain makes contact both with the C-terminal R subdomain and with a key loop in the NADPH-binding pocket of the N-terminal subdomain, previously defined as the "gating" loop by Kinatukara and co-workers (31) (Fig. 1*B*). The PCP domain itself is a four-helix bundle, like all previously determined PCP domain structures (Fig. 1*B*). Eight residues in the Ppant-modified structure belonging to the PCP domain–R domain linker were not modeled because of weak electron density. Ppant can be seen covalently attached to the conserved serine residue in the active site of the PCP domain (Fig. 1), but our model is pruned to atom N41 because of weak electron density beyond this atom (Fig. S2).

The N-terminal R subdomain harbors the conserved and cofactor-binding Rossmann fold consisting of a β-sheet made of seven parallel β-strands (in the order 3-2-1-4-5-6-7) flanked by α-helices on either side (Fig. 2*A*). This fold is present in all members of the SDR superfamily. The Rossmann fold contains a glycine-rich conserved motif: T–G–A–T–G–F–L–G

ASBMB

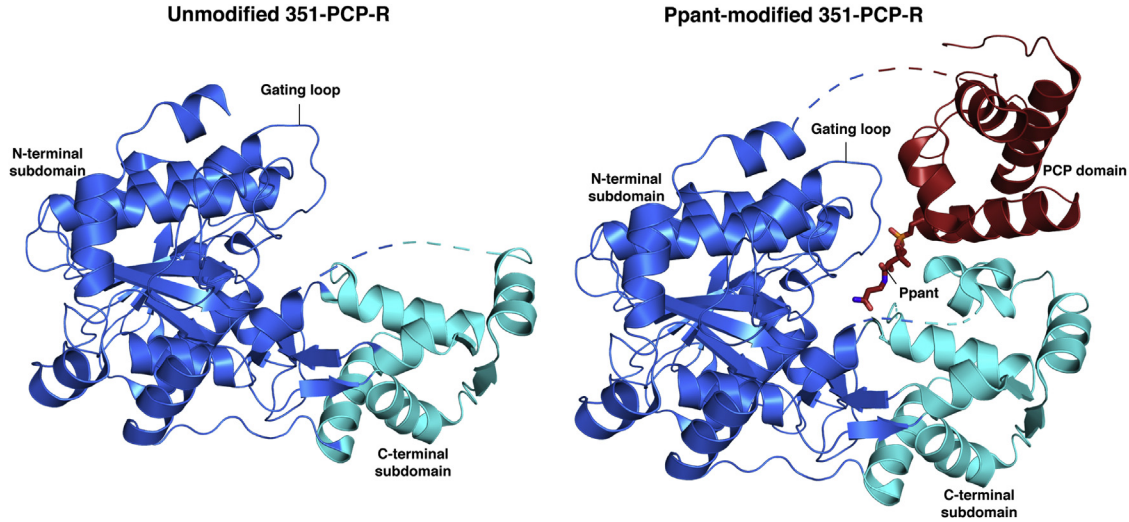

Figure 1. Ribbon diagram representations of the 351-PCP–R structures. *A*, the unmodified 351-PCP–R structures. *B*, the Ppant-modified 351-PCP–R structures (N-terminal reductase subdomain—*blue*, C-terminal R subdomain—*cyan*, and PCP domain—*red*). PCP–R, peptide carrier protein–reductase; Ppant, 4′-phosphopantetheine.

(residues T3828 to G3835 in 351-PCP–R), which is important in maintaining the structure of the central β-sheet and in cofactor binding (25). Another characteristic feature of the SDR superfamily is the presence of the triad of catalytic residues Y–T/S–K, which are present at positions Y3988, S3991, and K3992 in 351-PCP–R (Fig. 2*A*). The N-terminal residues of the subdomain (L3789–N3824 in 351-PCP–R), which are immediately downstream of the disordered PCP–R linker, wrap around the core of the R domain, tracing a path similar to that seen in AusA-R and AusA-PCP–R (8) but different to that seen in other R domain structures (6, 20, 27) (Fig. 2*B*).

The C-terminal R subdomain comprises six α-helices and two β-strands and harbors the putative substrate-binding pocket (6, 8, 27). However, the details of the substrate-binding site of NRPS R domains remain unknown, as there are no reported structures determined in complex with their substrates. The C-terminal subdomain shows more variability in tertiary structure than the N-terminal subdomain, as seen from the higher RMSD values in comparisons with other R domain structures (Table S1). The C-terminal subdomain is also typically less conserved in sequence than the N-terminal subdomain, when compared with other R domains (Table S1).

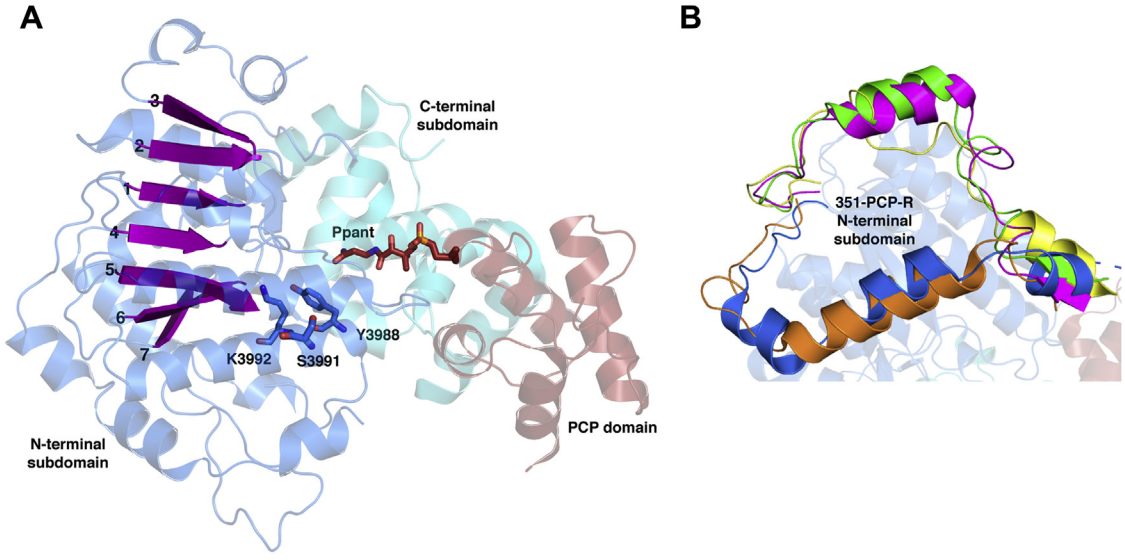

Figure 2. Details of the 351 reductase domain structure. *A*, features of the 351-PCP–R N-terminal subdomain (N-terminal reductase [R] subdomain—*blue*, C-terminal R subdomain—*cyan*, and PCP domain—*red*). The β-sheet in the cofactor-binding Rossmann fold is shown in *magenta*, active-site residues are shown as *blue* sticks, and Ppant is shown as *red* sticks. *B*, superimposed N-terminal regions (L3789–N3824 in 351-PCP–R) of various nonribosomal peptide synthetase and carboxylic acid reductase structures (351-PCP–R—*blue*, AusA-R—*orange*, carboxylic acid reductase PCP–R—*magenta*, MxaA-R—*yellow*, and Mtb-R—*green*). AusA-R, R domain of a dimodular nonribosomal peptide synthetase (NRPS) from the aureusimine biosynthetic cluster; MxaA-R, R domain of the NRPS module of the myxalamid biosynthetic pathway; PCP–R, peptide carrier protein–reductase; Ppant, 4′-phosphopantetheine; Mtb-R, R domain of a *Mycobacterium tuberculosis* NRPS.

Both the unmodified and Ppant-modified 351-PCP–R proteins were crystallized in the absence of the NADPH cofactor. Despite significant efforts, crystals of 351-PCP–R could not be obtained by cocrystallization with NADPH. Examination of the NADPH-binding pocket revealed the presence of small patches of positive difference density in the Ppant-modified 351-PCP–R structure but not in the unmodified structure (Fig. 3*A*). This could be due to partial occupancy from NADPH or could be unidentified small molecules that have occupied the pocket during crystallization or protein purification, although the density did not match any molecules known to be present during crystallization or purification. Previous investigations of the gating loop in Mtb-R structures have proposed that the gating loop is flexible and able to occupy the NADPH-binding pocket resulting in an "inactive" R domain conformation (27, 31, 32) (Fig. 3*A*). Interestingly, in both the unmodified and Ppant-modified 351-PCP–R structures, the gating loop adopts an orientation similar to that of the NADPH-bound MxaA-R structure (6) (Fig. 3*B*).

The 351-PCP–R domain interface

The C-terminal R subdomain and the gating loop from the N-terminal subdomain form an interface with the PCP domain. The region in the C-terminal R subdomain participating in this interface is an HTH motif, which includes residues D4110 to F4139 (Fig. 4*A*). It has previously been noted that this motif is unique to NRPS R domains (6), but its function was unknown. In this study, we have conducted a sequence alignment of a number of extended SDRs, confirming that this motif is an insertion that is present in all NRPS R domains but rarely found in other extended SDRs (Figs. S3 and S4). Thus, we denote this motif as the interface–HTH motif. Part of the interface–HTH motif, M4122 to F4139, is not modeled in the unmodified 351-PCP–R structure because of weak electron density and is likely disordered, whereas it is present in the Ppant-modified structure and forms part of the interface with the PCP domain and Ppant (Fig. 4*A*). Superposition of the interface–HTH motifs of all R domain structures reported to date displays variability in the length and orientation of the helices (Fig. 4*B*).

The region of the PCP domain participating in the PCP–R interface comprises α-helices 2 and 3 and the loop connecting them (Fig. 4*A*). This region also interacts with the partner A, C, and terminal TE domains during different steps of the catalytic cycle in other previously reported NRPS structures (14–16, 33). Our study therefore adds another interaction to this list, underscoring both the relevance of this crystallographic snapshot and the importance of this interface in the domain movements that guide the NRPS catalytic cycle.

The 351-PCP–R interface is distinct from that seen in the only other reported PCP–R structure, which is that of the CAR–PCP–R (20). The PCP domain in the CAR–PCP–R structure interacts exclusively with the N-terminal R subdomain, and no contacts can be seen with the interface–HTH motif (20) (Fig. 5*A*). The region of the CAR PCP domain participating in the interface is made up of parts of α-helix 2, α-helix 3, and the loops connecting α-helices 2, 3, and 4. This is in contrast to that of the NRPS PCP–partner domain interfaces, which involve α-helices 2 and 3 and the loop connecting them. The position of the conserved serine residue and Ppant is similar in both structures (Fig. 5*B*). Ppant is only modeled to atom N41 in the 351-PCP–R structure, whereas the complete Ppant molecule is modeled in the CAR–PCP–R

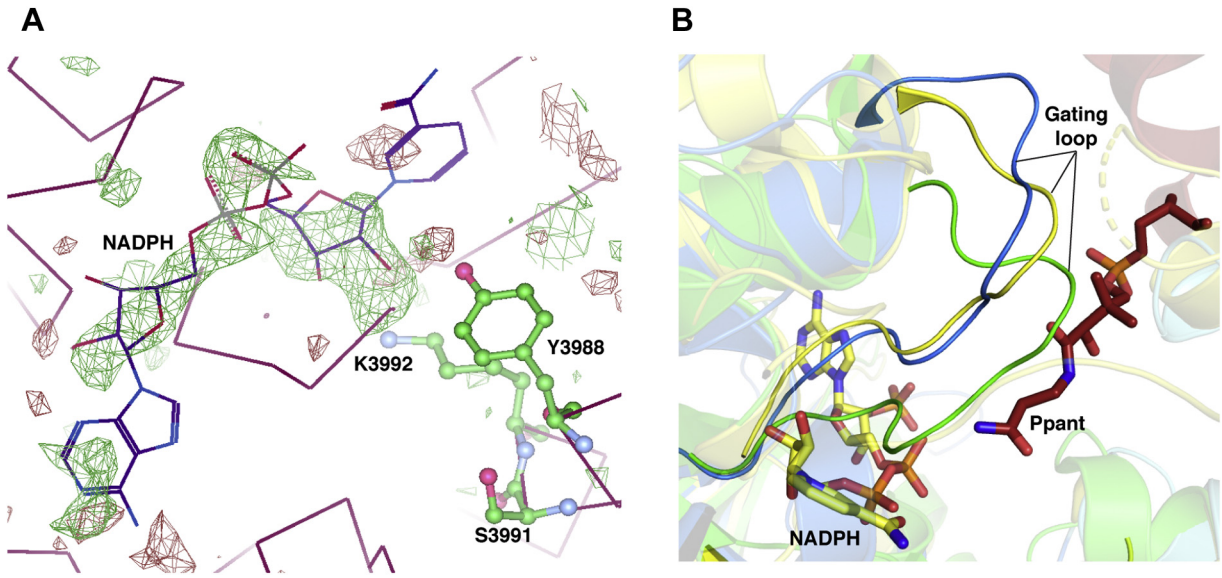

Figure 3. Details of the 351 reductase domain NADPH-binding site. *A*, Ppant-modified 351-PCP–R showing patches of positive difference density in the NADPH-binding site as *green mesh*. The active-site resides are shown in ball and stick representation. NADPH superimposed from structure of MxaA-R (6) is shown. The F_o–F_c map is contoured at 2.5 σ. *B*, comparison of gating loop orientations (351-PCP–R—*blue*, MxaA-R—*yellow* (6), and Mtb-R—*green* (27)). NADPH from MxaA-R and Ppant from 351-PCP–R are shown in stick representation. MxaA-R, R domain of the nonribosomal peptide synthetase module of the myxalamid biosynthetic pathway; PCP–R, peptide carrier protein–reductase; Ppant, 4′-phosphopantetheine; Mtb-R, R domain of a *Mycobacterium tuberculosis* nonribosomal peptide synthetase.

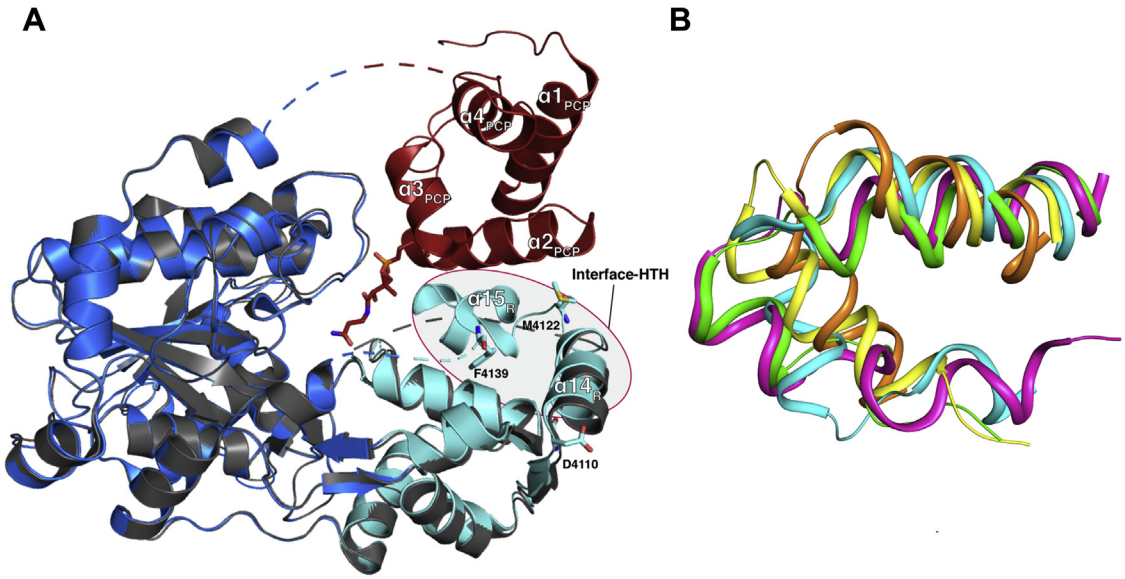

Figure 4. The peptide carrier protein–reductase interface is mediated by an interface of helix–turn–helix motif. *A,* structural alignment of unmodified (*gray*) and Ppant-modified 351-PCP–R (*blue*, *cyan*, and *red*). The interface–HTH motif is highlighted. *B,* structural alignment of the interface–HTH motif of all nonribosomal peptide synthetase reductase domain structures (351-PCP–R—*cyan*, AusA-R—*orange* (8), MxaA-R—*yellow* (6), Mtb-R—*green* (27), and carboxylic acid reductase–PCP–R—*magenta* (20)). AusA-R, R domain of a dimodular nonribosomal peptide synthetase from the aureusimine biosynthetic cluster; HTH, helix–turn–helix; Mtb-R, R domain of a *Mycobacterium tuberculosis* nonribosomal peptide synthetase; MxaA-R, R domain of the nonribosomal peptide synthetase module of the myxalamid biosynthetic pathway; PCP–R, peptide carrier protein–reductase; Ppant, 4′-phosphopantetheine.

structure, in which it extends to the active site with its thiol group positioned in close proximity with NADPH and the catalytic triad. The similar position of Ppant in both structures implies that the thiol group of the 351-PCP–R Ppant would also be positioned similarly in the active site.

The 351-PCP–R interface is characterized by a series of hydrogen bonds and hydrophobic interactions between residues of the PCP and R domains. Residues involved in hydrophobic interactions include F3736, L3743, I3749, I3750, L3753, Y3761, and F3765 from the PCP domain; Y4118, M4122, I4126, and I4130 from the interface–HTH motif; and Y3920 from the gating loop (Fig. 6). Most of these hydrophobic residues are conserved in all PCP and R domains, thereby corroborating the significance of this interface in the

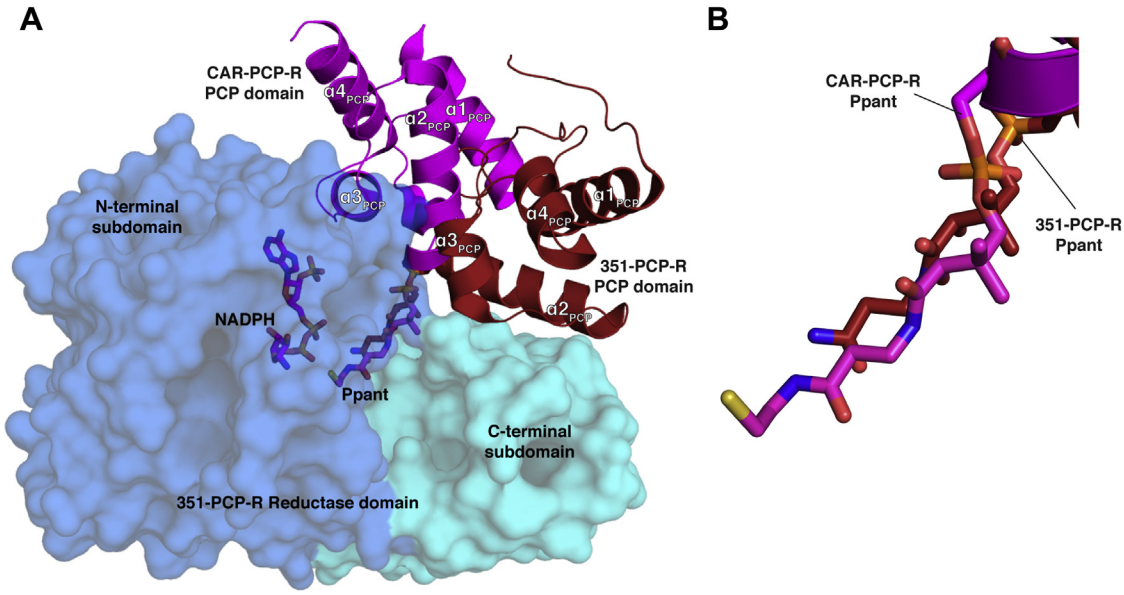

Figure 5. The Ppant-modified 351-PCP-R structure reveals a previously unobserved conformation. *A,* the position of PCP domain of 351-PCP–R compared with the PCP domain of CAR–PCP–R (20) following structural alignment of the reductase domains (351-PCP–R—*blue*, *cyan*, and *red*; CAR–PCP–R—*magenta*). The 351-PCP–R domain is shown in surface representation (*blue* and *cyan*). NADPH from CAR–PCP–R (*magenta*) is shown. *B,* comparison of the position of Ppant in Ppant-modified 351-PCP–R and CAR–PCP–R following structural alignment of the reductase domains. CAR–PCP–R, carboxylic acid reductase–peptide carrier protein–reductase; Ppant, 4′-phosphopantetheine.

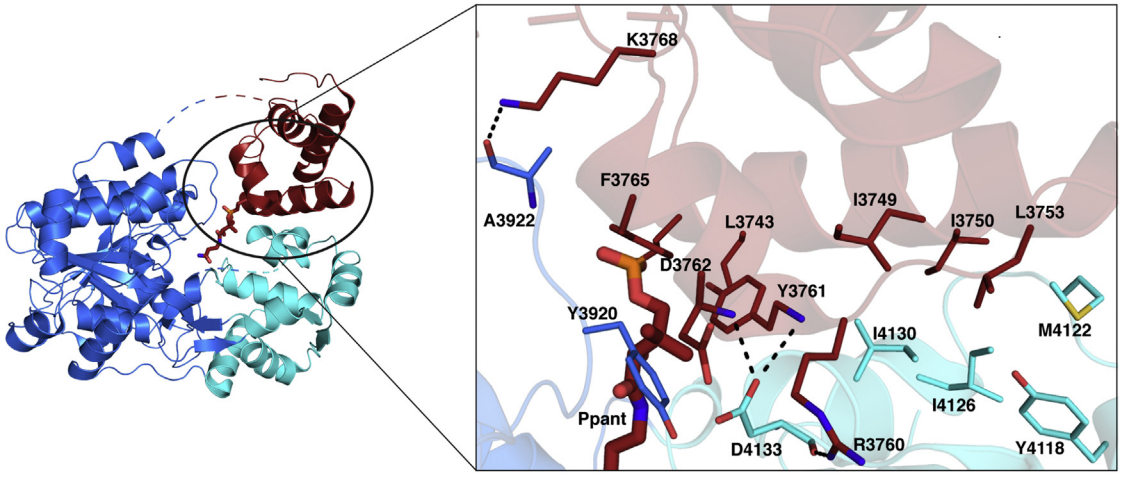

Figure 6. The 351-peptide carrier protein–reductase domain interface. Interacting residues in the peptide carrier protein domain (*red*), reductase C-terminal subdomain (*cyan*), and gating loop (*blue*) are shown in stick representation. Hydrogen bonds are shown as *black dotted lines*.

context of the peptide termination reaction. In addition to hydrophobic interactions, the side chain of D4133 from the interface–HTH motif is hydrogen bonded to the side chain of R3760 and the main chain amide groups of Y3761 and D3762 from the PCP domain, whereas the main chain carbonyl group of A3922 in the gating loop forms a hydrogen bond with the side chain of K3768 in the PCP domain. Structural water molecules present in the interface mediate further hydrogen bonds between PCP and the interface–HTH of the R domain (Fig. S5).

The conserved active-site serine residue (S3742) is present at the beginning of α-helix 2 in the PCP domain. Ppant extends from this residue into the active site of the R domain, forming extensive interactions. The geminal dimethyl group of Ppant sits in the hydrophobic pocket created by Y3920, L3743, and Y3761 (Fig. 7) in the 351-PCP–R structure and the structurally equivalent residues V901, P1130, and L1131 in the CAR–PCP–R structure. Ppant forms hydrogen-bonding interactions with the main chain amide of H3919 in the gating loop and T4032 and A4034 in the C-terminal subdomain (Fig. 7). Similar main chain hydrogen-bonding interactions between Ppant and equivalent residues in the R domain are seen in the CAR–PCP–R structure (20). Mutation to proline of residues involved in main chain hydrogen bonding with Ppant in CAR resulted in loss of catalytic activity (20). In the 351-PCP–R structure, the side chain of K3918 is hydrogen bonded to the phosphate group of Ppant. In CAR–PCP–R, the equivalent residue is an asparagine (N899), which is also hydrogen

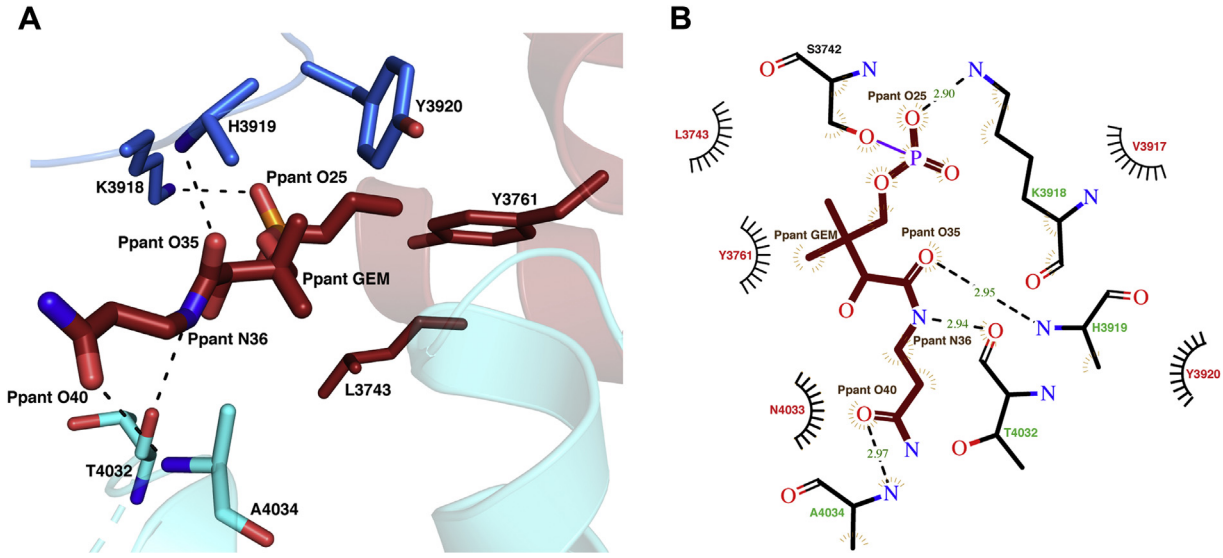

Figure 7. The interactions between the Ppant group and 351-peptide carrier protein–reductase protein. *A*, the interactions between Ppant group and peptide carrier protein domain (*red*), reductase C-terminal subdomain (*cyan*), and gating loop residues (*blue*). Hydrogen bonds are shown as *black dotted lines*. *B*, Ppant interface diagram calculated by LigPlot+ (55). Residues involved in hydrogen bonding and hydrophobic interactions with Ppant are represented in *green* and *red*, respectively. Hydrogen bonds between atoms are shown as *black dotted lines*. Ppant, 4′-phosphopantetheine; Ppant GEM, geminal dimethyl group of Ppant.

bonded to the phosphate group of Ppant. In both structures, the interactions of Ppant with the R domain are limited beyond the N41 atom, thereby explaining the weak electron density in the 351-PCP–R structure. In contrast, the presence of NADPH in the active site of CAR–PCP–R structure possibly limits the flexibility of the Ppant.

Discussion

R domains are one of the two main types of domains employed by NRPSs for peptide termination and release, the other one being terminal TE domains (5). TE domains, as the name suggests, cleave the thioester bond between Ppant and the peptide, whereas R domains use dinucleotide cofactors (NADH/NADPH) to reductively cleave the peptide tethered to the terminal PCP domain, consequently releasing aldehyde or alcohol products (5). Interactions that the PCP domain makes with its partner domains are particularly crucial for the assembly-line catalysis of NRPSs because PCP domains are tethered to the substrate amino acids and growing peptides and shuttle them between the upstream A domains and the downstream C domains (in peptide extending modules) or downstream termination domains (in peptide termination modules). Although the interactions of the PCP domain with A, C, and TE domains have been structurally characterized (14–16, 33), its interactions with the R domain are unexplored in NRPSs. The peptide termination reaction involving the PCP and TE domains differs from that of the R domain because the latter uses NAD(P)H cofactor and in some cases involves two rounds of reduction, which require cofactor recycling. This difference may result in domain movements that are distinct from those involved in the TE reaction. In this study, we describe the PCP–R didomain structure of an NRPS from *M. ruminantium*, which shows novel interdomain interactions between these "product release" domains.

Conformation of the gating loop

The gating loop in the N-terminal subdomain has been shown to be mobile in SDR family members and may block NADPH binding by occupying its binding pocket in some conformations, as seen in the Mtb-R structure (27, 31). It has been speculated that the gating loop may be stabilized by interactions with the PCP domain in conformations compatible with NADPH binding (31). Our study supports this hypothesis since the Ppant-modified 351-PCP–R and CAR–PCP–R structures reveal interactions between the gating loop, the PCP domain, and Ppant (Fig. 7*A*). It is also clear from these structures that the gating loop plays a role in fixing the otherwise flexible Ppant arm in a catalytically competent conformation in the active site. However, the gating loop in the unmodified 351-PCP–R domain structure also mimics an NADPH-bound conformation similar to that seen in the MxaA-R structure, in spite of it lacking the Ppant group and being crystallized without NADPH (Fig. 3*A*). The gating loop in the AusA-R structure, although partially disordered, also does not seem to block the NADPH-binding site (8). There is no evidence of the gating loop interacting with crystallographic symmetry–related molecules that might have resulted in the apparent NADPH-bound conformation seen in the 351-PCP–R and AusA-R structures. Hence, we suggest that even though the gating loop is mobile and more flexible in the absence of the PCP domain, Ppant, and NADPH, the NADPH-bound conformation is fairly stable and may be the predominant one.

Substrate recognition and binding in R domains

The mature peptide awaiting release is predicted to bind in the C-terminal "substrate-binding" subdomain of the R domain. While substrate binding in the C-terminal domain has been shown in some SDRs (2,4-dienoyl-CoA reductase [PDB ID: 1W6U] and CDP-D-glucose 4,6-dehydratase [PDB ID: 1WVG]) (34, 35), the substrate-binding site in NRPS R domains remains uncertain as there are no reported structures determined in complex with their substrates. Docking and molecular dynamics–based investigations to identify the putative substrate-binding pocket in NRPS R domains have predicted slightly different binding sites. The Ppant–peptide thioester in the Mtb-R structure was docked in the C-terminal subdomain based on comparison with the substrate-bound structure of 2,4-dienoyl-CoA reductase (27). The substrate was docked adjacent to a hydrophobic groove in the C-terminal subdomain, indicating that the substrate likely binds in this hydrophobic groove (27) (Fig. S6*A*). The analysis of substrate binding to the MxaA-R structure by *in silico* docking and molecular dynamics–predicted substrate binding occurs in a similar region of the protein as that predicted for Mtb-R but identified a different set of residues as interacting with the substrate (6) (Fig. S6*B*). This result was supported by biochemical evidence that showed mutation of the predicted substrate-interacting residues resulted in considerable reduction of catalytic activity compared with the wildtype enzyme (6). These differing results and the absence of a substrate-bound R domain structure leave the location of the substrate-binding pocket to further investigation. It is likely that the NRPS substrates, owing to their larger size and greater chemical and structural variability, orient themselves differently compared with the substrates of other SDRs and hence interact with different residues of the C-terminal subdomain.

A unique HTH motif forms the interface with the upstream PCP domain

We have identified that the unique HTH motif of the R domain (D4110 to F4139) forms the interface partner with the PCP domain. Our sequence alignments and structural comparisons of NRPS R domains with other extended SDRs confirmed that the interface–HTH motif is present in all NRPS R domains but is generally absent in extended SDRs (Figs. S3 and S4). SDRs themselves are subclassified into "classical" SDRs that are ∼250 amino acids long and "extended" SDRs that are identified by the presence of an additional ∼100 amino acid residues in their C-terminal subdomains (25, 36). Although NRPS R domains have been classified as extended SDRs, the presence of the unique interface–HTH motif further differentiates NRPS R domains from other members of the

extended SDR subclass, therefore suggesting that a new subclass nomenclature is required for NRPS R domains.

The interface–HTH motif shows strong electron density and interacts with the PCP domain in the Ppant-modified 351-PCP–R structure, whereas it shows weak electron density, and consequently part of it is not modeled in the unmodified structure. Examination of the electron density for the structures of AusA-PCP–R (which has a disordered PCP domain) and AusA-R (determined using a protein construct that lacked a PCP domain) (8) revealed that the electron density is weak in the interface–HTH motif region although the region could still be modeled (Fig. S7). Interestingly, this region is ordered and shows strong electron density in structures of MxaA-R (6) and Mtb-R (27), which were both determined using protein constructs that lacked PCP domains. An examination of the crystal packing in MxaA-R (Fig. S8) and Mtb-R (Fig. S9) structures revealed that the interface–HTH motifs are involved in close contacts with the ordered subdomains of symmetry-related molecules. In contrast, the interface–HTH motif shows no such contacts in the AusA-R (Fig. S10) and AusA-PCP–R structures (Fig. S11). In the unmodified 351-PCP–R structure, the partially disordered interface–HTH region is in contact with its partially disordered counterpart in the symmetry-related molecule. This suggests that the interface–HTH contacts with other disordered regions are not as stabilizing as the contacts with more ordered and rigid regions. Furthermore, molecular dynamics analysis of the MxaA-R domain suggested relatively high flexibility in the interface–HTH motif (6). Taken together, these observations strongly suggest that the interface–HTH motif is relatively mobile and becomes ordered or fixed when in contact with another ordered protein region, such as the upstream PCP domain in an intact NRPS enzyme, as reflected in the Ppant-modified 351-PCP–R structure. Protein regions that undergo disorder to order transitions upon binding are relatively common in nature (37). The functional significance of the partial disorder of the interface–HTH motif is uncertain. Interestingly, in addition to its function in forming an interface with the PCP domain, molecular dynamics and docking studies have indicated that the interface–HTH motif also plays a role in binding the chemically and structurally diverse substrates of NRPSs (6, 27). One functional advantage of disordered regions in proteins is known to be binding promiscuity (38–41), and it is possible that the partial disorder of the interface–HTH motif allows it to bind both the PCP domain and the peptide substrates.

How do NRPS R domains regulate the extent of reduction?

Investigations of the CAR–PCP–R structure (20) described a key aspartate residue present in the loop connecting N- and C-terminal R subdomains, which was previously defined as the "linker loop" by Kinatukara and co-workers (31). It was proposed that the orientation of this residue, and consequent orientation of the linker loop backbone, is dependent on the presence of the Ppant–peptide thioester in the active site. Once the product (a carboxylic acid in CAR) is reduced to an aldehyde, the resulting reduced affinity of the aldehyde product would reorient the aspartate residue and the amino acid backbone in the linker loop. This was proposed to lead to the nicotinamide moiety of NADPH becoming disordered, and thus the R domain entering an inactive state. In this inactive state, no further reduction of the aldehyde product to alcohol would occur, thereby ensuring the strict $2e^-$ reduction seen in CARs. Interestingly, this aspartate residue is replaced by a conserved glycine residue in all the known NRPS R domain structures, including the NRPS R domains that release aldehydes as their final products (8). Hence, it is still unclear how NRPS R domains regulate the extent of reduction of the peptide product. It is possible that the orientation of the product and its interactions with residues in the substrate-binding pocket influences the extent of reduction. Another possibility may be that alcohol-producing R domains may allow the dissociation of the aldehyde product after the first reduction and/or subsequent reassociation while aldehyde-producing domains do not. Structural evidence of the NRPS product orientation in the substrate-binding site and characterization of the substrate-binding pocket in general is therefore essential to understand the mechanisms that distinguish NRPS R domains based on the extent of reduction.

The novel conformation of 351-PCP–R reveals previously unexplored interdomain interactions

Examination of the NADPH-binding pocket in the Ppant-modified 351-PCP–R and CAR–PCP–R structures indicates that NADPH can likely access its binding pocket in spite of the presence of the PCP domain and Ppant (Fig. S12), suggesting that NADPH binding can occur following binding of the PCP domain. This speculation is supported by the NMR investigations of reaction mechanism of Mtb-R by Haque and co-workers (32) that suggested NRPS R domains follow a random bi–bi reaction mechanism unlike the sequential bi–bi reaction mechanism followed by SDRs. A random bi–bi mechanism implies that there is no strict order followed for binding of NADPH and the Ppant–peptide thioester. Based on the 351-PCP–R conformation and the interactions of Ppant with the gating loop, we propose a reaction mechanism for catalysis by NRPS R domains: the PCP domain tethered with the mature peptide forms an interface with the C-terminal R subdomain followed by NAD(P)H binding or vice versa; the Ppant–peptide thioester is anchored in the active site in a catalytically competent position by interactions with the gating loop and the C-terminal subdomain; the Ppant–peptide thioester is reduced, resulting in an aldehyde intermediate; in alcohol-producing domains, the oxidized NAD(P)H is replaced by its reduced counterpart, allowing a second round of reduction; $NADP^+$ and product exit the active-site pocket. It is unlikely that dissociation of the free aldehyde intermediate is necessary for cofactor exchange in the case of a $4e^-$ reduction, as the putative substrate-binding pocket is situated in the C-terminal subdomain; hence, not interfering with NADPH recycling. However, as the location of the peptide substrate-binding site is uncertain, structures of R domains complexed

ASBMB

Table 1
Gene-specific and generic primers used for PCR amplification of the PCP–R region of *mru_0351*

Name	Sequence
PCR1	
Gene-specific forward primer	5′-GGCAGCGGCGCG GTGGAACT-GAATTATGTTGCA-3′
Gene-specific reverse primer	5′-GAAAGCTGGGTG TTATTCAAAA-TAATCGAATTT-3′
PCR2	
Generic forward primer	5′-GGGGACAAGTTTGTA-CAAAAAAGCAGGCTTC-GAAAACCTGTATTTTCA GGGCAGCGGCGCG-3′
Generic reverse primer	5′-GGGGACCACTTTGTACAA GAAAGCTGGGTG-3′

The underlined residues indicate the common sequence in generic and gene-specific primers.

with Ppant–substrate thioester are essential to confirm this hypothesis.

The Ppant-modified 351-PCP–R structure interacts with the C-terminal R subdomain, in contrast to the CAR–PCP–R structure (20), where the PCP domain interacts with the N-terminal R subdomain. The 351-PCP–R interface is characterized by the participation of the interface–HTH motif. The interface–HTH motif is present in CAR–PCP–R but does not participate in the PCP–R interface. In the unmodified 351-PCP–R structure, the PCP domain is positionally disordered, whereas in the Ppant-modified structure, it is locked in the observed conformation as a consequence of Ppant modification. The Ppant-modified 351-PCP–R structure, although it does not possess a covalently tethered peptide, represents the native NRPS PCP–R didomain more closely than the unmodified structure, further reinforcing the relevance of this didomain conformation. Conserved residues from both domains participate in the interface, which encompasses a total buried surface area of ∼1250 Å^2. Analysis of the 351-PCP–R interface using the Proteins, Interfaces, Structures and Assemblies server hosted at the Protein Data Bank in Europe gave a complex formation significance score of 0.585, implying that the interface plays an essential role in complex formation (42). The CAR–PCP–R, on the other hand, has a smaller total buried area at the interface (∼950 Å^2) compared with that of 351-PCP–R, indicating a relatively weaker interface. In addition, only a small part of the "classic" PCP region involved in NRPS PCP–partner domain interfaces is seen to be participating in the CAR–PCP–R interface. Furthermore, computational docking of the MxaA-R domain with the PCP domain revealed a similar interface where α-helix 3 of the PCP domain is predicted to form interactions with the interface–HTH motif as seen in the Ppant-modified 351-PCP–R structure (6). Taken together, these results are compelling evidence that the conformation observed in Ppant-modified 351-PCP–R is likely to be the catalytic interface involved in the NRPS peptide termination reaction.

The significance of the two differing orientations observed for the PCP–R interaction is uncertain. It is likely that both orientations are biologically relevant given that the Ppant is oriented in the active site in a catalytically competent orientation in both 351-PCP–R and CAR–PCP–R. Furthermore, the PCP domain adopts the same orientation in all CAR–PCP–R didomain structures independent of the crystal packing (20), indicating that the orientation observed is unlikely to be an artifact of crystallization. Biochemical assays based on mutation of interface residues in the CAR–PCP–R structure have been limited to the Ppant–R domain interactions and did not investigate the effect of mutating interface residues of PCP–R interface on CAR catalysis (20). It is therefore unclear whether the conformation observed in CAR–PCP–R is unique to CARs or whether it is a conformation shared with NRPSs, possibly representing a transient state or an intermediate state. Since the product of Mru_0351 is unknown, it is currently not possible to carry out biochemical assays based on mutation of the interface residues to further validate the biological significance of the interface.

In conclusion, this study presents the first archaeal NRPS domain structure to be reported and the first NRPS R domain structure to be determined together with the upstream PCP domain, thereby elucidating the previously unknown NRPS PCP–R interactions. We have identified the function of the unique interface–HTH motif as the interface partner with the upstream PCP domain. Further structural and biochemical investigations of NRPS R domains with cognate substrates are essential to characterize the substrate-binding pocket. The key residues of the PCP–R interface identified in the 351-PCP–R structure together with knowledge of the substrate-binding pocket may have broad implications in understanding the mechanisms underlying R domain reactions and in engineering NRPS R domains for *in vitro* synthesis of novel peptide products.

Experimental procedures

PCR amplification and cloning of the PCP–R didomain gene

The PCP–R region (encoding residues 3701–4187) of the *mru_0351* gene was cloned using the proprietary Gateway Cloning Technology from Invitrogen. The genes encoding the R and PCP–R regions were amplified by a nested PCR approach. The first PCR reaction was performed using gene-specific primers and introduced a tobacco etch virus (TEV) protease cleavage site at the 5′ end of the construct (Table 1). The second PCR reaction was performed using generic primers and introduced *attB1* and *attB2* recombination sequences at both ends of the constructs (Table 1). The amplified gene comprising the *attB* sequences on their 5′ and 3′ ends was inserted into the pDONR221 donor entry vector. Subsequently, the gene was transferred to the destination expression vector pET-53-DEST, which possesses an N-terminal polyhistidine tag to facilitate purification of the recombinant protein.

Protein production and purification

Electrocompetent *Escherichia coli* BL21 Star(DE3) cells were transformed with the pET-53-DEST gateway destination expression vector containing the gene encoding the PCP–R region. Overexpression was carried out using an auto-induction strategy in high-density shaking cell cultures (43).

Cultures grown in ZYP-5052 medium were incubated for 4 h at 37 °C with shaking at 180 rpm followed by incubation at 18 °C for 20 h with shaking at 180 rpm. After a total of 24 h of incubation, the cells were harvested by centrifugation at 5300*g* for 30 min at 4 °C and stored at −20 °C.

Cell pellets were thawed and suspended in lysis buffer (10 ml lysis buffer per gram of cell pellet) containing 50 mM sodium phosphate buffer at pH 7.5, 500 mM NaCl, 10 mM imidazole, and 5 mM β-mercaptoethanol. Lysozyme (1 mg/ml final concentration), DNase (10 μg/ml final concentration), and a tablet of EDTA-free protein inhibitor cocktail (cOmplete from Roche Diagnostics) were added to the cell suspension and incubated for 30 min at 4 °C. The cells were lysed using a continuous cell disruptor (M-110P Microfluidizer from Constant Systems Ltd) at pressure of 18.5 kPa. The lysate was centrifuged at 20,000*g* for 30 min at 4 °C. The supernatant was loaded onto a 5 ml nickel–nitrilotriacetic acid agarose column pre-equilibrated with lysis buffer. The column was washed with five column volumes of lysis buffer. Protein bound to the column was eluted with a continuous gradient of imidazole from 100 to 500 mM. His-351-PCP–R eluted between 150 and 250 mM imidazole. In the next step, fractions containing His-351-PCP–R were pooled together and dialyzed in dialysis membrane tubing (CelluSep from Interchim) against the lysis buffer overnight at 4 °C to remove imidazole. Recombinant TEV protease was added during dialysis to cleave the polyhistidine tag with 5 mM β-mercaptoethanol added during dialysis to ensure reducing conditions required for activity of TEV protease. After overnight dialysis, the protein was loaded on to the nickel–nitrilotriacetic acid agarose column pre-equilibrated with lysis buffer. His-TEV protease as well as any remaining His-351-PCP–R were bound to the resin, whereas tag-free 351-PCP–R was retained in the flowthrough. The flowthrough was concentrated to a volume of 250 μl using Amicon ultra-15 centrifugal filter units with molecular weight cutoff of 30 kDa. Concentrated 351-PCP–R was loaded onto a HiLoad 16/600 Superdex 75 pg column from Cytiva pre-equilibrated with 25 mM Hepes buffer (pH 7.5) and 150 mM NaCl. The 351-PCP–R protein eluted as a single peak. Fractions containing the purified 351-PCP–R were pooled together and concentrated to 30 mg/ml.

Ppant modification

The modification of 351-PCP–R with Ppant (on the conserved serine residue S3742 of the PCP domain) was carried out enzymatically using recombinant Sfp, a phosphopantetheine transferase from *Bacillus subtilis*. Sfp was produced by recombinant expression in *E. coli*. The 351-PCP–R protein (0.5 mM) was incubated with 1 mM coenzyme A, 1 mM of Sfp, and 10 mM $MgCl_2$ for 90 min at 30 °C. This step was carried out before size-exclusion chromatography to facilitate separation of Sfp from 351-PCP–R. Tandem mass spectrometry was used to determine the presence and efficiency of Ppant modification. Mass spectrometric analysis showed that over 99% of the protein was modified by Ppant (data not shown).

Crystallization and X-ray diffraction

The 351-PCP–R protein (30 mg/ml in 25 mM sodium Hepes buffer at pH 7.5 and 150 mM NaCl) was crystallized using the sitting-drop vapor-diffusion method. Initial screening was performed using commercially available crystallization screens from Molecular Dimensions: JCSG-plus, Morpheus, and PACT premier with drop size of 0.4 μl (0.2 μl protein + 0.2 μl reservoir solution) equilibrated against reservoir volume of 50 μl at 20 °C. The unmodified 351-PCP–R formed small needle-shaped crystals in Morpheus screen well H5 (0.1 M Mops/Na–Hepes buffer at pH 7.5, 20% PEG monomethylether 550, and 10% PEG 20 K as precipitants in the presence of 0.02 M each of L-Na glutamate, DL-alanine, glycine, DL-lysine hydrochloride, and DL-serine as additives) after about 7 days of incubation. Fine screening was performed around this condition by varying the concentration of precipitants and increasing the drop size to 2 μl (1 μl protein + 1 μl reservoir solution) while keeping all the other parameters constant. Fine screening was able to produce bigger needle-shaped crystals under precipitant concentrations of 15% PEG monomethylether 550 and 14% PEG 20 K. The crystals were flash cooled in liquid nitrogen for data collection. The modified 351-PCP–R formed stacked hexagonal-shaped plate-like crystals overnight in JCSG-plus screen condition F8 (2.1 M DL-malic acid disodium salt at pH 7.0). Fine screening around this condition by varying the pH and concentration of salt and increasing the drop size to 2 μl (1 μl protein + 1 μl reservoir solution) resulted in the formation of unstacked hexagonal plate-like crystals in 1.7 M malic acid disodium salt at pH 6.0. The crystals were soaked in the reservoir solution with 20% glycerol as cryoprotectant for less than a minute before flash cooling in liquid nitrogen. Data were collected at the Australian Synchrotron, MX2 beamline.

Data processing, phasing, and refinement

X-ray detector software (XDS) was used for indexing, integration, and scaling of the diffraction data (44). Aimless from CCP4 suite was used for merging the data (45). The crystal structure of the aureusimine biosynthetic cluster R domain from *S. aureus* strain Mu50 (PDB ID: 4F6C; sequence identity 28% to Mru_0351 R domain) (8) was used as a template for molecular replacement for both unmodified and Ppant-modified 351-PCP–R. Phaser from the phenix package was used for molecular replacement, and phenix.refine was used for refinement (46). Efforts at conducting molecular replacement using existing PCP domain structures as search models were unsuccessful. In the case of the unmodified 351-PCP–R, the PCP domain could not be modeled as the corresponding density was weak or nonexistent, which is likely because of the domain being positionally disordered. In the case of the Ppant-modified 351-PCP–R, the initial model resulting from molecular replacement of the R domain was used as a template to autobuild the model using Arp/Warp (47, 48). The PCP domain was built in contact with the R domain with density for Ppant visible in the active-site region of the R domain. The electronic Ligand Builder and Optimization Workbench was used to generate restraints for Ppant (49). Both the unmodified

ASBMB

and Ppant-modified 351-PCP–R models were further refined *via* iterative cycles of automated refinement using phenix.refine and manual rebuilding using Coot (46, 50). The structural figures were generated using PyMOL (The PyMOL Molecular Graphics System, Version 2.3.4 Schrödinger, LLC) or Coot (50). The data collection, refinement, and final model statistics for modified and unmodified 351-PCP–R are shown in Table S2.

Data availability

The unmodified and Ppant-modified structures of 351-PCP–R have been deposited in the PDB with codes 6VTZ and 6VTJ, respectively.

Supporting information—This article contains supporting information (6, 8, 20, 27, 30, 51–54).

Acknowledgments—This research was undertaken in part using the MX2 beamline at the Australian Synchrotron, part of the Australian Nuclear Science and Technology Organization and made use of the Australian Cancer Research Foundation detector. We thank all the beamline staff of the MX beamline for their support. DNA sequencing was provided by the Center of Genomics, Proteomics and Metabolomics within the School of Biological Sciences. We thank Martin Middleditch for his support with mass spectrometry. We thank Dr Stephanie Dawes for her advice during writing of the article.

Author contributions—E. A., V. S., J. S. L., and T. V. L. conceived the project. All authors were involved in designing the experiments. S. D. conducted the experiments and determined the X-ray structures. All authors were involved in analysis and interpretation of the results. S. D. and T. V. L. wrote the original draft. All authors edited the article.

Funding and additional information—This research was supported by the New Zealand Ministry for Business, Innovation and Employment on contract UOAX1506.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: A, adenylation; AusA-R, R domain of a dimodular NRPS from the aureusimine biosynthetic cluster in *Staphylococcus aureus* strain Mu50; C, condensation; CAR, carboxylic acid reductase; HTH, helix–turn–helix; Mtb-R, R domain of a *Mycobacterium tuberculosis* NRPS; MxaA-R, R domain of the NRPS module of the myxalamid biosynthetic pathway; NRPS, nonribosomal peptide synthetase; PCP, peptide carrier protein; PDB, Protein Data Bank; Ppant, 4′-phosphopantetheine; R, reductase; SDR, short-chain dehydrogenase/reductase; TE, thioesterase; TEV, tobacco etch virus.

References

1. Marahiel, M. A., Stachelhaus, T., and Mootz, H. D. (1997) Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651–2673
2. Süssmuth, R. D., and Mainz, A. (2017) Nonribosomal peptide synthesis—principles and prospects. *Angew. Chem. Int. Edition* **56**, 3770–3821
3. Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9259–9264
4. Hur, G. H., Vickery, C. R., and Burkart, M. D. (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat. Product Rep.* **29**, 1074–1098
5. Du, L., and Lou, L. (2010) PKS and NRPS release mechanisms. *Nat. Product Rep.* **27**, 255–278
6. Barajas, J. F., Phelan, R. M., Schaub, A. J., Kliewer, J. T., Kelly, P. J., Jackson, D. R., Luo, R., Keasling, J. D., and Tsai, S.-C. (2015) Comprehensive structural and biochemical analysis of the terminal myxalamid reductase domain for the engineered production of primary alcohols. *Chem. Biol.* **22**, 1018–1029
7. Mullowney, Michael W., McClure, R. A., Robey, M. T., Kelleher, N. L., and Thomson, R. J. (2018) Natural products from thioester reductase containing biosynthetic pathways. *Nat. Product Rep.* **35**, 847–878
8. Wyatt, M. A., Mok, M. C. Y., Junop, M., and Magarvey, N. A. (2012) Heterologous expression and structural characterisation of a pyrazinone natural product assembly line. *ChemBioChem* **13**, 2408–2415
9. Conti, E., Stachelhaus, T., Marahiel, M. A., and Brick, P. (1997) Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S. *EMBO J.* **16**, 4174–4183
10. Lee, T. V., Johnson, L. J., Johnson, R. D., Koulman, A., Lane, G. A., Lott, J. S., and Arcus, V. L. (2010) Structure of a eukaryotic nonribosomal peptide synthetase adenylation domain that activates a large hydroxamate amino acid in siderophore biosynthesis. *J. Biol. Chem.* **285**, 2415–2427
11. Keating, T. A., Marshall, C. G., Walsh, C. T., and Keating, A. E. (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526
12. Weber, T., Baumgartner, R., Renner, C., Marahiel, M. A., and Holak, T. A. (2000) Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. *Structure* **8**, 407–418
13. Bruner, S. D., Weber, T., Kohli, R. M., Schwarzer, D., Marahiel, M. A., Walsh, C. T., and Stubbs, M. T. (2002) Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrfTE. *Structure* **10**, 301–310
14. Liu, Y., Zheng, T., and Bruner, S. D. (2011) Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem. Biol.* **18**, 1482–1488
15. Mitchell, C. A., Shi, C., Aldrich, C. C., and Gulick, A. M. (2012) Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* **51**, 3252–3263
16. Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Allen, C. L., Skiniotis, G., Aldrich, C. C., and Gulick, A. M. (2016) Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238
17. Reimer, J. M., Eivaskhani, M., Harb, I., Guarné, A., Weigt, M., and Schmeing, T. M. (2019) Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, 1–7
18. Tanovic, A., Samel, S. A., Essen, L. O., and Marahiel, M. A. (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663
19. Tarry, M. J., Haque, A. S., Bui, K. H., and Schmeing, T. M. (2017) X-Ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure* **25**, 783–793
20. Gahloth, D., Dunstan, M. S., Quaglia, D., Klumbys, E., Lockhart-Cairns, M. P., Hill, A. M., Derrington, S. R., Scrutton, N. S., Turner, N. J., and Leys, D. (2017) Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis. *Nat. Chem. Biol.* **13**, 975–981
21. Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C., and Gulick, A. M. (2012) Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem. Biol.* **19**, 188–198
22. Miller, B. R., Sundlov, J. A., Drake, E. J., Makin, T. A., and Gulick, A. M. (2014) Analysis of the linker region joining the adenylation and carrier

protein domains of the modular nonribosomal peptide synthetases. *Proteins Struct. Funct. Bioinformatics* **82**, 2691–2702
23. Bozhüyük, K. A. J., Linck, A., Tietze, A., Kranz, J., Wesche, F., Nowak, S., Fleischhacker, F., Shi, Y.-N., Grün, P., and Bode, H. B. (2019) Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661
24. Owen, J. G., Calcott, M. J., Robins, K. J., and Ackerley, D. F. (2016) Generating functional recombinant NRPS enzymes in the laboratory setting via peptidyl carrier protein engineering. *Cell Chem. Biol.* **23**, 1395–1406
25. Kavanagh, K. L., Jörnvall, H., Persson, B., and Oppermann, U. (2008) Medium- and short-chain dehydrogenase/reductase gene and protein families: The SDR superfamily: Functional and structural diversity within a family of metabolic and regulatory enzymes. *Cell Mol. Life Sci.* **65**, 3895–3906
26. Persson, B., and Kallberg, Y. (2013) Classification and nomenclature of the superfamily of short-chain dehydrogenases/reductases (SDRs). *Chemi. Biol. Interact.* **202**, 111–115
27. Chhabra, A., Haque, A. S., Pal, R. K., Goyal, A., Rai, R., Joshi, S., Panjikar, S., Pasha, S., Sankaranarayanan, R., and Gokhale, R. S. (2012) Non-processive [2 + 2]e- off-loading reductase domains from mycobacterial nonribosomal peptide synthetases. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 5681–5686
28. Janssen, P. H., and Kirs, M. (2008) Structure of the archaeal community of the rumen. *Appl. Environ. Microbiol.* **74**, 3619–3625
29. Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., Li, D., Kong, Z., McTavish, S., Sang, C., Lambie, S. C., Janssen, P. H., Dey, D., and Attwood, G. T. (2010) The genome sequence of the rumen methanogen Methanobrevibacter ruminantium reveals new possibilities for controlling ruminant methane emissions. *PLoS One* **5**, e8926
30. Krissinel, E. B., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. D Struct. Biol.* **60**, 2256–2268
31. Kinatukara, P., Patel, K. D., Haque, A. S., Singh, R., Gokhale, R. S., and Sankaranarayananan, R. (2016) Structural insights into the regulation of NADPH binding to reductase domains of nonribosomal peptide synthetases: A concerted loop movement model. *J. Struct. Biol.* **194**, 368–374
32. Haque, A. S., Patel, K. D., Deshmukh, M. V., Chhabra, A., Gokhale, R. S., and Sankaranarayanan, R. (2014) Delineating the reaction mechanism of reductase domains of nonribosomal peptide synthetases from mycobacteria. *J. Struct. Biol.* **187**, 207–214
33. Frueh, D. P., Arthanari, H., Koglin, A., Vosburg, D. A., Bennett, A. E., Walsh, C. T., and Wagner, G. (2008) Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903–906
34. Koropatkin, N. M., and Holden, H. M. (2005) Structure of CDP-D-glucose 4,6-dehydratase from Salmonella typhi complexed with CDP-D-xylose. *Acta Crystallogr. D Struct. Biol.* **61**, 365–373
35. Alphey, M. S., Yu, W., Byres, E., Li, D., and Hunter, W. N. (2005) Structure and reactivity of human mitochondrial 2,4-Dienoyl-CoA reductase: Enzyme-ligand interactions in a distinctive short-chain reductase active site. *J. Biol. Chem.* **280**, 3068–3077
36. Yvonne, K., Udo, O., and Bengt, P. (2010) Classification of the short-chain dehydrogenase/reductase superfamily using hidden Markov models. *FEBS J.* **277**, 2375–2386
37. Chakravarty, D., Janin, J., Robert, C. H., and Chakrabarti, P. (2015) Changes in protein structure at the interface accompanying complex formation. *IUCrJ* **2**, 643–652
38. Fonin, A. V., Darling, A. L., Kuznetsova, I. M., Turoverov, K. K., and Uversky, V. N. (2019) Multi-functionality of proteins involved in GPCR and G protein signaling: Making sense of structure–function continuum with intrinsic disorder-based proteoforms. *Cell Mol. Life Sci.* **76**, 4461–4492
39. Weng, J., and Wang, W. (2020) Dynamic multivalent interactions of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **62**, 9–13
40. Oldfield, C. J., and Dunker, A. K. (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* **83**, 553–584
41. Tompa, P., and Fuxreiter, M. (2008) Fuzzy complexes: Polymorphism and structural disorder in protein–protein interactions. *Trends Biochem. Sci.* **33**, 2–8
42. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
43. Fox, B. G., and Blommel, P. G. (2009) Autoinduction of protein expression. *Curr. Protoc. Protein Sci.* Chapter 5, Unit-5.23
44. Kabsch, W. (2010) Xds. *Acta Crystallogr. D Struct. Biol.* **66**, 125–132
45. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 235–242
46. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkoczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L.-W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: Recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877
47. Langer, G., Cohen, S. X., Lamzin, V. S., and Perrakis, A. (2008) Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat. Protoc.* **3**, 1171–1179
48. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Struct. Biol.* **D67**, 355–367
49. Moriarty, N. W., Grosse-Kunstleve, R. W., and Adams, P. D. (2009) Electronic Ligand Builder and Optimization Workbench (eLBOW): A tool for ligand coordinate and restraint generation. *Acta Crystallogr. D Biol. Crystallogr.* **65**, 1074–1080
50. Emsley, P., and Cowtan, K. (2004) Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Struct. Biol.* **60**, 2126–2132
51. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539
52. Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C., and Adams, P. D. (2017) Polder maps: Improving OMIT maps by excluding bulk solvent. *Acta Crystallogr. D Struct. Biol.* **73**, 148–157
53. Pei, J., and Grishin, N. V. (2014) PROMALS3D: Multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information. *Methods Mol. Biol.* **1079**, 263–271
54. [preprint] Pluskal, T., Torrens-Spence, M. P., Fallon, T. R., De Abreu, A., Shi, C. H., and Weng, J.-K. (2018) The biosynthetic origin of psychoactive kavalactones in kava. *bioRxiv*. https://doi.org/10.1101/294439
55. Laskowski, R. A., and Swindells, M. B. (2011) LigPlot+: Multiple ligand–protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **51**, 2778–2786