New Structural Data Reveal the Motion of Carrier Proteins in Nonribosomal Peptide Synthesis

Tiia Kittilä, Aurelio Mollo, Louise K. Charkoudian,* and Max J. Cryle*

Dedicated to Mohamed Marahiel

The nonribosomal peptide synthetases (NRPSs) are one of the most promising resources for the production of new bioactive molecules. The mechanism of NRPS catalysis is based around sequential catalytic domains: these are organized into modules, where each module selects, modifies, and incorporates an amino acid into the growing peptide. The intermediates formed during NRPS catalysis are delivered between enzyme centers by peptidyl carrier protein (PCP) domains, which makes PCP interactions and movements crucial to NRPS mechanism. PCP movement has been linked to the domain alternation cycle of adenylation (A) domains, and recent complete NRPS module structures provide support for this hypothesis. However, it appears as though the A domain alternation alone is insufficient to account for the complete NRPS catalytic cycle and that the loaded state of the PCP must also play a role in choreographing catalysis in these complex and fascinating molecular machines.
condensation (C), and peptidyl carrier protein (PCP, also known as thiolation, (T); Figure 1B) domains.

The A domains select the amino acid monomer \([5,6]\) activate it by using ATP, and then load it onto the adjacent PCP domain. NRPS A domains comprise two subdomains, with the motions of the smaller (C-terminal) subdomain allowing both selection/activation and PCP-loading steps to occur within the same active site of the larger subdomain. \([7]\) Bacterial A domains have highly conserved structures, and their sequences can generally be used to predict the backbone peptide structure of the NRP. This is also a highly useful property for potential rational redesign of NRPSs. \([5,6,8]\)

C domains catalyze peptide bond formation between PCP-bound substrates: two PCPs bind to the C domain and the amine of a downstream aminoacyl-PCP attacks the thioester of the upstream peptidyl-PCP to form a new peptide bond. In this manner, the NRP is transferred downstream and elongated by one residue each cycle. \([9]\) The C domain plays an important role in the maintenance of the stereochemistry of the growing NRP, \([10]\) although our current understanding is limited by a lack of structures for substrate-bound C domains.

Shuttling of intermediates between the A and C domains is performed by the PCP domain, a small \((ca. 10 \text{ kDa})\) catalytically silent domain bearing an 18 Å phosphopantetheine (Ppant) arm that is added post-translationally to a conserved serine residue at the N terminus of helix II (Figure 2A). \([11]\) The Ppant moiety acts as a swinging arm to increase the “reach” of the PCP domain into the active sites of adjacent domains. Crucially, the Ppant arm terminates in a thiol moiety, which allows intermediates to be shuttled as thioesters whilst remaining reactive enough to support peptide bond formation in the C domain.

The minimal C/A/PCP architecture is supplemented by an array of additional NRPS domains: the most common are epimerization (E) domains, which epimerize the l-amino acids activated by A domains into their D-form, and thioesterase (TE) domains, which release the completed NRP from the assembly line. \([12]\) Further diversity can be achieved through additional cis domains, such as formylation, oxidation, and methylation domains, \([13]\) or through interaction with enzymes that act in trans to the main NRPS machinery. \([11–13]\) In both cases, these interactions occur on PCP-bound substrates, with PCP domains playing a major role in the recruitment of enzymes in trans. \([11–13]\) An exception to PCP-driven recruitment is the X domain (an additional C/E-type domain) from glycopeptide antibiotic biosynthesis, which is required to recruit cytochrome P450s to the PCP-bound substrate to perform crosslinking of aromatic side chains. \([14,15]\) Diversification of “standard” domain chemistry is also increasingly being observed in NRPS assembly lines: recent discoveries include β-lactam formation, \([16]\) and TE-catalyzed epimerization \([17]\) in norcardicin biosynthesis, and transesterification in salinamide biosynthesis. \([18]\)

3. The Peptidyl Carrier Protein

The secondary and tertiary structures of PCP domains are highly conserved, with only minor deviations from the prototypical four-helix bundle being documented to date. \([19–27]\) At the level of primary structure, PCPs are more variable, \([21]\) which gives rise to variations in local shape and charge distributions of the exposed and buried surfaces. This affects how individual PCPs interact with catalytic partners, \([21]\) especially in trans, where PCP recognition is crucial to selectivity. \([11,20]\)

An early hypothesis to explain the mechanism of NRPSs was based around large changes in PCP tertiary structure as
a result of changes in PCP loading state, which would in turn affect and govern PCP interactions. This model was based on observations that a PCP domain from the tyrocidine synthetase exists in three different conformations (termed A, A/H, and H) depending on its loaded state. However, this has since been dismissed as an artifact of PCP domain excision from the larger synthetase since all other PCP structures elucidated to date adopt the A/H state. The role of Ppant and substrate loading on PCP tertiary structure is therefore unclear. Analogous carrier proteins from polyketide synthetases can sequester their molecular cargo, although the biochemical relevance of this mechanism is not known. In the case of PCPs, the Ppant arm does not appear to interact appreciably with the protein core nor to alter the tertiary structure of the PCP in a significant way. This observation supports the “swinging arm hypothesis”, in which the flexible Ppant arm delivers substrates to adjacent domains and the PCP domain serves as a largely rigid and chemically inert platform. Two recent NMR studies suggest that PCPs can interact weakly with the Ppant arm (Figure 2A), although these interactions were observed in atypical PCPs: the aryl-acid-loaded PCP from yersiniabactin synthetase and the pyrrole-loaded PCP from pyoluteorin synthetase. In both cases, the Ppant arm was found to interact with helices II and III. While these findings suggest that the loaded state of the PCP could modulate NRPS interactions, our current understanding of PCP chain sequestration is limited by a lack of data on both prototypical and non-excised PCP domains.

4. PCP Domain interactions With NRPS Catalytic Domains

Structural studies have shown that PCP domains use a similar protein surface in interactions with A, C, and TE domains and that these interactions are mainly mediated by helices II/III and the loop connecting them (Figure 2B). Hydrophobic interactions are commonly described, but since cognate domains are selective to their partner PCP(s), further interactions, such as variable charge distribution, must also play important roles in recognition. Structures of holo-PCPs in multidomain environments revealed that Ppant arm loading strengthens domain interac-
This appears to be especially crucial for PCP interactions in trans, for example with P450 enzymes. Given the restricted motion available to PCP domains in a multidomain context, it is perhaps unsurprising that further interaction interfaces between PCPs and other partner domains have been demonstrated. For example, C and TE domains can interact with a PCP domain simultaneously at distinct interaction sites. Furthermore, in the recent multidomain AB3403 structure, the PCP domain is “packed” between the C and TE domains, revealing alternative helix I interactions with the TE domain when the Ppant arm is located in the binding site of the C domain. Given the subtle nature of substrate-induced PCP conformational changes, it appears unlikely that this is the sole driving force for partner-domain recruitment: clearly, another process must be driving NRPS catalysis.

5. Adenylation Domain Conformational Changes Guide PCP Interactions

The first structure of a complete NRPS module (SrfA-C, C/A/PCP/TE) showed that while the Ppant linker site of the PCP is 16 Å from the C domain active site (and therefore within reach of its 18 Å Ppant arm), the same site is 57 and 43 Å away from the A and TE domain active sites, respectively. Substantial movements and domain rearrangements must thus take place within each module to allow the PCP domain to interact with all of its partners. It has been suggested that these PCP movements are coupled directly to the catalytic cycle of the A domains (Figure 3). During substrate activation, the small subdomain of the A domain is located close to the large subdomain (closed state), from where it rotates 140° to reveal the active site for subsequent PCP thiolation. The cycle is completed with a third, open state to allow substrate binding. Gulick and co-workers identified a conserved LPxP motif in the linker region between the PCP and A domains that forms a stable interaction with the small subdomain of the A domain, thereby shortening the effective length of the linker. This interaction could serve to couple conformational changes in the A domain to movements of the PCP domain. In support of this model, the efficiency of A domains is enhanced by the presence of the PCP domain. Further, A-domain activity is faster in intact A/PCP di-domains compared to excised domains, thus suggesting that the functions of the A and PCP domains are closely intertwined. The coupling of A-domain activity to PCP motion has important implications for NRPS redesign, since alterations to the rate of A-domain activity made during efforts to alter substrate selection would impact the overall efficiency of NRPS catalysis.

Recently published structures of two NRPS termination modules crystallized in different states (thiolation and condensation) further support the alternation theory (Figure 3). Nonetheless, the structures do raise questions as to how A and PCP domain movements are coupled to each other. For example, in the holo-AB3404 structure (C/A/PCP/TE, condensation state), the A domain is in a closed state, which is inconsistent with the open state seen in the preceding SrfA-C structure. Also, in the case of the recent gramicidin synthetase structures (LgrA, formylation (F)/A/PCP), the PCP interaction with the F domain would clash with the open state conformation of the A domain seen in the SrfA-C structure. Unfortunately, the structure of the PCP is distorted...
in the LgrA structure, which displays a closed state of the A domain, thus leaving the question of PCP localization open to further analysis.

### 6. PCP Movements Guided by Substrate State

The apparent contradiction of PCP localization in these whole-module structures appears to require an expansion of the theory that domain alternation is the sole driving force underlying NRPS catalysis. Conformational changes in C domains have been reported, but the function of these changes is unclear. The structure of AB3403 showed the C domain in a closed state, with strong interactions between the Ppant arm and the C domain, whilst in the apo-SrfC-A structure, the C domain is in an open state. Recent evidence shows that an A domain can catalyze two activation cycles (one to load the PCP and one to activate the next amino acid) on an intact module, which then halts in the absence of an upstream donor substrate. These observations, combined with the A-domain states observed in complete-module structures, suggest that interaction of the PCP with the upstream

---

*Figure 3.* The A domain guides PCP domain movement. Large conformational rearrangements are needed for the PCP domain to move between the A- and C-domain active sites. This movement is mediated by the domain alternation cycle of the A domain. Substrate loading state could mediate further movements to downstream domains. A) The catalytic cycle of an initiation module. B) The catalytic cycle of a termination module. Yellow = A, large subunit (A); gray = A, small subunit (A); multicolored cylinders/pink circles = T (PCP) domain, brown = F domain, blue = C domain.
C domain “freezes” the A domain at the adenylation step if peptide is not available at the C-domain donor site. This makes mechanistic sense, since further movement of the PCP to the thiolation state is not desirable before the bound substrate is released from the PCP domain.

A recent structure of a C domain bound to a tethered amino acid Ppant mimic revealed strong interactions between the binding pocket and the ε-amino group of the substrate.\[44\]

Since peptide bond formation alters this interaction, this could provide a sensor signal for C-domain closing and PCP release, although this hypothesis needs further study. The mechanistic details that underpin C-domain activity and the PCP interactions that support peptide bond formation are an area of great importance for future NRPS research, with biochemical studies to date lagging behind in vivo experiments.\[45,46\]

Although significant insight into the NRPS mechanism has been gained from recent complete-module structures, knowledge of PCP domain interactions with downstream C and TE domains remains limited. Insight into these interactions can be gained however from a chemically trapped PCP/TE di-domain structure.\[24\] Overlaying the TE domain from this structure with the TE domain from the AB3403 structure shows how the lid region of the TE domain changes upon substrate binding (Figure 4). Strikingly, a simple rotation of the PCP domain appears to be sufficient to deliver the substrate from the C domain to the TE domain, and this could be accomplished without major structural rearrangement of the NRPS. Furthermore, a PCP/C di-domain structure revealed the upstream PCP domain located close to the donor binding site of the C domain, but with the Ppant attachment site rotated away by 180° when the PCP domain is not loaded.\[47\] This closely resembles the interaction between the PCP and TE domains in the AB3403 structure.\[26\]

Therefore, PCP loading state and acceptor-domain flexibility, rather than A-domain rearrangement, might well be the driving force for interaction between the PCP and C/TE domains. This model would be supported by the free movement of the TE domain seen in the EntF structure,\[28\] since free rotation of the TE domain would facilitate binding of the substrate once it is released from the upstream C domain. A similar strategy could potentially be used to guide PCP domains to tailoring enzymes, but to date no structural information exists to examine this hypothesis.

7. Future Directions

Recent whole-module structures provide insight into the mechanisms underpinning NRPS function and indicate that PCP domain movement is achieved through a combination of conformational changes and substrate loading states. The next challenge for the field is to understand the interactions between NRPS modules, since no structure of a multimodular NRPS has been resolved thus far. Marahiel recently proposed a model for a complete NRPS system based on multidomain structures, whereby a helical organization is achieved by rotating each module 120° along the helical axis, thereby leaving PCP domains close to the axis in such a way that intermediates would be protected from hydrolysis.\[48\] A helical structure would orientate the C domains from adjacent modules in close proximity, thus allowing a simple rotation of the PCP domain to deliver substrate from one C domain to the next, which fits well with the current proposed catalytic model of an NRPS assembly line. While structural studies provide important insight into NRPS systems, they now need to be complemented with further biochemical characterization of complete modules, since it has become obvious that domain interactions both between and within modules play vital roles in catalysis.\[36,41,45,46\] Such studies are especially crucial to enable the effective redesign of NRPS machineries, since understanding the sources of substrate specificity for complete NRPS modules is currently lagging behind in vivo NRPS redesign.\[8\] Given the importance of NRPs, these efforts are clearly needed to improve our access to derivatives of these medically relevant natural products.

Acknowledgements

The authors are grateful for financial support from Haverford College (to A.M. and L.K.C.), the Research Corporation for Science Advancement Cottrell Scholars Award (to L.K.C.), the Deutsche Akademischer Austausch Dienst (Graduate School Scholarship Program to T.K.) and Monash University, the Deutsche Forschungsgemeinschaft Emmy-Noether Program, CR 392/1-1 (to M.J.C.).

How to cite: Angew. Chem. Int. Ed. 2016, 55, 9834–9840

Angew. Chem. 2016, 128, 9988–9995

[1] D. J. Newman, G. M. Cragg, J. Nat. Prod. 2007, 70, 461.
[2] E. A. Felnagle, E. E. Jackson, Y. A. Chan, A. M. Podeval, A. D. Berti, M. D. McMahon, M. G. Thomas, Mol. Pharmaceutics 2008, 5, 191.
[3] G. H. Hur, C. R. Vickery, M. D. Burkt, Nat. Prod. Rep. 2012, 29, 1074.
[4] C. T. Walsh, R. V. O’Brien, C. Khosla, Angew. Chem. Int. Ed. 2013, 52, 7098; Angew. Chem. 2013, 125, 7238.
[5] G. L. Challis, J. Ravel, C. A. Townsend, Chem. Biol. 2000, 7, 211.
[6] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, Chem. Biol. 1999, 6, 493.
[7] A. M. Gulick, ACS Chem. Biol. 2009, 4, 811.
[8] M. Winn, J. K. Fyans, Y. Zhuo, J. Micklefield, Nat. Prod. Rep. 2016, 33, 317.
[9] P. J. Belshaw, C. T. Walsh, T. Stachelhaus, Science 1999, 284, 486.
[10] U. Linne, M. A. Marahiel, Biochemistry 2000, 39, 10439.
[11] B. Kokona, E. S. Winezet, A. N. von Kruenstern, M. J. Cryle, R. Fairman, L. K. Charkoudian, Anal. Biochem. 2016, 495, 42.
[12] K. Haslinger, C. Bricke, S. Uhlmann, L. Sieverling, R. D. Süssmuth, M. J. Cryle, Angew. Chem. Int. Ed. 2014, 53, 8518; Angew. Chem. 2014, 126, 8658.
[13] S. Uhlmann, R. D. Süssmuth, M. J. Cryle, ACS Chem. Biol. 2013, 8, 2586.
[14] K. Haslinger, M. Peschke, C. Bricke, E. Maximowitsch, M. J. Cryle, Nature 2015, 521, 105.
[15] M. Peschke, K. Haslinger, C. Bricke, J. Reinstein, M. J. Cryle, J. Am. Chem. Soc., 2016, 138, 6746–6753.
[16] N. M. Gaudelli, D. H. Long, C. A. Townsend, Nature 2015, 520, 383.
[17] N. M. Gaudelli, C. A. Townsend, Nat. Chem. Biol. 2014, 10, 251.
[18] L. Ray, K. Yamanaaka, B. S. Moore, Angew. Chem. Int. Ed. 2016, 55, 364; Angew. Chem. 2016, 128, 372.
[19] P. Tufar, S. Rahighi, F. I. Kraas, D. K. Kirchner, F. N. Lohr, E. Henrich, J. Kopke, I. Dikic, P. Giuntert, M. A. Marahiel, V. Dotsch, Chem. Biol. 2014, 21, 552.
[20] J. R. Lohman, M. Ma, M. E. Cuff, L. Bigelow, J. Bearden, G. Babnigg, A. Joachimiak, G. N. Phillips, Jr., B. Shen, Proteins 2014, 82, 1210.
[21] K. Haslinger, C. Redfield, M. J. Cryle, Proteins 2015, 83, 711.
[22] S. Zimmermann, S. Pfennig, P. Neumann, H. Yonus, U. Weininger, M. Kevermann, J. Balbach, M. T. Stubbs, FEBS Lett. 2015, 589, 2283.
[23] A. C. Goodrich, B. J. Harden, D. P. Fruch, J. Am. Chem. Soc. 2015, 137, 12100.
[24] Y. Liu, T. Zheng, S. D. Bruner, Chem. Biol. 2011, 18, 1482.
[25] M. J. Jaremko, D. J. Lee, S. J. Opella, M. D. Burkt, J. Am. Chem. Soc. 2015, 137, 11546.
[26] E. J. Drake, B. R. Miller, C. Shi, J. T. Tarrasch, J. A. Sundlow, C. L. Allen, G. Skiniotis, C. C. Aldrich, A. M. Gulick, Nature 2016, 529, 235.
[27] J. M. Reimer, M. N. Aloise, P. M. Harrison, T. M. Schmeing, Nature 2016, 529, 239.
[28] A. Koglin, M. R. Mofid, F. N. Lohr, B. Schafer, V. V. Rogov, M. Blum, T. Mittag, M. A. Marahiel, F. N. Bernhard, V. Dotsch, Science 2006, 312, 273.
[29] J. R. Lai, A. Koglin, C. T. Walsh, Biochemistry 2006, 45, 14869.
[30] J. A. Sundlow, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, Chem. Biol. 2012, 19, 188.
[31] J. Crosby, M. P. Crump, Nat. Prod. Rep. 2012, 29, 1111.
[32] J. Beld, H. Cang, M. D. Burkt, Angew. Chem. Int. Ed. 2014, 53, 14456; Angew. Chem. 2014, 126, 14684.
[33] S. E. Evans, C. Williams, C. J. Arthur, E. Ploskon, P. Wattanamorn, R. J. Cox, J. Crosby, C. L. Willis, T. J. Simpson, M. P. Crump, J. Mol. Biol. 2009, 389, 511.
[34] A. C. Goodrich, D. P. Fruch, Biochemistry 2015, 54, 1154.
[35] D. P. Frueh, H. Arthanari, A. Koglin, D. A. Vosburg, A. E. Bennett, C. T. Walsh, G. Wagner, Nature 2008, 454, 903.
[36] C. A. Mitchell, C. Shi, C. C. Aldrich, A. M. Gulick, Biochemistry 2012, 51, 3252.
[37] A. Tanovic, S. A. Samel, L. Essen, M. A. Marahiel, Science 2008, 321, 659.
[38] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, EMBO J. 1997, 16, 4174.
[39] H. Yonus, P. Neumann, S. Zimmermann, J. J. May, M. A. Marahiel, M. T. Stubbs, J. Biol. Chem. 2008, 283, 32484.
[40] B. R. Miller, J. A. Sundlow, E. J. Drake, T. A. Makin, A. M. Gulick, Proteins 2014, 82, 2691.
[41] T. Kitti, F. Schoppet, M. J. Cryle, ChemBioChem 2016, 17, 576.
[42] J. C. Henderson, C. D. Fage, J. R. Cannon, J. S. Brodbelt, A. T. Keatinge-Clay, M. S. Trent, ACS Chem. Biol. 2014, 9, 2382.
[43] K. I. Bloudoff, D. Rodionov, T. M. Schmeing, J. Mol. Biol. 2013, 425, 3137.
[44] K. I. Bloudoff, D. A. Alonzo, T. M. Schmeing, Cell Chem. Biol. 2016, 23, 331.
[45] M. C. Eguru, C. Milne, M. Borg, F. Flett, C. P. Smith, J. Mickelfield, J. Am. Chem. Soc. 2004, 126, 5032.
[46] J. Thirlway, R. Lewis, L. Nunns, M. Al Nakeeb, M. Styles, A.-W. Struck, C. P. Smith, J. Mickelfield, Angew. Chem. Int. Ed. 2012, 51, 7181; Angew. Chem. 2012, 124, 7293.
[47] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L.-O. Essen, Structure 2007, 15, 781.
[48] M. A. Marahiel, Nat. Prod. Rep. 2016, 33, 136.

Received: March 15, 2016
Published online: July 20, 2016