Piecing Together Nonribosomal Peptide Synthesis

Janice M. Reimer[a,*], Asfarul S. Haque[a,*], Michael J. Tarry[a,*], and T. Martin Schmeing[a,b]

[a]Department of Biochemistry, McGill University, Montréal, QC H3G 0B1, Canada

[b]E-mail: martin.schmeing@mcgill.ca

[*]Equal contribution

Abstract

Nonribosomal peptide synthetases (NRPSs) produce peptide products with wide-ranging biological activities. NRPSs are macromolecular machines with modular assembly-line logic, a complex catalytic cycle, moving parts and multiple active sites. They are organized into repeating sets of domains, called modules. Each module contains all functionality to introduce a building block into the growing peptide, many also perform co-synthetic tailoring. Structures of individual domains have provided insights into their catalytic mechanisms, but with one exception, larger NRPS proteins were refractory to structure determination. Recently, structure determination succeeded for four multi-domain NRPS proteins: an alternative formylating initiation and two termination modules as well as a large cross-module construct. This review highlights how these data, together with novel didomain structures, contribute to a holistic view of the architecture, domain-domain interactions and conformational changes in NRPS megaenzymes.

Introduction

Nonribosomal peptide synthetases (NRPSs) are a family of microbial megaenzymes that produce natural products that are useful to society as therapeutics (antibiotics, antivirals, antitumours, and immunosuppressants) and green chemicals (agricultural agents, emulsifiers, siderophores, and research tools) (Figure 1a) [3]. NRPSs typically synthesize their products through amide bond formation between aminoacyl (or other acyl) monomers. Their architecture is dissimilar to the more famous peptide maker. Whereas ribosomes use the same active sites for each amino acid added to the ribosomal peptide, NRPSs typically employ a dedicated set of enzyme domains for each amino acid added to the nonribosomal peptide. This set of domains is termed a module, and the synthetic strategy dictates that, normally, the number and specificity of the modules correspond to the length and sequence of amino acids in the peptide product. NRPSs can consist of a single polypeptide of between 2 and 18 modules, with a mass of ~220 kDa – 2.2 MDa, or be split over multiple proteins that assemble non-covalently (Figure 1)[4-7].

Within a module, the domains work together to incorporate the incoming amino acid into the growing peptide (Figure 1b) [5,8]. A basic elongation module contains a condensation (C) domain, an adenylation (A) domain and a peptidyl carrier protein (PCP) domain. The A domain selects and adenylates the cognate amino acid, then attaches it by a thioester link to a prosthetic phosphopantetheinyl (PPE) group on the PCP domain. The PCP domain then transports the amino acid to the C domain, which catalyzes amide bond formation between this amino acid and the peptide attached to the PCP domain of the preceding module, elongating the peptide by a single residue. Next, the PCP domain brings the elongated peptide to the downstream module, where it is passed off and further elongated in the next condensation reaction. Once a PCP domain has donated its peptide, it can accept a new amino acid from the A domain and participate in the next cycle of assembly-line synthesis. Initiation modules lack the C domain, and termination modules usually contain a thioesterase (Te) domain, which releases the peptide by cyclization or hydrolysis. A canonical organization of a basic NRPS is A-PCP-(C-A-PCP)$_n$-Te (Figure 1c). Additionally, NRPS modules very often have tailoring domains, including oxidase, reductase, epimerisation, ketoreductase, aminotransferase and methyltransferase domains, and the action of these domains is incorporated into the catalytic cycle of the module [9]. NPRSs can alternatively end in a reductase [10] or terminal C domain [11,12]. This wide range of tailoring domains, combined with the over five hundred monomers that can be used as substrates, including D-amino acids, aryl acids, hydroxy acids, and fatty acids, allows nonribosomal peptides to occupy a diverse area of chemical space [13].

Starting with Conti's determination of an A domain in 1997[14], each core NRPS domain has been structurally characterized [5,15]: The A domain has a large A_{core} (~450 amino acids) portion with binding sites for ATP and substrate amino acid, and a small A_{sub} (~100 amino acids) portion that changes position depending on functional state [14,16-20]. (A_{core} and A_{sub} are also called the large/N-terminal and small/C-terminal subdomains [21].) The C domain is a ~450 amino acid V-shaped pseudodimer of chloramphenicol acetyltransferase folds, with an active site at the middle of a tunnel connecting binding sites for donor and acceptor PCP domains. The PCP domain is an oblong 4-helix bundle of ~80 amino acids, homologous to FAS and PKS acyl carrier protein domains, with the PPE attached to a conserved serine at one end. The Te domain is a ~275 amino acid α/β hydrolase domain with an active site topped by a variable "lid" region. Studies investigating the catalytic mechanisms of each individual reaction in the NRPS cycle are reviewed elsewhere [5,15,16,22,23]. Furthermore, some tailoring domains have been structurally characterized, several didomain structures have been determined, and a landmark study in 2008 from Marahiel's lab produced a snapshot of an entire termination module (reviewed in [5]). However, a most intriguing aspect of NRPS function is how these domains transiently and productively interact during the catalytic cycle and how domains are built up into modules and full NRPS megaenzymes [8]. Multiple structures of large constructs of NRPSs in recognizable functional states are required to understand these megaenzymes. Over the last ~2 years, several studies, including on full modules and module-sized constructs of NRPSs, have started to reveal the catalytic cycle in the context of the large macromolecular machine. In this review, we follow a hypothetical nascent peptide from initiation to termination, with special emphasis on insight gained from these recently-determined structures of large NRPS proteins.

Start it off

Minimal initiation modules contain only A and PCP domains. However, it is not uncommon for NRPSs to start with more complex initiation modules. Lipopeptides synthetases for peptide like daptomycin contain starter C domains that use acyl-PCPs as donor substrates; depsipeptide synthetases for valinomycin and cereulide contain ketoreductase domains; and kolossin A (Figure 1) and linear gramicidin synthetase contain formylation (F) domains. Recently, we determined a series of crystal structures of the initiation module of linear gramicidin synthetase (LgrA-D) [24] which visualized the synthetic cycle for this initiation-tailoring module (Figure 2a) [25,26].

The LgrA initiation module structures are the first visualization of a cis-acting tailoring domain embedded into the NRPS architecture (Figure 2a). The F domain is recognizable by its homology to free-standing formyl-transferase (FT) proteins. In LgrA, the C-terminus of the F domain is fused to the

N-terminus of the A domain, with little disruption of the conformation of either domain. Only ~8 residues are not recognizable as clearly belonging to F or A domain, and they form a small helix-turn, meaning there is no extended linker between F and A domains. Accordingly, a fairly extensive interface is formed the between F and A domains, burying 831 $Å^2$ of surface area per domain and placing the two domains in an extended conformation. This extended conformation is seen in ~6 different crystal packing environments and by small angle X-ray scattering, and places the formylation and adenylation active sites, locations for successive steps in the catalytic cycle, ~50 Å apart from one another. Fortunately, the structures capture each major conformation.

The initiation cycle starts when the A domain in the "open" conformation (Figure 2a-i) binds amino acid and ATP, and the A_{sub} rotates 30° to close and catalyze amino acid adenylation (Figure 2a-ii) [16]. This reaction triggers the 140° rotation of the A_{sub} to allow PCP domain binding and transfer of the amino acid to the PPE arm (Figure 2a-iii) [20,27]. In LgrA crystals of the open and closed states, the PCP domain was present but disordered, in accordance with it not being involved until the thiolation sate, which we trapped using an aminoacyl-aminopantetheine mimic of the aminoacyl-PPE. Previous A and A-PCP structures had visualized these states, but here all states were observed within the same initiation module [28,29]. For the next step, formylation [30], an impressive conformational change must occur: the PCP domain rotates 75° and translates 61 Å to deliver the aminoacyl-PPE to the F domain [26]. The A_{sub} domain plays several roles in this transition: First, the linker between PCP and A_{sub} is not long enough to allow this movement, so the A_{sub} translocates 21 Å and rotates 180° to act as an extension of the linker (or "linker domain" [8]). Second, the A_{sub} helps to orient the PCP domain for binding to the F domain, which appears required because the PCP-F domain binding surface is very small (~219 $Å^2$). Aided by the A_{sub}, the PCP positions the amino nitrogen of the aminoacyl-PPE into the F domain active site to accept a formyl group from the formyltetrahydrofolate cofactor. This formylation event completes the set of reactions that occur in the initiation module, and produces the substrate for the first reaction in an elongation module, the formyl-aminoacyl-PCP. We and others are working to visualize that step in dimodular NRPSs, but for now excised didomains can provide insight into the transition to elongation.

Pass it on

The (formyl-)aminoacyl-PCP moves from the initiation module to the donor site of the C domain. This donor site was identified by apo structures of the C domain and biochemically [31,32]. The structure of an excised PCP-C didomain from TycC visualized the two domains required for

donation, but was in a non-productive conformation [33]. However, recently two didomain structures of PCP and C domain homologues, a terminal cyclizing C (C_T) [11] and an epimerization (E) domain [34] produced the first glimpses of this donation interaction (Figure 2b). The C_T domain catalyzes internal cyclization and release of the peptide from peptidyl-PCP and the E domain catalyzes epimerization of the first amino acid residue of peptidyl-PCP. Both the PCP-C_T structure from TqaA and the PCP-E structure from GrsA contain PCP domains in the broad and somewhat shallow donor site, and visualize PPE arms lining the C domain tunnel toward the active sites. However, the two conformations of the PCP domains are rotated ~30° to one another and make contact with opposite faces of the donor site depression. The differences between the E and C_T domains and bona fide C domains are most pronounced in the area of the acceptor PCP binding site, as this is blocked by domain-specific sequences in E and C_T domains. There are also differences near the donor site relative to the C domain, so it is not yet clear whether the PCP bound to a canonical C domain would sample both the observed positions or assume a new orientation, or whether the rest of the upstream and downstream modules would influence the PCP:C interaction. However, these structures certainly do represent the broad strokes of PCP delivery to the donor site of a C domain in an elongation module.

Keep it going

The first view of any module was that of the termination module of surfactin synthetase, SrfA-C, solved in the peptide-accepting state almost a decade ago[35]. With a domain architecture of C-A-PCP-Te, it represents both a minimal C-A-PCP elongation domain and the most common type of bacterial termination module. The SrfA-C structure showed very large distances between active sites in NRPSs indicating that substantial conformations changes would occur in the catalytic cycle (Figure 3a-i). Another key finding was that the C-terminal lobe of the C domain and the A_{core} form a "catalytic platform", burying a large (765 Å^2) area of surface. This interaction defines the overall rectangular shape of an elongation module, was by far the most extensive interdomain contact seen in the module, and was proposed to be a fixed interface.

Our knowledge of structures of elongation / termination modules was greatly enhanced with the recent determination of two more C-A-PCP-Te termination modules, of EntF (involved in enterobactin production) and AB3403 (from an uncharacterized pathway) (Figure 3a) [36,37]. Firstly, there was important insight about the C:A interface: Despite similar interaction surfaces donated by the C domain and the A_{core} in SrfAC, EntF and AB3403, a variation of ~20° is propagated over the C domain, and combined with differences in conformation of the C domain lobes, analogous atoms in the far side of

the C domains assume positions >30 Å apart. This is not solely due to differences between synthetases, as two separate structures of the EntF module show somewhat shifted interfaces that cause a ~15 Å difference in position of the N terminus of the C domain. Thus, the C:A catalytic platform is more plastic than first thought. However, even with these movements, the C:A interface is by far the most constant of any between two canonical domains, and defines an NRPS module structurally.

The SrfA-C, EntF and AB3403 crystallography also provides views of catalytically-relevant states, allowing us to continue following the synthesis of our hypothetical nascent peptide [36,37]. As is the case in initiation modules, the elongation module starts its cycle with the A domain binding amino acid and ATP, and then adenylation. AB3403 shows the adenylation state of this module, with substrate analogues bound and the A_{sub} domain in the closed conformation (Figure 3a-ii). The next stage, thiolation, is captured in the structure of EntF stalled by a mechanism-based aminoacyl-adenosine-vinylsulfonamide inhibitor [38] (Figure 3a-iii). The aminoacyl-PCP then travels ~45 Å, rotating ~75°, to bind the acceptor site of the C domain (Figure 3a-iv) and elongate the nascent peptide in the condensation reaction. Both SrfA-C and AB3403 were visualized with PCP domain bound at the acceptor site of the C domain, but key differences do exist, highlighting the advantages of obtaining multiple similar structures (compare Figure 3a-i and 3a-ii). Firstly, the PCP domain of SrfA-C is unmodified, but in AB3403, the PPE arm was seen making specific hydrogen bonds with the side of the C domain tunnel. Furthermore, the PCP domain of AB3403 is rotated 30° relative to that of SrfA-C. The PCP domain of SrfA-C is unable to take the position seen in AB3403 because it would overlap with the A_{sub} domain and a loop of the C domain. One or the other position could be influenced by crystal packing, but the acceptor site of the C domain is even more shallow than the donor site, and it is likely that particular PCP-C domain pairs have their own preferred binding orientations, and that even within a particular PCP-C pair, multiple orientations should allow productive substrate delivery to the catalytic center of the C domain. This delivery permits the aminoacyl-PCP to accept the peptidyl group from the donor, transforming it to a new, elongated peptidyl-PCP. In an elongation module, this peptidyl-PCP would travel to the next module's C domain donor site, as described above for the initiation module. Interestingly, the AB3403 structure shows that the first stages of the next cycle need not wait for condensation, as the adenylation state is observed simultaneously with the peptide acceptance state [36]. Accordingly, Figure 3a-ii and iv are the same structure.

Finish him!

This elongation cycle (or an expanded cycle with a tailoring step) occurs in every canonical elongation module along the NRPS assembly line, until the termination module. Here, after one more elongation cycle, the PCP domain delivers the elongated peptide to the chain-terminating domain. In fungal termination modules, that is often a C_T domain, and the PCP-C_T structure is as discussed above, but for bacterial NRPSs, it is most commonly a Te domain and second most commonly a terminal reductase (R) domain. SrfA-C, EntF and AB3403 all have domains C-A-PCP-Te, yet their structures show markedly different Te domain positions. (Compare the Te domain in Figure 3a-i, ii and iii.) In SrfA-C, the Te domain floats near the acceptor side of the C domain and makes almost no contact with the other domains (Figure 3a-i). The Te active site is facing away from the rest of the module and is partially blocked by the C domain, so both the Te domain and the PCP domain have to move for productive binding. Te domain movement seems simplistic however, as in AB3403, the Te domain is displaced by >50 Å and only contacts the "back face" of the PCP domain (another contact that would need to break for termination) (Figure 3a-ii). Furthermore, in EntF the Te domain is >80 Å away from either of these positions, making modest contacts with the A domain (Figure 3a-iii). Electron microscopy also shows this domain to be very mobile [36,39]. Thus, the Te domain is loosely tethered to the termination module via the mobile PCP domain.

Text Box: In between domains

Unlike many other macromolecular machines with moving parts, NRPSs do not have a power stroke: No NTP hydrolysis drives PCP domain movement. Rather, it is likely the PCP domain moves largely randomly through tethered Brownian motion. The interdomain linkers are of variable length across NRPSs, but combined with other domain movements, provide enough freedom for the PCP domain to travel the long distances between active sites. Though the PCP domain likely samples binding to all its partners, the acyl moiety on the PPE arm should increase affinity for the appropriate partner. The presence of the various partner domains and the acyl moiety could aid progression to some extent [1,2], but it is the unidirectionality of the condensation reaction that dictates the direction of synthesis. The lack of power stroke, high degree of flexibility and largely undirected tethered Brownian motion result in a rate of NRPS synthesis of peptides ~3 orders of magnitude slower than that of the ribosome. Nonetheless, NRPSs appears to be fast enough, and the products important enough, for the host microbes to keep their massive genes in their genome.

None of the structures of the termination modules report on the release step in NRPS synthesis, but once again, didomain structures provide insight. The excised EntF didomain PCP-Te was determined by NMR without PPE modification [40] and by X-ray

crystallography with an alpha-chloro-aminoacyl-PPE designed to crosslink to the Te domain active site serine [41]. The PCP binds the edge of a crevice formed by the main core of the Te domain and the mobile lid segment, and the PPE extends into the catalytic center. The domain-domain interface buries 745 Å of surface area, which is more than PCP binding to other domains. The configuration allows the Te domain active site serine to attack and accept the nascent peptide in the thioesterase first half-reaction. The second half-reaction can be oligomerization, for which a presumably similar PCP-Te interaction occurs, or hydrolysis or cyclization, for which the PCP domain presumably departs.

A recent structure of the PCP-R didomain from the NRPS-like carboxylic acid reductase provides insight into release in terminally reducing NRPSs [10]. The PCP is again attached to the rest of the protein by flexible linkers. The PCP:R interface buries a modest 473 $Å^2$ of surface area. The PPE runs along the interface between the main body of the domain, the substrate-binding subdomain, and NADPH, and may help organize the active site for reduction. Despite the absence of peptide and the disorder of the nicotinamide moiety, the observed PCP:R interaction appears representative of the conformation NRPSs with terminal R domains use to reduce the peptidyl-PCP and release peptide as an aldehyde product.

Bring it all together

The above-described structures, in context of excellent existing functional studies, provide a wealth of insight into each step of the NRPS cycle. They are less informative about how modules form intact NRPS megaenzymes. High resolution structures of multimodular NRPSs are an outstanding goal in the field, and multiple such structures are required to answer questions of NRPS architecture, but a view of higher order structure is beginning to form.

We recently solved the structure of a cross-domain construct of bacillibactin synthetase protein DhbF (Figure 4a) [39]. The A1-PCP1-C2 construct contained the A and PCP domains from the first module and the C domain from the second module. In this structure, the vinylsulfonamide inhibitor is present in the A domain active site and attached to the PCP domain, but the A_{sub} and PCP domain are somewhat shifted from a true thiolation conformation, likely due to crystal contacts. This structure and the second structure of the EntF module (Figure 3a-iii) both contain MbtH-like proteins (MLPs) complexed to the A_{core} in the same position as in a fused A-MLP protein [42]. MLPs are necessary partners to some A domains, promoting activity in a manner that does not alter the average structure of the A domain [37].

A structure with the last large domain from module 1 and the first large domain from module 2 should allow visualization of intermodule interactions not formed by the PCP domain (which is

unlikely to contribute to consistent higher-order structure because of its mobility). However, the structure showed no contact between the A1 and C2 domains. Instead, the only non-covalent interaction between module 1 and 2 was the back face of the PCP domain contacting the donor site of the C domain. A similar interaction was observed in TycC PCP-C didomain structure and is echoed by the back face of the PCP domain interacting with the Te domain in AB4304. The lack of intermodular A:C contacts is in stark contrast to the extensive C:A contacts within a module and implies that DhbF may not assume a single module–module conformation. This was confirmed by negative stain EM of dimodular C1-A1-PCP1-C2-A2-PCP2-Te2 and C1-A1-PCP1-C2-A2-PCP2 DhbF (Figure 4b). Despite stalling with vinylsulfonamide inhibitors, DhbF was in multiple conformations. Five separate envelopes were reconstructed from this data and each envelope could be fitted with two models of C-A-PCP modules, as expected. However, there were large differences in the relative orientations of the two modules and no consistent module–module interface was observed, strongly suggesting that for DhbF at least, no regular, repeating supermodular architecture is present. This seems to also be the case for cyclosporin synthetase, a 12-module NRPS, micrographs of which show it to adopt either a "ball of balls" or uneven "balls on a string" morphology [43]. It remains to be seen whether lack of higher-order architecture is a general theme for NRPSs, or whether elegant and attractive regular architecture, which can be modelled based on single conformational states, can occur in some NRPSs (Figure 4c)[44]. High-quality structures of NRPSs of two or more modules, towards which we and others are working, will start to answer these questions.

Onward

Recent work has illuminated many aspects of NRPSs structure. The next several years will likely be equally insightful. It will be interesting to see whether the coming structures of multimodular NRPSs will be determined by crystallography or cryo-EM. Cryo-EM is now the technique of choice for large macromolecules but the moderate size of NRPS domains and especially the massive array of domain-domain and module-module conformations which are absolutely required for NRPS function (many of which remain even after targeted crosslinking), make NRPSs extremely challenging targets. A sample with a near continuum of conformations is problematic, even for the "new EM" [45], but will eventually be conquered. Chemical biology tools were key to obtaining structures, and new chemical biology tools will likely be needed access novel structures [15,46-48].

Structures can also greatly inform non-canonical NRPS systems, including those that perform optional module skipping and iteration. All these static snapshots will be brought to life with experiments that report on movement more directly, including fluorescence approaches [49], NMR

experiments [1,50,51] and molecular dynamics simulations[52]. Altogether, the work reviewed here and ongoing research provides an increasingly excellent view of these megaenzymes, which are among nature's most elegant and fascinating macromolecular machines.

References

1. Goodrich AC, Harden BJ, Frueh DP: **Solution structure of a nonribosomal peptide synthetase carrier protein loaded with its substrate reveals transient, well-defined contacts.** *J Am Chem Soc* (2015) **137**(37):12100-12109.

2. Linne U, Marahiel MA: **Control of directionality in nonribosomal peptide synthesis: Role of the condensation domain in preventing misinitiation and timing of epimerization.** *Biochemistry* (2000) **39**(34):10439-10447.

3. Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD, Thomas MG: **Nonribosomal peptide synthetases involved in the production of medically relevant natural products.** *Mol Pharm* (2008) **5**(2):191-211.

4. Walsh CT: **Polyketide and nonribosomal peptide antibiotics: Modularity and versatility.** *Science* (2004) **303**(5665):1805-1810.

5. Weissman KJ: **The structural biology of biosynthetic megaenzymes.** *Nat Chem Biol* (2015) **11**(9):660-670.

6. Schwarzer D, Finking R, Marahiel MA: **Nonribosomal peptides: From genes to products.** *Nat Prod Rep* (2003) **20**(3):275-287.

7. Dehling E, Volkmann G, Matern JC, Dorner W, Alfermann J, Diecker J, Mootz HD: **Mapping of the communication-mediating interface in nonribosomal peptide synthetases using a genetically encoded photocrosslinker supports an upside-down helix-hand motif.** *J Mol Biol* (2016) **428**(21):4345-4360.

8. Gulick AM: **Structural insight into the necessary conformational changes of modular nonribosomal peptide synthetases.** *Curr Opin Chem Biol* (2016) **35**:89-96.

9. Sundaram S, Hertweck C: **On-line enzymatic tailoring of polyketides and peptides in thiotemplate systems.** *Curr Opin Chem Biol* (2016) **31**:82-94.

10. Gahloth D, Dunstan MS, Quaglia D, Klumbys E, Lockhart-Cairns MP, Hill AM, Derrington SR, Scrutton NS, Turner NJ, Leys D: **Structures of carboxylic acid reductase reveal domain dynamics underlying catalysis.** *Nat Chem Biol* (2017) **13**(9):975-981.

•A series of mono- and didomain structures of a carboxylic acid reductase (CAR) which has domains A-PCP-R, shows high mobility. The release of product as an aldehyde, not an alcohol as in some R domains, may be because of active site rearrangement upon PPE disengagement.

11. Zhang J, Liu N, Cacho RA, Gong Z, Liu Z, Qin W, Tang C, Tang Y, Zhou J: **Structural basis of nonribosomal peptide macrocyclization in fungi.** *Nat Chem Biol* (2016) **12**(12):1001-1003.

•The crystal structure of the terminal PCP-C_T didomain of TqaA provides insight into the protein-protein interactions with facilitate C_T domain-catalyzed termination by intramolecular cyclization, which is common in fungal nonribosomal peptide synthesis.

12. Bloudoff K, Fage CD, Marahiel MA, Schmeing TM: **Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis.** *Proc Natl Acad Sci U S A* (2017) **114**(1):95-100.

13. Caboche S, Pupin M, Leclere V, Fontaine A, Jacques P, Kucherov G: **NORINE: A database of nonribosomal peptides.** *Nucleic Acids Res* (2008) **36**(Database issue):D326-331.

14. Conti E, Stachelhaus T, Marahiel MA, Brick P: **Structural basis for the activation of phenylalanine in the non-ribosomal biosynthesis of gramicidin S.** *EMBO J* (1997) **16**(14):4174-4183.

15. Miller BR, Gulick AM: **Structural biology of nonribosomal peptide synthetases.** *Methods Mol Biol* (2016) **1401**:3-29.

16. Gulick AM: **Conformational dynamics in the acyl-coa synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase.** *ACS Chem Biol* (2009) **4**(10):811-827.

17. May JJ, Kessler N, Marahiel MA, Stubbs MT: **Crystal structure of DhbE, an archetype for aryl acid activating domains of modular nonribosomal peptide synthetases.** *Proc Natl Acad Sci U S A* (2002) **99**(19):12120-12125.

18. Gulick AM, Starai VJ, Horswill AR, Homick KM, Escalante-Semerena JC: **The 1.75 A crystal structure of acetyl-coa synthetase bound to adenosine-5'-propylphosphate and coenzyme a.** *Biochemistry* (2003) **42**(10):2866-2873.

19. Du L, He Y, Luo Y: **Crystal structure and enantiomer selection by D-alanyl carrier protein ligase DltA from Bacillus cereus.** *Biochemistry* (2008) **47**(44):11473-11480.

20. Yonus H, Neumann P, Zimmermann S, May JJ, Marahiel MA, Stubbs MT: **Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains.** *J Biol Chem* (2008) **283**(47):32484-32491.

21. Gulick AM: **Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens.** *Nat Prod Rep* (2017) **34**(8):981-1009.

22.   Payne JA, Schoppet M, Hansen MH, Cryle MJ: **Diversity of nature's assembly lines - recent discoveries in non-ribosomal peptide synthesis.** *Mol Biosyst* (2016) **13**(1):9-22.

23.   Bloudoff K, Schmeing TM: **Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: Discovery, dissection and diversity.** *Biochimica et Biophysica Acta* (2017) **1865**(11 Pt B):1587-1604.

24.   Kessler N, Schuhmann H, Morneweg S, Linne U, Marahiel MA: **The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains.** *J Biol Chem* (2004) **279**(9):7413-7419.

25.   Reimer JM, Aloise MN, Harrison PM, Schmeing TM: **Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase.** *Nature* (2016) **529**(7585):239-242.
••Every major catalytic step in the initiation process is visualized in a series of crystal structures for the linear gramicidin synthetase. The structures emphasize the substantial conformational changes that occur between sequential steps in the synthetic cycle.

26.   Reimer JM, Aloise MN, Powell HR, Schmeing TM: **Manipulation of an existing crystal form unexpectedly results in interwoven packing networks with pseudo-translational symmetry.** *Acta Crystallographica Section D* (2016) **72**(10):1130-1136.

27.   Reger AS, Wu R, Dunaway-Mariano D, Gulick AM: **Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase.** *Biochemistry* (2008) **47**(31):8016-8025.

28.   Sundlov JA, Shi C, Wilson DJ, Aldrich CC, Gulick AM: **Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains.** *Chem Biol* (2012) **19**(2):188-198.

29.   Mitchell CA, Shi C, Aldrich CC, Gulick AM: **Structure of PA1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains.** *Biochemistry* (2012) **51**(15):3252-3263.

30.   Schoenafinger G, Schracke N, Linne U, Marahiel MA: **Formylation domain: An essential modifying enzyme for the nonribosomal biosynthesis of linear gramicidin.** *J Am Chem Soc* (2006) **128**(23):7406-7407.

31.   Keating TA, Marshall CG, Walsh CT, Keating AE: **The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains.** *Nat Struct Biol* (2002) **9**(7):522-526.

32.   Bloudoff K, Rodionov D, Schmeing TM: **Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases.** *J Mol Biol* (2013) **425**(17):3137-3150.

33.   Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen LO: **Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase.** *Structure* (2007) **15**(7):781-792.

34.   Chen W-H, Li K, Guntaka NS, Bruner SD: **Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases.** *ACS chemical biology* (2016) **11**(8):2293-2303.

•This first structure of a PCP-E didomain structure shows the PCP domain docked at the E domain for substrate delivery. E domains are structurally similar to C domains and the structure is similar what a peptide donation state of a C domain would be.

35.   Tanovic A, Samel SA, Essen LO, Marahiel MA: **Crystal structure of the termination module of a nonribosomal peptide synthetase.** *Science* (2008) **321**(5889):659-663.

36.   Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Allen CL, Skiniotis G, Aldrich CC, Gulick AM: **Structures of two distinct conformations of holo-non-ribosomal peptide synthetases.** *Nature* (2016) **529**(7585):235-238.

••Two structures of full holo termination modules highlight the role of the adenylation domain in guiding the synthetic cycle. The dynamic nature of modules is illustrated by concurrent catalytic conformations and observed flexibility between domains.

37.   Miller BR, Drake EJ, Shi C, Aldrich CC, Gulick AM: **Structures of a nonribosomal peptide synthetase module bound to MbtH-like proteins support a highly dynamic domain architecture.** *J Biol Chem* (2016) **291**(43):22559-22571.

•New structures of the EntF termination module are presented, with MLPs bound and the Te domain ordered. MLP binding does not alter the A domain structure. The location of the Te domain is different to that in other termination module structures, supporting the idea that it is highly mobile.

38.   Qiao C, Wilson DJ, Bennett EM, Aldrich CC: **A mechanism-based aryl carrier protein/thiolation domain affinity probe.** *J Am Chem Soc* (2007) **129**(20):6350-6351.

39.   Tarry MJ, Haque AS, Bui KH, Schmeing TM: **X-ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture.** *Structure* (2017) **25**(5):783-793 e784.

••The crystal structure of an A1-PCP1-C2 construct of DhbF and EM reconstructions of dimodular DhbF suggests DhbF lacks a defined, multimodular architecture. This study presents the first 3D view of a multimodular NRPS, albeit at low-resolution.

40.   Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G: **Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase.** *Nature* (2008) **454**(7206):903-906.

41.   Liu Y, Zheng T, Bruner SD: **Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases.** *Chem Biol* (2011) **18**(11):1482-1488.

42.   Herbst DA, Boll B, Zocher G, Stehle T, Heide L: **Structural basis of the interaction of MbtH-like proteins, putative regulators of nonribosomal peptide biosynthesis, with adenylating enzymes.** *J Biol Chem* (2013) **288**(3):1991-2003.

43.   Hoppert M, Gentzsch C, Schorgendorfer K: **Structure and localization of cyclosporin synthetase, the key enzyme of cyclosporin biosynthesis in tolypocladium inflatum.** *Arch Microbiol* (2001) **176**(4):285-293.

44. Marahiel MA: **A structural model for multimodular NRPS assembly lines.** *Nat Prod Rep* (2016) **33**(2):136-140.

45. Frank J: **Advances in the field of single-particle cryo-electron microscopy over the last decade.** *Nat Protoc* (2017) **12**(2):209-212.

46. Meier JL, Burkart MD: **The chemical biology of modular biosynthetic enzymes.** *Chem Soc Rev* (2009) **38**(7):2012-2045.

47. Tarry MJ, Schmeing TM: **Specific disulfide cross-linking to constrict the mobile carrier domain of nonribosomal peptide synthetases.** *Protein Engineering, Design & Selection* (2015) **28**(6):163-170.

48. Bloudoff K, Alonzo DA, Schmeing TM: **Chemical probes allow structural insight into the condensation reaction of nonribosomal peptide synthetases.** *Cell Chem Biol* (2016) **23**(3):331-339.

49. Alfermann J, Sun X, Mayerthaler F, Morrell TE, Dehling E, Volkmann G, Komatsuzaki T, Yang H, Mootz HD: **FRET monitoring of a nonribosomal peptide synthetase.** *Nat Chem Biol* (2017) **13**(9):1009-1015.

50. Goodrich AC, Meyers DJ, Frueh DP: **Molecular impact of covalent modifications on nonribosomal peptide synthetase carrier protein communication.** *J Biol Chem* (2017) **292**(24):10002-10013.

51. Jaremko MJ, Lee DJ, Opella SJ, Burkart MD: **Structure and substrate sequestration in the pyoluteorin type ii peptidyl carrier protein PltL.** *J Am Chem Soc* (2015) **137**(36):11546-11549.

52. Dowling DP, Kung Y, Croft AK, Taghizadeh K, Kelly WL, Walsh CT, Drennan CL: **Structural elements of an nrps cyclization domain and its intermodule docking domain.** *Proc Natl Acad Sci U S A* (2016) **113**(44):12432-12437.

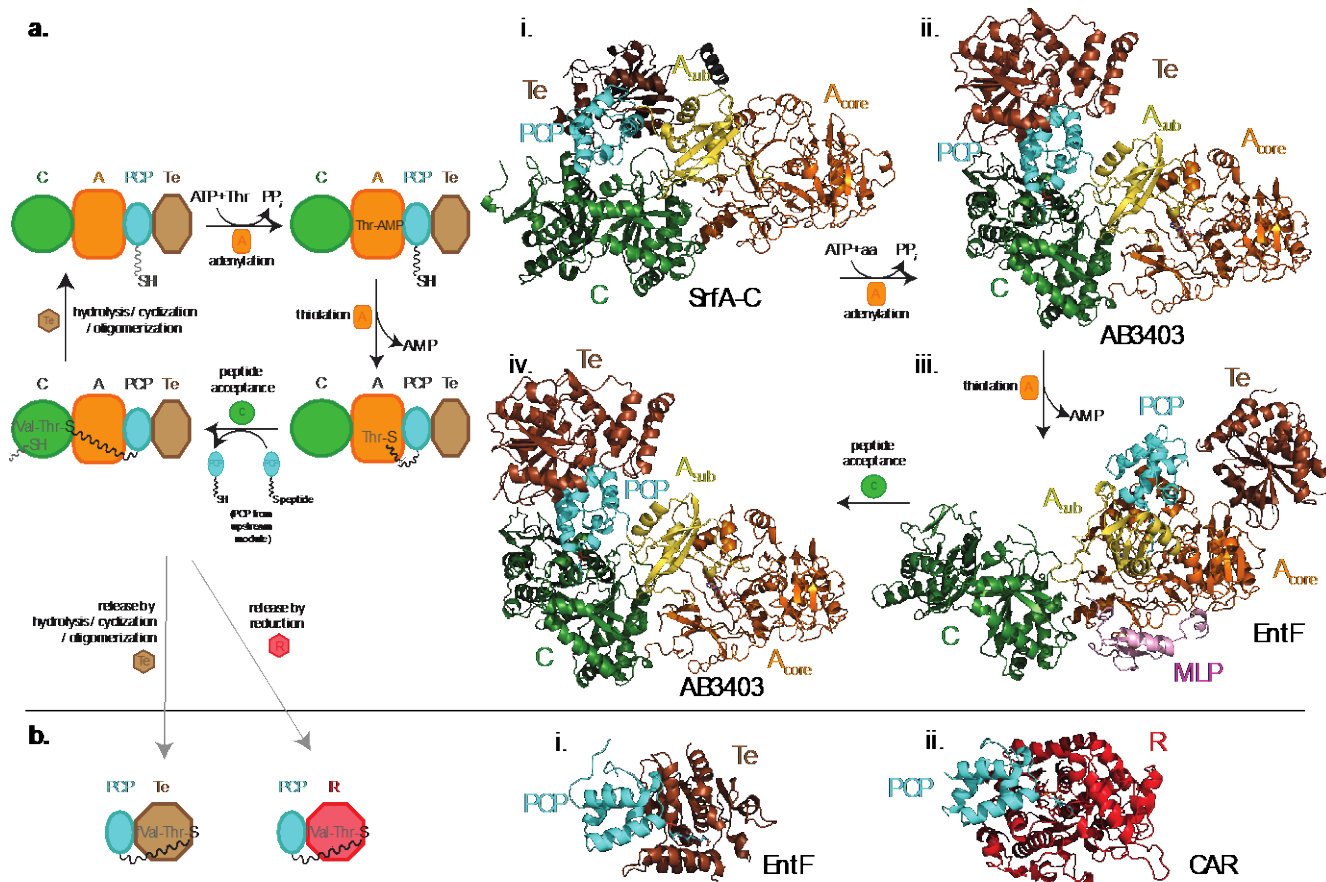**Figure 1. Nonribosomal peptides and peptide synthetases**

A. Some examples of nonribosomal peptides: gramicidin A (topical antibiotic), cyclosporin A (immunosuppressant), enterobactin (siderophore), yersinibactin (siderophore), bacillamide D (anti-algael), daptomycin (antibiotic, trade name Cubicin). B. A schematic diagram of the elongation cycle of a canonical NRPS elongation module. C. A generic synthetase. In this review, we follow the synthesis of a hypothetical formylated dipeptide that would be synthesized by the F-A-PCP-C-A-PCP-Te NRPS depicted (where n=1; note that most NRPSs do not contain an F domain, the tailoring domain which formylates the N-terminal amino acid). The domains included in some important crystal structures discussed in this review are indicated below the schematic. D. Some examples of NRPSs. (ArCP – Aryl carrier protein domain; Cy – heterocyclization domain; E/C – bifunctional epimerization / condensation domain.)
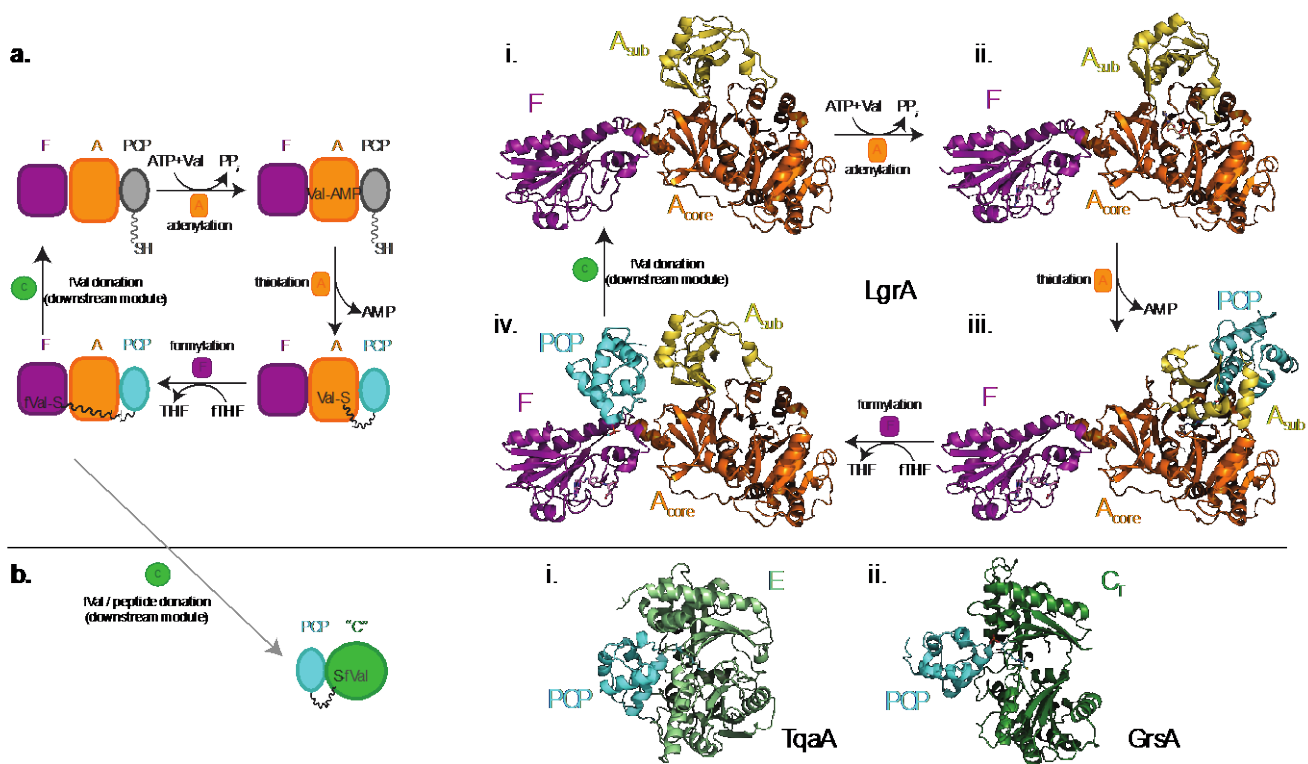
Figure 2. The first stages of nonribosomal peptide synthesis

A. The catalytic cycle of the initiation module of LgrA, in schematic form and as illustrated by crystal structures (i – iv) [25,26]. The F and A_{core} regions are quite static, while the A_{sub} and PCP domain appear in dramatically different positions in the various catalytic states. Disordered portions are shown in grey. The module selects and adenylates valine (i,ii), covalently tethers it to the PCP domain by via a thioester bond (iii), and formylates it (iv). B. The PCP domain transports the formyl-valine (or in the general case, peptide) to the next module and delivers it to the C domain acceptor site. This delivery is represented by didomain structures of PCP domain and C domain homologues, E domain (i) [34] and C_T domain (ii) [11].

Figure 3. The elongation and termination of nonribosomal peptide synthesis

A. The catalytic cycle of the elongation and termination modules in schematic form and as illustrated by crystal structures of SrfA-C, EntF and AB3404 (i – iv) [35-37]. The module selects and adenylates an amino acid, here depicted as threonine, the EntF substrate (i,ii). (The SrfA-C A domain (i) is in a pseudo-open state and the AB3403 is in a canonical closed state). The A domain covalently tethers the amino acid to the PCP domain via a thioester (iii). This aminoacyl-PCP accepts (iv) the peptidyl group from the upstream PCP, elongating the peptide. B. The PCP domain transports the formyl-valine-threonine (or in the general case, peptide) to the termination domain, represented by didomain structures of PCP domain and Te domain (i) [41] or R domain [10].
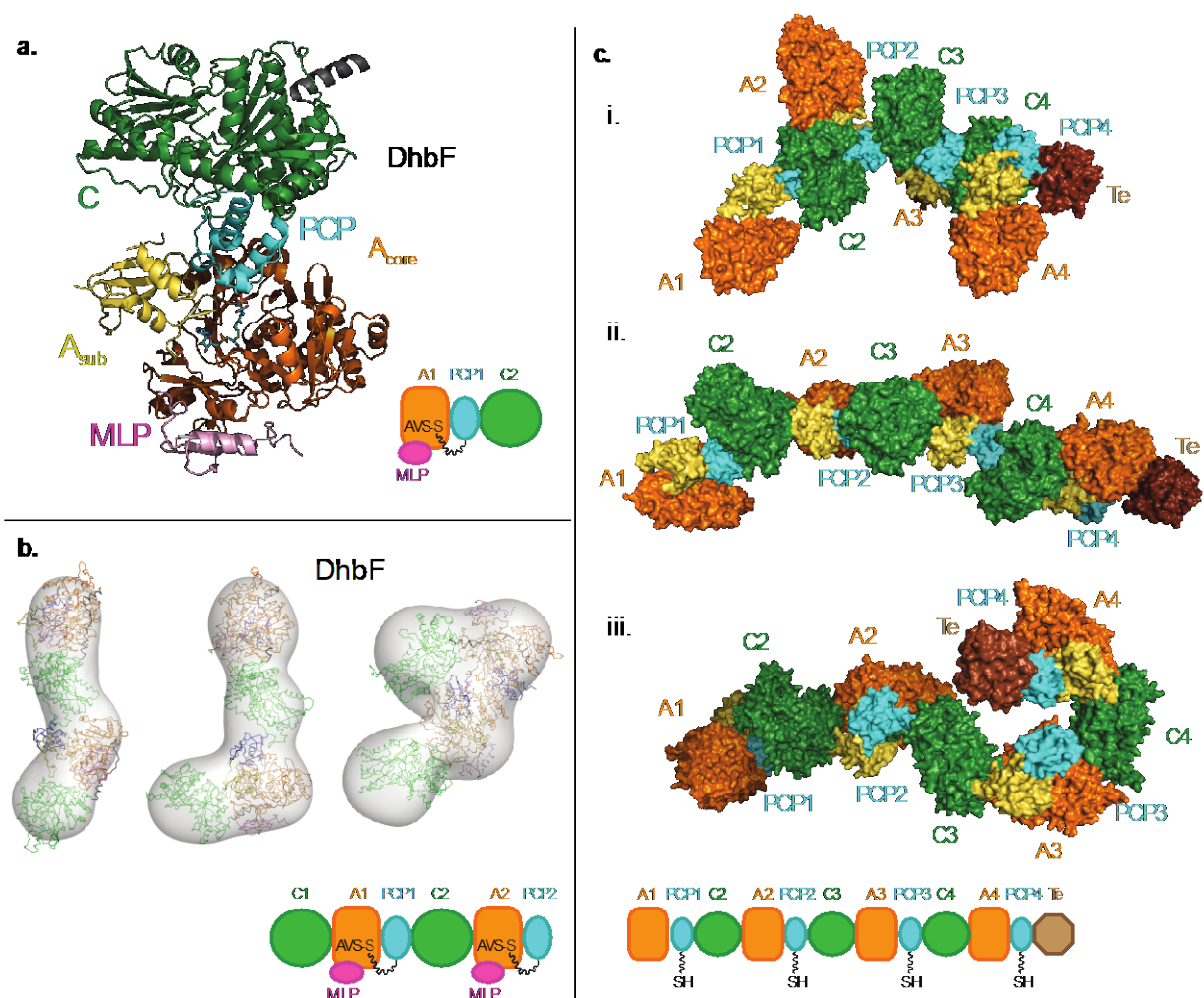
Figure 4. Higher order architecture of nonribosomal peptide synthetases

A. Crystal structure of an A1-PCP1-C2 cross-module construct of DhbF shows no contact between the large domains of adjacent molecules. B. Three different low-resolution envelopes of dimodular DhbF calculated from the same negative stain EM data set [39]. C. Three models of a hypothetical tetramodular NRPS, assembled from X-ray (i, ii) and EM (iii) studies. Both the helical model (i) (recreated from Marahiel [44]) based on serially overlapping TycC PCP-C (PDB 2JGP) [33]and SrfA-C C-A-PCP-Te (PDB 2VSQ) [35] structures, and the shallow curved model (ii) based on serially overlapping EntF C-A-PCP [36,37] (PDB 5JA1) and DhbF A1-PCP1-C2 [39] (PDB 5U89), show regular architecture, but rely on relative conformations known not to be fixed. The model created from serially overlapping dimodules fit to low-resolution negative stain envelopes (iii) [39] is irregular, and is intended to represent one of a near-continuum of irregular structures.

Highlights

1. Nonribosomal peptide synthetase (NRPSs) are modular biosynthetic factories

2. Most functional states of NRPS modules have now been visualized at high resolution

3. Massive conformational changes occur within a module at different stages of catalysis

4. X-ray and EM studies hint at a lack of regular higher order multimodular architecture.