T. Martin Schmeing PhD

Department of Biochemistry, McGill
University, Montréal, QC, Canada

Visualizing A Natural Antibiotic Nanofactory

Abstract

Environmental samples are excellent sources of natural products that possess numerous kinds of therapeutic activities. One important family of natural products is the nonribosomal peptides, which includes penicillin, cyclosporin, viomycin and daptomycin (Cubicin). These peptides are made in bacteria or fungi by large enzymes called nonribosomal peptide synthetases (NRPS). NRPSs are true macromolecular machines or nanofactories, with modular assembly-line logic, a complex catalytic cycle, moving parts and many active sites. Visualization of large fragments of NRPSs at various functional states is required to understand the manner in which NRPSs synthesize their important products. Many excellent structural experiments have been performed to date. Recently, we added to the structural knowledge by visualizing the first module of the NRPS, which makes linear gramicidin, a clinical topical antibiotic, in all its major functional states. These experiments show how the individual domains, including an unusual tailoring domain, function together in assembly-line synthesis. Along with the ever-expanding body of biophysical, biochemical and genetic work, this work brings us closer to a fundamental understanding of these natural antibiotic nanofactories, and perhaps the ability to exploit them to produce novel therapeutics.

Manuscript submitted 24th October, 2016

Manuscript accepted 25th October, 2016

Clin Invest Med 2016; 39 (6): E220-E226.

Correspondence to:

T. Martin Schmeing PhD

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

E-mail: martin.schmeing@mcgill.ca

Nonribosomal peptide synthetases - Natural antibiotic nanofactories

Microorganisms exist in complex and hostile environments. To obtain growth advantages, allow communication or inhibit competitors, fungi and bacteria produce small molecules with powerful and diverse activities. The chemical and biological properties of these compounds often make them useful to society as therapeutics (antifungals, antibacterials, antivirals, antitumours and immunosuppressants) and as natural green chemicals (agricultural agents, pesticides, emulsifiers, siderophores and research tools) [1-4]. Many such molecules (Figure 1) are synthesized in microbes by complex enzymes called nonribosomal peptide synthetases (NRPSs) [1-4].

NRPSs are large macromolecular machines that link substrates by peptide bond formation. Instead of making proteins, these enzymes produce a vast variety of small molecules with important and diverse biological activity, including well-known compounds such as penicillin, cyclosporin, daptomycin, viomycin and bacitracin. These NRPS products are of critical importance to human health:

penicillin has saved millions of lives worldwide [5]; organ transplantation would likely not have become so viable and widespread without cyclosporin and related compounds [6]; and, daptomycin (Cubicin) is a billion-dollar drug successful against hospital gram-positive infections [7]. As pathogens and other challenges to human health are dynamic, new derivatives of these and other NRPS compounds are needed.

NRPS products often do not resemble proteinaceous peptides. As well as the 20 standard amino acids, over 500 others can be used as building blocks, including D-amino acids, N-methyl amino acids and hydroxy acids [8]. In addition, the peptide products are often cyclic and can be branched (Figure 1) [1-4]. These differences between NRPS products and proteinaceous peptides result from the dissimilar mechanisms of reactions and processivity of the ribosome and NRPSs. NRPSs resemble assembly lines, with dedicated active sites for each amino acid added to the peptide, and a particular NRPS can make only one kind of peptide.

NRPSs are organized into modules of ~1000 amino acids, with each module responsible for the addition of one specific amino acid. In typical, "type 1 NRPSs", the number and order

FIGURE 1. Some representative nonribosomal peptides. Shown are the compounds cyclosporin (an immunosuppressant and antiviral agent), penicillin, bacitracin and linear gramicidin (all antibiotics). The stereochemistry of the α -carbon is displayed for linear gramicidin (gramicidin A) to highlight that the chirality in each amino acid alternates between L- and D-.

FIGURE 2. Reaction cycle of a typical elongation module and the linear gramicidin synthetase initiation module. Figure based on that in Reimer et al. [22].

FIGURE 3. Four structures of the LgrA linear gramicidin synthetase initiation module represent every major conformation in the module's catalytic cycle. Each structure shown is determined from crystals of the entire F-A-PCP initiation module. The PCP domain is not functionally required for the steps performed by the open (substrate binding) and closed (adenylation) steps, and is also not observed in the electron density of those states.

of the modules correspond to the length and sequence of amino acids in the peptide product (the co-linearity rule [9]). An NRPS system can be a single huge polypeptide chain or split over multiple chains that assemble through non-covalent interactions. It can be as short as one module or as long as eighteen modules, with a molecular weight of over 2 MDa in a single protein chain [10].

Modules contain multiple domains, each with a specific function [1-4]: a basic module contains a condensation (C) domain, an adenylation (A) domain and a peptidyl carrier protein (PCP) domain (Figure 2a). The A domain is responsible for recognition and binding of the amino acid (Figure 2a). It catalyzes first the adenylation of the amino acid using ATP, then the transfer of this activated amino acid to the PCP domain. The amino acid is ligated to the PCP domain via a prosthetic phosphopantetheinyl moiety which has been covalently attached to a serine in the PCP domain by a dedicated phosphopantetheinyl transferase[11]. The PCP domain transports the amino acid to the C domain, which catalyzes peptide bond formation between this amino acid and an amino acid or peptide attached to the PCP domain of the preceding module, thus elongating the peptide chain. Peptide bond formation results in attaching the growing peptide to this PCP domain, which then transports it to the downstream C domain and donates it through peptidyl transferase reaction to the downstream amino acid PCP domain. Once a PCP domain has donated its attached peptide, it is free to accept a new amino acid from the A domain and participate in the next cycle of assembly line peptide synthesis.

Other domains are also sometimes present in a module, and modify the amino acid or peptide, providing additional diversity [3, 12]. These include epimerisation (E) domains, which catalyze the racemization of L to D amino acids, methyltransferase domains, which conjugate methyl groups to carbons or nitrogens in the peptide, cyclization domains, which replace C domains and catalyze both condensation and intramolecular heterocyclization, reductase (R) domains, which reduce intermediates, and several others. The action of these tailoring domains introduces functionalities into the peptide product, including protease resistance and improved affinity.

NRPSs often terminate with a thioesterase (TE) domain, which cleaves off covalently-tethered peptide by catalyzing an attack of water or an internal nucleophile [13]. Reductase domains can also be used for termination. The typical organization of an NRPS is A-T-(C-A-T)_n-TE/R, with additional domains sometimes inserted.

Many NRPS-synthesized products have therapeutic or commercial value. Since NRPSs are modular, it is possible to construct novel NRPSs by adding, removing or substituting domains or modules. Engineered NRPSs can then be used to produce new compounds which may have increased potency or new biological activity. The strategy has high potential and could affect diverse fields such as agriculture, medicine and industrial chemistry, and would be very welcome as a source of new antibiotics in the face of growing resistance. Efforts to exploit this system have long been underway [14-16], but have been hampered by very low product yields [17-19]. To realize the full potential of bioengineering of NRPSs, we believe that the already excellent body of structural knowledge, which includes structures of isolated domains, di-domain constructs, and even full termination modules (reviewed by last year by Weissman [4], and continuing to grow [20-23]), must be augmented to provide increased understanding of the architecture and mechanisms of these huge machines.

Structures of the initiation module of linear gramicidin synthetase

Linear gramicidin is a clinical topical antibiotic, effective against gram-positive bacteria, which is used in dermatologic and ophthalmic topical formulations including several Polysporin ointments [24]. It is a 15 amino acid-long linear peptide capped at the N-terminus by a formyl group and at the C-terminus by ethanolamine. The stereochemistry of the α -carbon of each amino acid alternates between L- and D- along the pentadecapeptide chain, allowing it to assume a conformation known as a β helix, where the main chain atoms are on the inside of the helix and all side chains face outward [25]. For antibacterial activity, two gramicidin molecules dimerize in a head-to-head orientation in the bacterial membrane, resulting in a β helical pore that spans the membrane. This pore allows diffusion of monovalent cations across the membrane, which destroys the ion gradient and kills the bacteria. The N-terminal formyl group is essential for this activity [26].

The NRPS system that makes gramicidin was discovered by Marahiel and colleagues [26, 27]. It is made up of four NRPS proteins, linear gramicidin synthetase subunits A-D (LgrA-D). Notably, LgrA was found to contain a domain with homology to free-standing formyltransferase (FT) proteins, which catalyzes formylation using the cofactor N¹⁰-formyltetrahydrofolate (fTHF). There are three families of homologous FTs: 1) FTs that formylate initiator Met-tRNA^{fMet} for bacterial translation initiation[28]; 2) FTs that catalyze formylation steps in the purine synthesis

pathway[29]; and, 3) FTs that formylate sugar moieties for the bacterial outer membrane [30]. The F domain in LgrA was the first of several dozen to be found (others are in kolossin A synthetase [10], anabaenopeptilide synthetases [31], oxazolomycin synthetase [32], and many NRPSs in deposited genomic sequences). This F domain is the most N-terminal domain of the initiation module of LgrA, which has the domain organization F-A-PCP[26] (Figure 2b). The formylation event occurs after the adenylation and thiolation of the valine substrate in the A domain and before it is donated in the condensation reaction in the C domain of the downstream module[26] (Figure 2b).

We undertook biophysical and structural study of the LgrA initiation module to address outstanding questions pertaining to the architecture and catalytic cycle of this tailoring NRPS initiation module [22], including the following: How similar in structure are F domains and FTs? How is the F domain incorporated into the NRPS architecture? How does the PCP domain interact with and deliver the substrate to the F domain? To answer these questions, we initiated work with multiple constructs of LgrA from several bacterial strains. Our first attempts centered on excised constructs of the F domain in isolation, but these constructs expressed extremely poorly and readily aggregated. Only when the A domain was also included in the constructs were we able to produce well-behaved protein which we were able to crystallize. We determined a structure of this F-A di-domain to 2.5 Å resolution by molecular replacement.

The structure of the LgrA F-A di-domain is elongated, with the F domain fused onto the N-terminus of the A domain (Figure 3). The fusion point is surrounded by a large interface that buries 830 Å² of surface area, and is the only interdomain contact. Accompanying small angle X-ray scattering experiments showed that the F-A di-domain likely maintains this elongated and rigid conformation in solution. The F domain superimposes very well with structures of FTs, and likewise the A domain superimposes with other NRPS A domains. All but ~eight residues between the two domains are recognizable as either FT or A domain. These residues include a single α -helix and form part of the F domain–A domain interface. Furthermore, the putative Val-phosphopantetheine binding site of the F domain overlaps in position, size and hydrophilic nature with the sugar-dTDP binding site of sugar FTs [27, 30, 31]. The small differences in protein sequence between the F domain and sugar FTs and the length and hydrophilic nature of the substrates make it likely that the FT becomes an active NRPS F domain soon after the horizontal

gene transfer event that fuses a sugar FT upstream of the NRPS A domain.

The A domain is composed of a large A_{core} subdomain and a smaller C-terminal A_{sub} subdomain, which is mobile and donates residues for the adenylation reaction [4, 33–37]. In this F-A di-domain structure, the A_{sub} is disordered, and no corresponding electron density was observed in our maps. This is likely because all positions that had previously been observed for the A_{sub} were blocked by an adjacent symmetry-related molecule in the crystal [33–38]. Since this crystal form displays 82% apparent solvent content and was the only highly diffracting form we had then established, and since the A_{sub} subdomain was not known to be required for the PCP domain–F domain interaction, we attempted to manipulate this crystal form to study the PCP domain–F domain interaction [39]. We hypothesized that a new construct that did not contain the A_{sub} domain could be used to produce a crystal that displayed extremely large solvent channels and in which the F domain could bind a PCP domain introduced after crystallization (by “soaking experiment”). The F-A_{Δsub} construct we produced did express well and crystallize readily in the same crystallization conditions; however, F-A_{Δsub} crystals had reduced symmetry (from space group P4₁2₁2 to P4₁), and when we determined the structure, we found that the desired large solvent channels were filled with duplicate copies of F-A_{Δsub} which, interestingly, formed a packing network identical to the first copies [22, 39]. These two packing networks make next to no contact with each other, but completely block all potential access of the PCP domain to the F domain; thus, this crystal form did not hold promise to answer the question of how the PCP domain interacts with and deliver the substrate to the F domain.

Fortunately, we were then able to grow a series of new crystal forms of the entire F-A-PCP LgrA initiation module and determine their structures, which showed every major conformation required for its complex catalytic cycle (Figure 3). Several important conformations have been previously observed in experiments with single domain or di-domain constructs from disparate NRPS systems [4, 33–37], but it is informative to observe all these conformations within the context of the same, full module. To begin, the fTHF cofactor and the ATP and Val substrates must bind LgrA. The F domain does not make a large conformational change upon binding fTHF. The A domain, however, has an “open” conformation, suitable for substrate binding, where the A_{sub} domain is retracted from the substrate binding site. Substrate binding apparently favours the next, “closed”, conformation where the A_{sub} subdomain rotates ~30° to participate in

catalysis of the valine adenylate formation [36]. The A_{sub} then rotates 140° and allows the PCP domain to accept the Val onto the thiol of its phosphopantetheine arm [35, 37]. The next step is transportation of this Val to the F domain active site fully ~50 Å away. In this massive movement, the PCP domain translocates its center of mass 61 Å and rotates 75°, and the A_{sub} subdomain rotates 180°. In the resulting formylation conformation, we observe a quite small interaction surface between the PCP domain and the F domain. Interestingly, A_{sub} subdomain also repositions to donate binding surface to the PCP domain, which likely helps it achieve the required conformation for formylation. This is strikingly similar to the function of a second domain in methionyl-tRNA^{fMet} formyltransferase, which is responsible for presenting the methionyl-tRNA^{fMet} to its FT active site [28]. Once the Val is formylated by the F domain, the PCP can productively bind the next module's C domain, where fVal is donated in the condensation reaction. This frees the thiol of the PCP domain to participate in the next round of assembly line synthesis.

Studies such as these give a detailed, increased view of the complex catalytic cycle of NRPSs and give hope that, in future, bioengineering experiments could be very robustly used to produce novel NRPS-produced compounds with improved therapeutic activity; however, these structures also show why bioengineering undertakings may be difficult. The interfaces formed between the larger domains within modules are extensive and differ appreciably from one NRPS to another [21, 22, 38], which makes it difficult to simply swap out whole domains. Furthermore, the interactions that are necessarily transient for the catalytic cycle are fine-tuned through evolution. The PCP domain and the A_{sub} subdomain are only ~100 amino acids each, and yet the LgrA PCP1 domain must bind three different partners in its cycle. In other NRPSs, PCP domains may need to bind up to five other domains. Meanwhile, the A_{sub} subdomain must open and close for substrate binding, donate catalytic residues for adenylation, provide protein-protein interactions for positioning of the PCP for thiolation and formylation, and act as a linker to allow the PCP to bind distal active sites. Almost all the surface residues of the PCP domain and A_{sub} subdomain are involved in one or more of these jobs; therefore, experiments that introduce new domains or change the domains with which the PCP domain and A_{sub} subdomain have to interact must be designed considering these parameters. It is likely the successful bioengineering workflow will always require rounds of

evolution and selection after rational or biocombinatorial NRPS design to routinely produce successful designer NRPS therapeutics.

Acknowledgments

This work was supported by grants from the Canadian Institutes of Health Research and Canada Research Chairs. Thanks to Canadian Society for Clinical Investigations for the generous bestowal of the Joe Doupe Young Investigator Award. Thanks also to the co-authors of the original studies discussed herein, Janice Reimer, Martin Aloise, Harold Powell and Paul Harrison.

References

1. Walsh CT Polyketide and nonribosomal peptide antibiotics: Modularity and versatility. *Science* 2004; 303(5665):1805-1810.
2. Fischbach MA, Walsh CT Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: Logic, machinery, and mechanisms. *Chem Rev* 2006; 106(8):3468-3496.
3. Hur GH, Vickery CR, Burkart MD Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* 2012; 29(10):1074-1098.
4. Weissman KJ The structural biology of biosynthetic megaenzymes. *Nat Chem Biol* 2015; 11(9):660-670.
5. Zaffiri L, Gardner J, Toledo-Pereyra LH History of antibiotics. From salvarsan to cephalosporins. *J Invest Surg* 2012; 25(2):67-77.
6. Tedesco D, Haragsim L Cyclosporine: A review. *J Transplant* 2012; 2012:230386.
7. Robbel L, Marahiel MA Daptomycin, a bacterial lipopeptide synthesized by a nonribosomal machinery. *J Biol Chem* 2010; 285(36):27501-27508.
8. Caboche S, Pupin M, Leclerc V, et al. Norine: A database of nonribosomal peptides. *Nucleic Acids Res* 2008; 36(Database issue):D326-331.
9. Marahiel MA, Stachelhaus T, Mootz HD Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem Rev* 1997; 97(7):2651-2674.
10. Bode HB, Brachmann AO, Jadhav KB, et al. Structure elucidation and activity of kolossin A, the D-/L-pentadecapeptide product of a giant nonribosomal peptide synthetase. *Angewandte Chemie International Edition* 2015; 54(35):10352-10355.
11. Lai JR, Koglin A, Walsh CT Carrier protein structure and recognition in polyketide and nonribosomal peptide biosynthesis. *Biochemistry* 2006; 45(50):14869-14879.

12. Walsh CT, Chen H, Keating TA, et al. Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr Opin Chem Biol* 2001; 5(5):525-534.
13. Trauger JW, Kohli RM, Mootz HD, Marahiel MA, Walsh CT Peptide cyclization catalysed by the thioesterase domain of tyrocidine synthetase. *Nature* 2000; 407(6801):215-218.
14. Kittendorf JD, Sherman DH Developing tools for engineering hybrid polyketide synthetic pathways. *Curr Opin Biotechnol* 2006; 17(6):597-605.
15. Challis GL, Naismith JH Structural aspects of non-ribosomal peptide biosynthesis. *Curr Opin Struct Biol* 2004; 14(6):748-756.
16. Hahn M, Stachelhaus T Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc Natl Acad Sci U S A* 2006; 103(2):275-280.
17. Calcott MJ, Owen JG, Lamont IL, Ackerley DF Biosynthesis of novel pyoverdines by domain substitution in a nonribosomal peptide synthetase of *Pseudomonas aeruginosa*. *Appl Environ Microbiol* 2014; 80(18):5723-5731.
18. Kries H, Niquille DL, Hilvert D A subdomain swap strategy for reengineering nonribosomal peptides. *Chem Biol* 2015; 22(5):640-648.
19. Calcott MJ, Ackerley DF Portability of the thiolation domain in recombinant pyoverdine non-ribosomal peptide synthetases. *BMC Microbiol* 2015; 15:162.
20. Gaudelli NM, Long DH, Townsend CA Beta-lactam formation by a non-ribosomal peptide synthetase during antibiotic biosynthesis. *Nature* 2015; 520(7547):383-387.
21. Drake EJ, Miller BR, Shi C, et al. Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* 2016; 529(7585):235-238.
22. 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.
23. 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.
24. Wallace BA Common structural features in gramicidin and other ion channels. *Bioessays* 2000; 22(3):227-234.
25. Townsley LE, Tucker WA, Sham S, Hinton JF Structures of gramicidins a, b, and c incorporated into sodium dodecyl sulfate micelles. *Biochemistry* 2001; 40(39):11676-11686.
26. 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.
27. 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.
28. Schmitt E, Panvert M, Blanquet S, Mechulam Y Crystal structure of methionyl-tRNA^{fMet} transformylase complexed with the initiator formyl-methionyl- tRNA^{fMet}. *EMBO J* 1998; 17(23):6819-6826.
29. Almassy RJ, Janson CA, Kan CC, Hostomska Z Structures of apo and complexed *Escherichia coli* glycylamide ribonucleotide transformylase. *Proc Natl Acad Sci U S A* 1992; 89(13):6114-6118.
30. Thoden JB, Goneau MF, Gilbert M, Holden HM Structure of a sugar N-formyltransferase from *Campylobacter jejuni*. *Biochemistry* 2013; 52(35):6114-6126.
31. Rouhiainen L, Paulin L, Suomalainen S, et al. Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena* strain 90. *Molecular Microbiology* 2000; 37(1):156-167.
32. Zhao C, Coughlin JM, Ju J, et al. Oxazolomycin biosynthesis in *Streptomyces albus* ja3453 featuring an "acyltransferase-less" type I polyketide synthase that incorporates two distinct extender units. *J Biol Chem* 2010; 285(26):20097-20108.
33. 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.
34. Yonus H, Neumann P, Zimmermann S, et al. Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J Biol Chem* 2008; 283(47):32484-32491.
35. 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.
36. 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.
37. 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.
38. 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.
39. 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.