A structural model for multimodular NRPS assembly lines

Journal:	*Natural Product Reports*
Manuscript ID	NP-PER-07-2015-000082.R1
Article Type:	Viewpoint
Date Submitted by the Author:	08-Sep-2015
Complete List of Authors:	Marahiel, Mohamed A ; Philipps-Universität Marburg, Chemie/Biochemie

Journal Name

ARTICLE

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

www.rsc.org/

A structural model for multimodular NRPS assembly lines

Mohamed A. Marahiel[a]

1-Introduction

Non-ribosomal peptide synthetases (NRPSs) are exceptional megaenzymes that have evolved in bacteria and fungi to assemble highly complex, bioactive secondary metabolites of peptide origin. In contrast to the nucleic acid-dependent ribosomal peptide/protein synthesis that relies on the 20-22 proteinogenic amino acids, NRPSs can process several hundreds of monomers, including D- and β-amino acids, α-keto- and α-hydroxy acids as well as heterocycles and fatty acids, among other moieties.[1,2,3] To perform this complex chemistry, NRPSs rely on an array of large, repetitive catalytic units called modules, each comprised of several catalytic domains covalently linked within a single polypeptide chain. A simple NRPS elongation module contains at least three essential domains: An adenylation (A) domain responsible for selecting the substrate and activating it as an aminoacyl-adenylate, a small peptidyl carrier protein (PCP) domain that carries all acyl-intermediates on the terminal -SH group of its 4'-phosphopantetheine (Ppant) cofactor, and a condensation (C) domain which forms peptide bonds between the acyl-S-PCP intermediates of two adjacent modules. Therefore, all NRPS elongation modules are composed of at least three essential domains in the order C-A-PCP. A termination module contains an additional thioesterase (TE) domain, responsible for product release, either by hydrolysis or by cyclization, to generate either cyclic or cyclic-branched molecules. C domains are notably absent from initiation modules (A-PCP) of NRPS assembly lines. As an example for the modular organization of an NRPS machinery, the surfactin synthetase is depicted in figure 1.

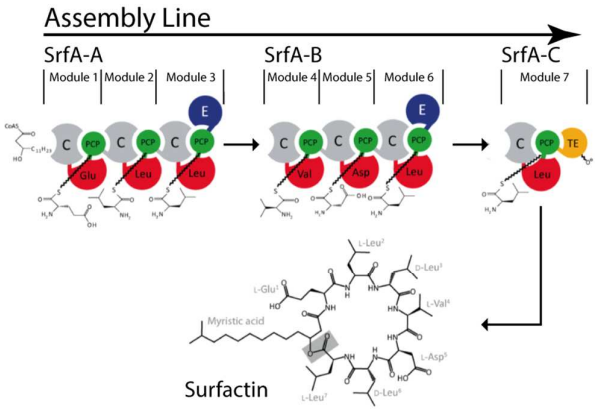

Fig.1 Surfactin synthetase, a prototypical NRPS machinery, assembles the lipoheptapetide surfactin. The synthetase consists of three subunits, SrfA-A, SrfA-B and SurfA-C (a pair of three extension modules and one termination module, respectively). The indicated building blocks incorporated by each module are selected and activated by the A domains (red). The reaction is initiated by the C domain (grey), which catalyses the transfer of the CoA-activated fatty acid onto the PCP-attached glutamate of the first module. During elongation (peptide bond formation, catalysed by the C domains) all biosynthetic intermediates remain attached to the PCP domain (green) as aminoacyl- or peptidyl-S-Ppant. The final product is released from the termination module (SrfA-C) by a regio- and stereoselective macrocyclization reaction catalysed by the TE domain (orange). Epimerization of the residues attached to modules three and six is catalysed by the E domains (blue). In total, 24 domains are included in the surfactin assembly line.

2-Structures and dynamics of the essential NRPS domains

Significant progress has been made in the past 20 years or so towards understanding the molecular basis that governs the reaction mechanisms of NRPSs as well as the structures and dynamics of their associated catalytic domains. Several crystal structures of dissected substrate selecting/activating A domains and NMR structures of PCP domains were solved in different catalytic states, providing insights into their active

[a] *Philipps University Marburg, Department of Chemistry/Biochemistry, Hans-Meerwein Str. 4, D-35032 Marburg.*

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

sites and reaction dynamics.[4,5] Also, the structures of NRPS didomains such as PCP-A, PCP-C, PCP-TE and an entire module (fig. 2) gave valuable insights into the dynamics of domain-domain interactions during catalysis.[6,7]

A domains (about 550 residues) were found to employ a novel, conserved fold distinct from that of aminoacyl-tRNA synthetases, despite the similar reactions—amino acid recognition and activation—catalysed by enzymes from both families. The A domain is characterized by its large, N-terminal subdomain and small, flexible, C-terminal subdomain, which enclose a hydrophobic active site for binding the amino acid and ATP substrates. Upon substrate binding and adenylate formation, the C-terminal subdomain rotates by some 140° relative to the large, N-terminal subdomain, thus adopting two distinct conformational states specific for the adenylation (aa-AMP) and thiolation (PCP-S-aa) reactions.[8] Extensive biochemical and structural studies on dissected A domains combined with phylogenetic studies revealed the substrate selectivity-conferring code and delineated the distinct evolutionary relationship between the domains and to the superfamily of adenylate-forming enzymes.[9,10] From the so-called A domain code, derived from ~10 amino acid residues that line the hydrophobic, phenylalanine-binding pocket in the structure of PheA[11], corresponding residues in the sequence of any A domain can be easily identified. Subsequent *in silico* studies demonstrated the power of the A domain code in predicting the substrate specificity of biochemically uncharacterized A domains and allowed for a rational switch of substrate specificity.[12,13]

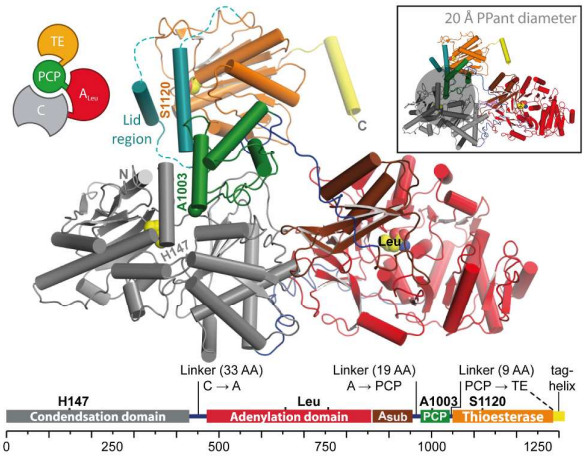

Fig.2 Overall structure of the termination module SrfA-C (C-A-PCP-TE) of surfactin synthetase (PDB code 2VSQ). The V-shaped C domain is shown in gray and its active site His147 in yellow spheres. The large, N-terminal A subdomain is shown in red and the leucine bound to its active site is depicted in yellow spheres. The C domain and the large A subdomain are connected via a 32-residue linker (blue) that makes numerous interactions with both domains, generating the C-A platform. The PCP domain (green) and the small, C-terminal A subdomain (brown) are located on top of the platform and are connected by a flexible linker shown in blue. The TE core domain is shown in orange and its lid region in light sea green. In the upper left panel a cartoon of the terminal module is shown with the same color code, and in the upper right panel the gray circle defines the 20-Å radius of the Ppant arm. The diagram below the structure depicts the relative sizes of the domains and linker regions and also indicates active site residues

The small PCP domain (about 100 aa), located downstream of the A domain, is the site of Ppant cofactor modification, installed on the side chain of a highly conserved serine residue by a dedicated Ppant-transferase (like Sfp) that utilizes Coenzyme A for cofactor donation.[14] Inactivation of this essential post-translational modification in any NRPS module renders the corresponding enzyme inactive by interrupting the transport of PCP-S-acyl intermediates. Several NMR and crystal structures of PCP domains, either as dissected domains or in association with other neighbouring domains, have defined the PCP as a four-helix bundle with an extended loop region between helices 1 and 2.[15] The conserved serine residue to which the Ppant cofactor is attached is located at the end of the loop region C-terminal to helix 2. NMR structural studies also showed conformational rearrangement for the PCP domain following the transition from apo- to holo- and from holo- to acyl-S-PCP.[16] NMR titration studies of ^{15}N-labeled PCP with other modifying enzymes confirmed that such conformational changes are important for partner protein recognition during post-translational modification and substrate shuffling. Other complex multimodular enzymes that use acyl carrier proteins (ACPs), such as polyketide and fatty acid synthases, may rely on a similar strategy to spatiotemporally coordinate the shuffling of substrates/intermediates.

After substrate selection, activation and covalent attachment as a thioester to the PCP domain, the third essential NRPS domain, the C domain (about 450 aa), catalyses peptide bond formation between the acyl-S-PCPs of two adjacent modules. Crystal structures of dissected C domains, PCP-C didomains and a C domain within an intact module revealed a highly conserved fold. The C domain is comprised of two related subdomains of the CAT (chloramphenicol acetyl transferase) fold, which together assume a V-shaped structure.[17,18] The active site histidine residue is located at the floor of the canyon and can be reached from both sides of the V-shaped C domain. This fold generates the biochemically defined acceptor and donor sites that accommodate the aminoacyl-S-PCP and the peptidyl-S-PCP, respectively. In the crystal structure of a PCP-C didomain shown in figure 3, the carrier PCP domain is attached to the donor site of the C domain in an unproductive interaction, as its active site serine residue (site of Ppant binding, in this structure changed to alanine) is 50 Å away from the active site histidine residue of the C domain. Therefore, the PCP domain in this structure may be in the appropriate conformational state for productive interaction with the acceptor site of the proceeding C domain. Consequently, large conformational transitions are necessary for the flexible PCP domains to shuttle the acyl-S-PCP intermediates between acceptor and donor sites of the C domains.

For most bacterial NRPSs, product release is catalysed by a unique thioesterase (TE) domain (about 280 aa), which is only found within termination modules (C-A-PCP-TE). Although most TE domains share low sequence similarity, their α/β-hydrolase fold and catalytic triad (His, Ser, Asp) are highly conserved.[19] In general, TE domains catalyse two reactions:

First, transfer of the peptidyl-S-PCP of the termination module onto the active site serine residue, generating an acyl-O-TE; second, product release, either by hydrolysis or a regio- and stereoselective cyclization that generates a cyclic or cyclic-branched molecule. In the case of cyclization, a lid region excludes water from acting as a nucleophile in the TE active site following formation of acyl-O-TE. A selective, internal nucleophile attacks this acyl-O-TE intermediate to yield either a lactam or lactone as the final product.

Most of the biochemical and structural studies conducted on dissected catalytic domains of NRPSs indicate that they are folded and active as independent catalytic units. This supports the hypothesis that NRPSs may have evolved as multimodular enzymes from individual catalytic units. Though costly for the host microorganism to assemble, such complex, multifunctional enzymes, covalently tethered to their intermediates, greatly facilitate the unusual peptide chemistry that results in diverse secondary metabolites, ultimately conferring a pronounced survival advantage to the host.

3-From the structure of a module to a model of a multimodular NRPS assembly line

The termination module of surfactin synthetase, SrfA-C (144 kDa), catalyses the incorporation of leucine and the release of the lipoheptapeptide surfactin product. Recently, structural elucidation of the entire termination module SrfA-C (C-A-PCP-TE) showed clearly that each of the four catalytic domains shares the fold of its dissected counterpart[20] (see fig. 2). This whole-module structure provides the unique opportunity to visualize linker regions between domains as well as domain-domain interactions. The C domain and the large, N-terminal A subdomain (about two-thirds of the module) are tightly connected by a 32-residue linker that strongly interacts with the surfaces of both domains. This tight association generates a large and rigid platform, with the active sites of the C and A domains pointing in the same direction. On top of this platform (workbench) are the small A subdomain and the PCP domain. Due to the flexible linker (15 residues) connecting them, they can presumably reach the active sites of both the C and A domains.

In the SrfA-C module structure, the PCP domain is positioned at the C domain acceptor site and its cofactor-binding serine residue (in this structure mutated to alanine) is only 16 Å away from the active site histidine residue of the C domain. This distance is within the radius of the 20-Å Ppant arm. However, for the PCP domain to reach the active site of the A domain (about 57 Å away) or the TE domain (about 43 Å away), large conformational changes are necessary. The TE domain of the SrfA-C module, displaying the conserved α-/β-hydrolase fold, is connected to the PCP domain via a short, flexible (9-residue) linker.

In summary, our understanding of the structure and function of the essential NRPS domains (A, C, PCP and TE) is now highly advanced. Knowledge also exists about how these domains are linked to each other in the context of an entire catalytic module. However, no structural details regarding the multimodular organization of large NRPS assembly lines are currently available. Here, we propose a structural model for a multimodular NRPS, based on the known structures of a PCP-C dimodule and an entire C-A-PCP elongation module. In both structures, the PCP domains assume the so-called acceptor site conformation; therefore, in the model generated for a multimodular NRPS, the PCP domain resides in the acceptor site of each module. As shown in figure 3, superposing the C-domains of the PCP-C and C-A-PCP structures produces a PCP-C-A-PCP four-domain structural model. In this structure, the relative orientation of the two PCP domains to the A and C domains is shown. Next, PCP domains from several C-A-PCP elongation modules are successively superposed onto the PCP-C-A-PCP structural model, yielding a helical organization wherein each module is rotated 120° relative to its neighbour along the helical axis.[21] For a full helical turn (360°), three modules are required. Figure 4 shows a model of a seven-module NRPS comprised of six elongation modules and one termination module. In this helical model, the majority of the C-A catalytic platform is oriented outwards, whereas the PCP domains alternate along the helical axis. All PCP domains are in the acceptor site orientation and, after substrate loading and peptide bond formation by the C-A workbench should move to the donor site of the downstream C domain. In this model, the PCP domains should be able to interact with the corresponding A domain for substrate loading and to shuttle the acyl-intermediates to the acceptor and donor sites of the C domains. The side and front views of the multimodular NRPS model (fig. 4) show the linear organization of the PCP domains as an assembly line within the helical structure, and suggest how these carrier domains are shielded from solvent to prevent substrate/intermediate hydrolysis.

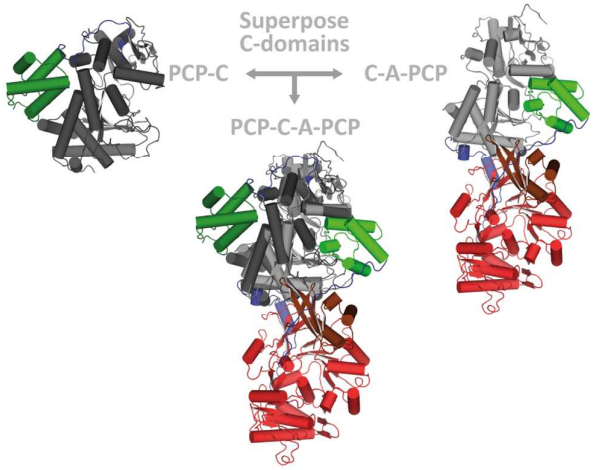

Fig. 3 Superposition of the C domains from PCP-C and C-A-PCP structures (PDB codes 2JGP and 2VSQ, respectively) generates the structural model for the four-domain (PCP-C-A-PCP) module in which both PCP domains are in the acceptor site orientation. Next, superposition of the PCP of several elongation (C-A-PCP) modules onto PCP-C-A-PCP generates the helical multimodular NRPS model shown in Figure 4.

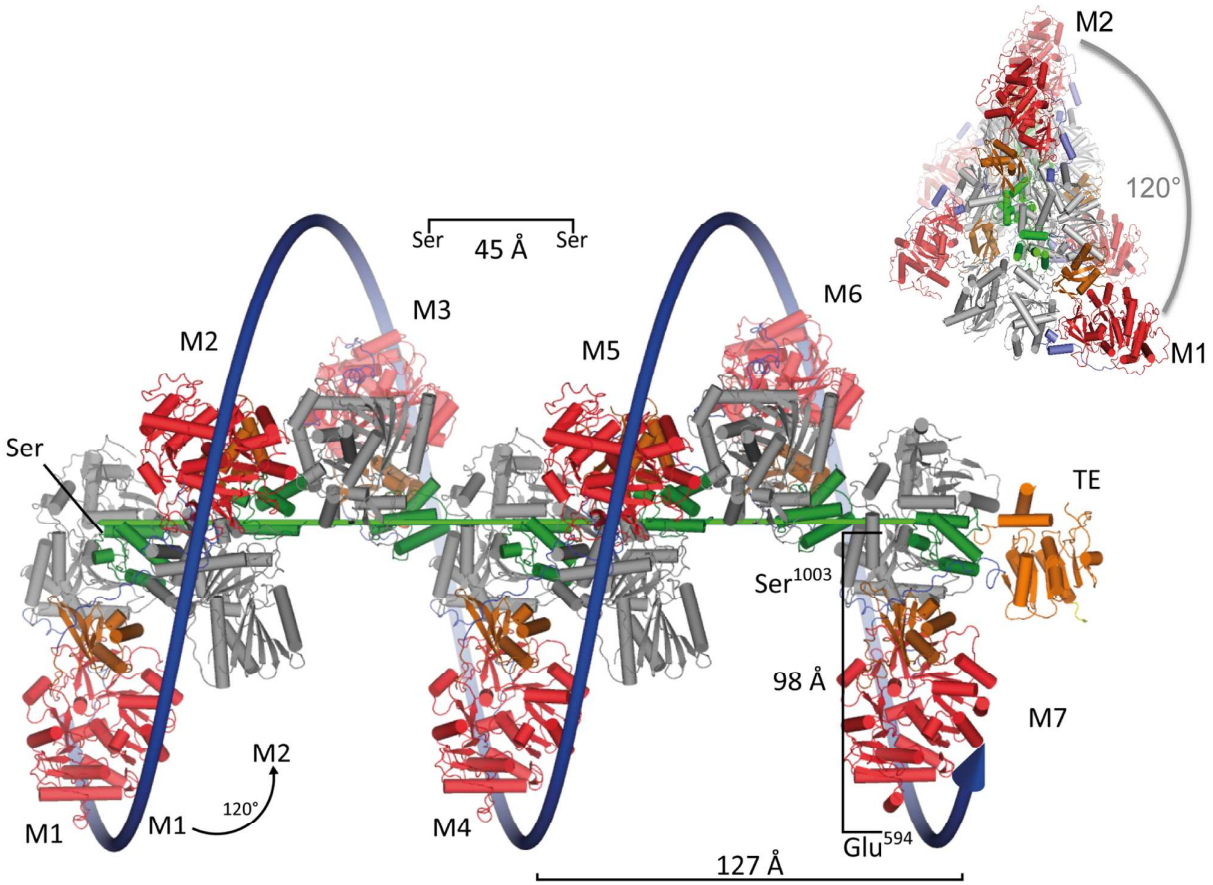

Fig. 4 Side view of a helical model of a heptamodular NRPS. In this model, the single elongation modules are 120° (counterclockwise) apart along a helical axis. A major portion of the C-A platform (gray-red) of each module is oriented outwards toward the solvent, whereas the PCP domains (green) alternate along the helical axis, generating an assembly line-like structure. The termination module (M7) carries a C-terminal TE domain. A front view of the model is show in the right upper panel.[21]

In this model the active site serine residues (Ppant binding sites) are about 45 Å apart, suggesting that a substantial rearrangement of the PCP-bound Ppant arm is needed. Although this multimodular NRPS model is based on known PCP-C didomain and C-A-PCP module structures, it must be validated by structural elucidation of an NRPS containing three or more modules, either by X-ray crystallography or cryo-EM. Until then, the suggested model is debatable.

Acknowledgements

I am very grateful to all past and present members of my group for their brilliant contributions and to our collaborators for their engagement. I thank Alan Tanović and Julian D. Hegemann for support and Christopher D. Fage for critical reading of the manuscript. I also gratefully acknowledge the financial support by the DFG.

Notes and references

1 M. A. Marahiel, T. Stachelhaus, H. D. Mootz, Chem Rev., 1997 , **7**, 2651-2674.
2 D. Konz, M. A. Marahiel, Chem Biol. 1999, **6**, 39-48.
3 R. Finking, M. A. Marahiel, Annu Rev Microbiol. 2004, **58**, 453-488
4 X. F. Tan, Y. N. Dai, K. Zhou, Y. L. Jiang, Y. M. Ren, Y. Chen, C. Z. Zhou, Acta Crystallogr D Biol Crystallogr. 2015, **4**, 873-881
5 A. Koglin, C. T. Walsh, Nat Prod Rep. 2009, **8**, 987-1000.
6 M. A. Marahiel, J Pept Sci. 2009, **15**, 799-807.
7 M. Strieker, A. Tanović, M. A. Marahiel, Curr Opin Struct Biol. 2010 , **20**, 234-240.
8 J. A. Sundlov, C. Shi, D. J. Wilson, C. C. Aldrich, A. M. Gulick, Chem Biol. 2012, **19**, 188-198.
9 T. Stachelhaus, M. D. Mootz, M. A. Marahiel, Chem Biol. 1999, **6**, 493-505.
10 G. L. Challis, J. Ravel, C. A. Townsend, Chem Biol. 2000, **7**, 211-224.
11 E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, EMBO J. 1997, **16**, 4174-4183.
12 C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben, D. H. Huson, Nucleic Acids Res. 2005, **33**, 5799-5808
13 H. Jenke-Kodama, E. Dittmann, Nat. Prod. Rep. 2009, **26**, 874-883.

14 K. Reuter, M. R. Mofid, M. A. Marahiel, R. Ficner, EMBO J. 1999, **18**, 6823-6831

15 T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, Structure. 2000, **8**, 407-418.

16 A. Koglin, M. R. Mofid, F. Löhr, B. Schäfer, V. V. Rogov, M. M. Blum, T. Mittag, M. A. Marahiel, F. Bernhard, V. Dötsch, Science. 2006, **312**, 273-276.

17 T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating, Nat Struct Biol. 2002, **9**, 522-526.

18 S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L. O. Essen, Structure. 2007, **15**, 781-792.

19 S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs, Structure. 2002, **10**, 301-310.

20 A. Tanović, S. A. Samel, L. O. Essen, M. A. Marahiel, Science. 2008, **321**, 659-663.

21 A. Tanović, PhD Thesis, 2012, Philipps University Marburg.