Biosynthesis

DOI: 10.1002/anie.200803293

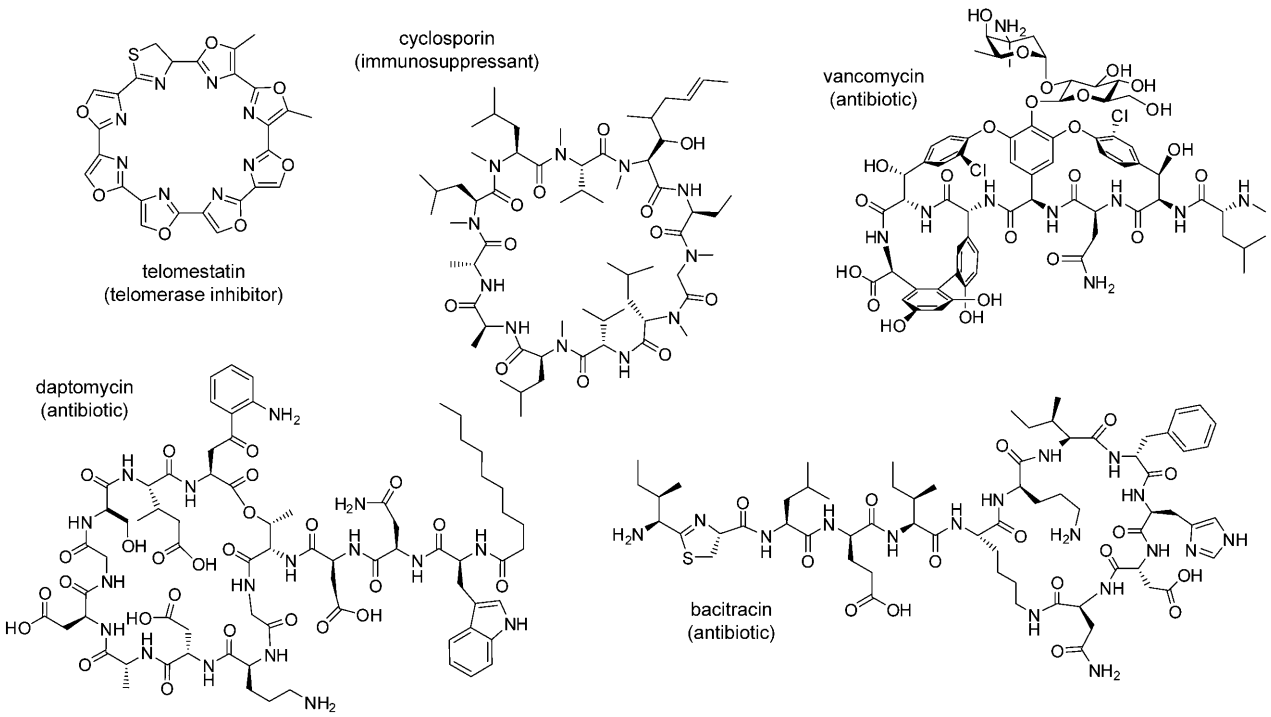
Crystal Structure of a Molecular Assembly Line**

*Kira J. Weissman and Rolf Müller**

biosynthesis · enzymes · natural products · structure elucidation

As a testimony to the power of persistence, Essen, Marahiel and co-workers report in *Science*,[1] the first crystal structure of an entire module of domains derived from a non-ribosomal peptide synthetase (NRPS) system—a landmark in the field. Although Marahiel and co-workers succeeded in solving the structure of a single NRPS-derived domain as early as 1997,[2] a further eleven years were required to obtain diffracting crystals of a complete module.

Non-ribosomal peptides (NRPs) are a diverse group of typically cyclic natural products (Scheme 1), which contain not only proteinogenic amino acids, but hundreds of other biosynthetic building blocks.[3] These metabolites are assembled in microbes by non-ribosomal peptide synthetases (NRPSs)—nature's nucleic-acid-independent route to peptide synthesis. In this mechanism, the sequence of amino acids incorporated into a product is determined by the order of autonomously folding enzymatic domains within gigantic multienzymes. Each domain performs a specific task in the pathway, and so NRPSs act like assembly lines on a molecular scale. Domains are grouped together into functional units called "modules", where each module catalyzes a single chain-extension step. This minimally requires an adenylation (A) domain, for selecting and activating the amino acid monomer as its adenylate, and a condensation (C) domain, to join the building block to the growing chain. Between these domains are noncatalytic domains called peptidyl carrier



Scheme 1. Structures and biological activities of representative non-ribosomal peptides.

proteins (PCPs), which are each equipped with the prosthetic group phosphopantetheine (Ppant). The terminal thiol of the Ppant is the attachment site for the incoming residue, and also the site of chain elongation as the growing peptidyl chain is translocated through the biosynthetic machinery. This thioester-based tethering activates the amino acid for the condensation reaction and enables substrate channeling to the various active sites, increasing the overall efficiency of the

[*] Dr. K. J. Weissman, Prof. R. Müller
Department of Pharmaceutical Biotechnology
Saarland University
P.O. Box 151150, 66041 Saarbrücken (Germany)
Fax: (+49) 681-3027-0201
E-mail: rom@mx.uni-saarland.de

[**] This research was supported by the Bundesministerium für Bildung und Forschung and the Deutsche Forschungsgemeinschaft.

NRPS. The basic domain set can be augmented by enzymes which modify and decorate the amino acids, for example by epimerization, *N*-methylation, and cyclization/dehydration. Chain release by macrocyclization is typically effected by a thioesterase (TE) domain, fused to the end of the last module. Using this division-of-labor principle, nature converts simple building blocks into products of high structural complexity. A similar biosynthetic stategy is exploited to construct another large family of secondary metabolites, the polyketides.[4]

Discovering the underlying modular architecture of NRPSs was encouraging, as it suggested that novel variants of these compounds could be generated by reprogramming the synthetases. This idea is not just of academic interest, as NRP metabolites are notable for their useful therapeutic activities, ranging from anticancer, to anti-infective, to immunosupressant properties (Scheme 1), making analogues attractive as drug leads. However, biological systems are rarely as simple as we would hope: many of the early, even modest attempts to reconfigure NPRS machineries by exchange of domains, proved disappointing.[5–7] Consequently, deciphering the detailed molecular enzymology and structural biology of NRPSs became a major focus of work. Although the X-ray crystal structure of an entire module has always been a target, the first portion of an NRPS to yield crystallographic analysis was a discrete A domain,[2] followed closely by several TEs,[8,9] and a C domain.[10] Subsequently, a PCP–C didomain, spanning the junction between NRPS modules, was reported.[11] The structure of a representative PCP was also determined by NMR spectroscopy.[12,13] The crystal structures provided important insights into the catalytic mechanisms of the various domains, as well as revealing a predictive 10-residue specificity code for the A domains (the so-called non-ribosomal code) which is now widely used to predict the products of orphan NRPSs (those where the function is unknown).[14,15] The PCP was shown to populate three alternative conformational states,[13] suggesting that conformational switching may be used to program alternative interactions with the domain's multiple partners. Critically, however, these structures could not reveal the three-dimensional relationship between the domains within a typical module, information which is key to understanding how the PCP-bound substrate is ferried among the active sites, and how these movements are orchestrated.[16]

The crystal structure (at a resolution of 2.6 Å) of SrfA–C, a prototypical C–A–PCP–TE module from surfactin synthetase (Figure 1),[1] goes a considerable way to addressing these deficiencies. To obtain diffracting crystals, Marahiel, Essen, and co-workers had to substitute the active serine of the PCP catalytic center with alanine. This mutation, which eliminates the possibility of post-translational addition of the Ppant moiety, was shown previously to trap the PCP domain in one of its three conformations.[13] In this way, they were able to reduce the overall conformational heterogeneity of the module, which presumably had stymied previous crystallographic attempts. The core of the module is a stable, rectangular catalytic platform formed by the C domain and the major N-terminal core (A_{core}) of the A domain, with both active sites arrayed on

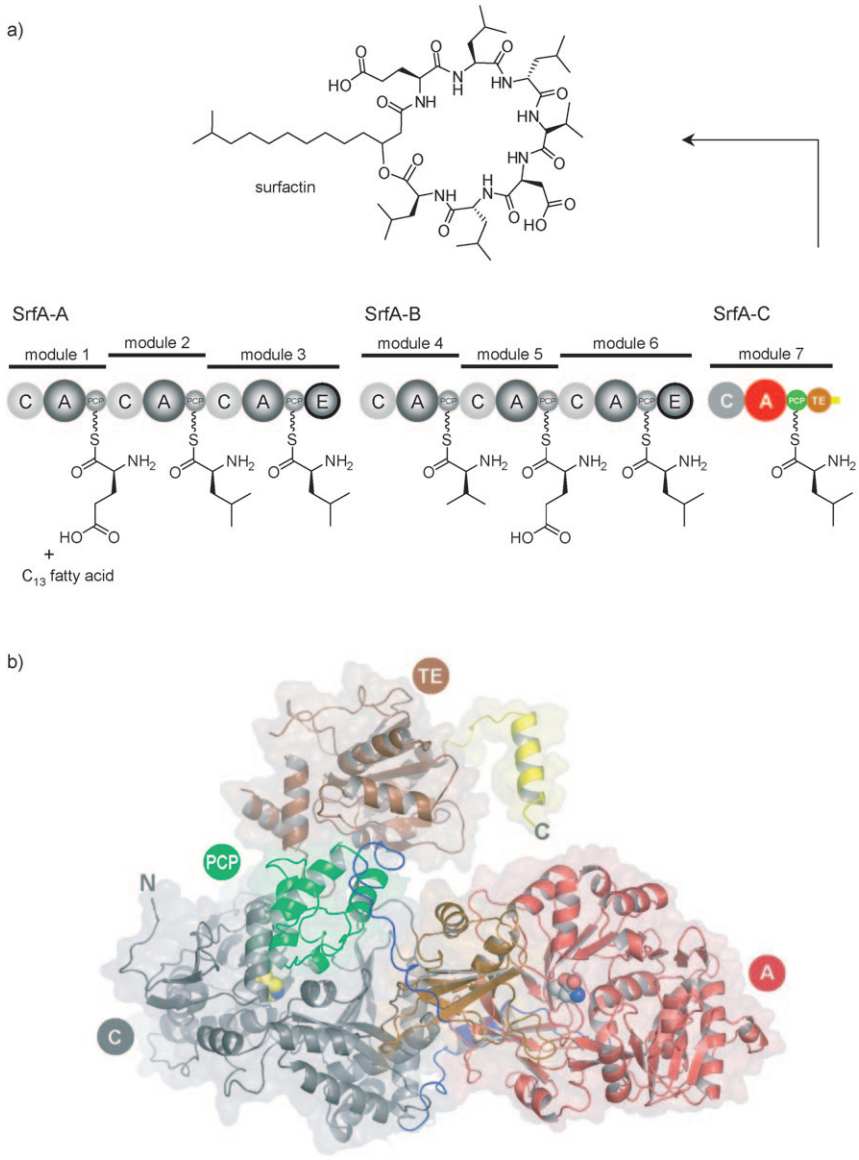


Figure 1. Biosynthesis of surfactin in *Bacillus subtilis*. a) Surfactin is assembled by an NRPS consisting of three subunits, SrfA-A, SrfA-B, and SrfA-C. SrfA-A and SrfA-B each include three modules, while SrfA-C comprises the termination module. The building blocks incorporated by each module are indicated (epimerization is likely to occur after peptide bond formation). b) Overall structure of SrfA-C solved at 2.6 Å resolution. Coloring of the domains is as in (a). The active site His of the C domain and a Leu bound within the A domain active site, are shown in space-filling representation. Reprinted with permission from ref. [1]; copyright 2008, American Association for the Advancement of Science.

the same side. The complex is "glued" together by an extensive interface between the C and A_{core} domains, as well as by close association of the catalytic regions with the well-defined, intervening linker. The rest of the A domain and the PCP domain are tethered to the platform and to each other with short, flexible linkers. This observation leads to a model in which the two domains move relative to the static C–A_{core} platform, in the course of each catalytic cycle. Such mobility is clearly required, as the catalytic sites within the C and A_{core} domains are separated by 63 Å, exceeding by some measure the reach of a static phosphopantetheine moiety (ca. 20 Å). The TE is essentially identical to the solved structures of the discrete domain,[8,9] and forms a distinct region within the module.

The modular structure also provides two important insights into how successive modules interact with each other to carry out chain elongation. In the structure, the downstream PCP (acceptor) is lodged at the C domain's acceptor site. The site for recognition of the upstream (donor) PCP is situated just across an active site canyon running through the C domain, so that both PCPs are simultaneously within reach of the catalytic histidine. In many cases, however, acceptor and donor PCPs are located on separate multienzymes. Thus, constructing the correct product requires that the two polypeptides specifically associate (or "dock") with each other, while resisting incorrect contacts with other multienzymes on the assembly line. Previous work identified matched sequence regions (called communication-mediating (COM) domains) at the extreme C- and N-terminal ends of the proteins, which contribute to partner selectivity.[17] According to the docking model, five residues on each α-helical COM domain interact with complementary amino acids on the partner COM domain, to form an overall antiparallel coiled-coil.[17] Serendipitously, the *myc*-His_6 purification tag appended to the C-terminus of SrfA-C resembles the putative COM-helices, in terms of both sequence and hydrophobicity. In the structure, the tag on one module is enveloped by a hand-shaped motif within the C domain of an adjacent module, contacting an array of largely hydrophobic residues. The "COM-hand" includes the previously identified N-terminal COM region, but additionally a three-stranded β-sheet contributed by the C domain. Evidently, the structural basis for specific docking between successive NRPS multienzymes is more complex than was originally appreciated. It is thus noteworthy that a single point mutation in the N-terminal COM was apparently sufficient to redirect docking specificity.[17]

Overall, the structure reveals that extensive domain rearrangements must occur during NRPS operation. However, this is also the main weakness of the work, as a crystal structure can only capture a single snapshot of a much more complex reaction sequence. We cannot deduce, for example, how the structure will evolve to allow the PCP to interact with the A domain and with its downstream counterpart (in this case the TE, but for internal modules, a C domain), nor how these rearrangements are programmed. The absence of the Ppant moiety (and thus of substrate) is also a significant issue, particularly as prosthetic-group binding modulates the conformational state of the PCP domain.[13] Clearly, further structures are required, in which an activated PCP interacts with its remaining modular partners. Recent work suggests a possible way forward.[18] By exploiting the inherent "promiscuity" of phosphopantetheinyl transferase enzymes, the Ppant arm of the PCP can be replaced by an inhibitor of a target active site. This modification directs the PCP to interact with a specific domain, trapping the overall complex in one of its catalytically relevant conformations. In future, it may be possible to combine a series of such static images to clarify the complete NRPS catalytic cycle.

In the meantime, what lessons can be gleaned for attempts to manipulate NRPS systems? The intimate association of the C and A domains suggests that they should be exchanged into new modular contexts as a functional pair, a strategy which simultaneously accommodates the relatively tight specificity of the C domain at its acceptor site.[19] The structure also highlights amino acid residue positions in the PCP, which can be modified to create new, productive C–PCP interactions.[20] Engineering of COM domains has already been exploited to generate artificial combinations of NRPS subunits.[17] Tailoring of the remaining β-sheet portion of the COM-hand should further optimize such non-native interfaces, increasing the efficiency of hybrid synthetases. The results of such structure-based engineering are eagerly anticipated.

Published online: September 9, 2008

[1] A. Tanovic, S. A. Samel, L.-O. Essen, M. A. Marahiel, *Science* **2008**, 321, 659.
[2] E. Conti, T. Stachelhaus, M. A. Marahiel, P. Brick, *EMBO J.* **1997**, *16*, 4174.
[3] J. Grünewald, M. A. Marahiel, *Microbiol. Mol. Biol. Rev.* **2006**, *70*, 121.
[4] J. Staunton, K. J. Weissman, *Nat. Prod. Rep.* **2001**, *18*, 380.
[5] S. Doekel, M. A. Marahiel, *Chem. Biol.* **2000**, *7*, 373.
[6] H. D. Mootz, D. Schwarzer, M. A. Marahiel, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 5848.
[7] A. Schneider, T. Stachelhaus, M. A. Marahiel, *Mol. Gen. Genetics* **1998**, *257*, 308.
[8] S. D. Bruner, T. Weber, R. M. Kohli, D. Schwarzer, M. A. Marahiel, C. T. Walsh, M. T. Stubbs, *Structure* **2002**, *10*, 301.
[9] S. A. Samel, B. Wagner, M. A. Marahiel, L. O. Essen, *J. Mol. Biol.* **2006**, *359*, 876.
[10] T. A. Keating, C. G. Marshall, C. T. Walsh, A. E. Keating, *Nat. Struct. Biol.* **2002**, *9*, 522.
[11] S. A. Samel, G. Schoenafinger, T. A. Knappe, M. A. Marahiel, L. O. Essen, *Structure* **2007**, *15*, 781.
[12] T. Weber, R. Baumgartner, C. Renner, M. A. Marahiel, T. A. Holak, *Structure* **2000**, *8*, 407.
[13] 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.
[14] T. Stachelhaus, H. D. Mootz, M. A. Marahiel, *Chem. Biol.* **1999**, *6*, 493.
[15] G. L. Challis, J. Ravel, C. Townsend, *Chem. Biol.* **2000**, *7*, 211.
[16] K. J. Weissman, R. Müller, *ChemBioChem* **2008**, *9*, 826.
[17] M. Hahn, T. Stachelhaus, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 275.
[18] Y. Liu, S. D. Bruner, *ChemBioChem* **2007**, *8*, 617.
[19] G. L. Challis, J. H. Naismith, *Curr. Opin. Struct. Biol.* **2004**, *14*, 748.
[20] J. R. Lai, A. Koglin, C. T. Walsh, *Biochemistry* **2006**, *45*, 14869.

