

Review



Cite this article: Bonhomme S, Dessen A, Macheboeuf P. 2021 The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities. *Open Biol.* **11**: 200386. <https://doi.org/10.1098/rsob.200386>

Received: 26 November 2020

Accepted: 5 May 2021

Subject Area:

biochemistry/structural biology

Keywords:

non-ribosomal peptide synthetases, flexibility, supramodular architecture

Author for correspondence:

Pauline Macheboeuf

e-mail: pauline.macheboeuf@ibs.fr

The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities

Sarah Bonhomme¹, Andréa Dessen^{1,2} and Pauline Macheboeuf¹

¹Univ. Grenoble Alpes, CNRS, CEA, IBS, F-38000 Grenoble, France

²Brazilian Biosciences National Laboratory (LNBio), CNPEM, Campinas 13084-971, São Paulo, Brazil

AD, 0000-0001-6487-4020; PM, 0000-0003-4523-9923

Non-ribosomal peptide synthetases (NRPSs) are multienzymes that produce complex natural metabolites with many applications in medicine and agriculture. They are composed of numerous catalytic domains that elongate and chemically modify amino acid substrates or derivatives and of non-catalytic carrier protein domains that can tether and shuttle the growing products to the different catalytic domains. The intrinsic flexibility of NRPSs permits conformational rearrangements that are required to allow interactions between catalytic and carrier protein domains. Their large size coupled to this flexibility renders these multi-domain proteins very challenging for structural characterization. Here, we summarize recent studies that offer structural views of multi-domain NRPSs in various catalytically relevant conformations, thus providing an increased comprehension of their catalytic cycle. A better structural understanding of these multienzymes provides novel perspectives for their re-engineering to synthesize new bioactive metabolites.

1. Introduction: the inherent flexibility of non-ribosomal peptide synthetases

Natural products are secondary metabolites synthesized by microorganisms in order to adapt to their environment [1]. Many of these natural products have been used for medical purposes, such as the antibiotics daptomycin and vancomycin [2] and the anti-cancer molecule bleomycin [3]. Among these natural products, non-ribosomal peptides (NRPs) make up a vast class of peptide-based metabolites, synthesized independently from the ribosome by large machineries named non-ribosomal peptide synthetases (NRPSs). For instance, the surfactin lipopeptide is synthesized by the 1 MDa surfactin NRPS from *Bacillus subtilis* (figure 1) [4]. Recent progress in whole genome sequencing has revealed the existence of numerous NRPS gene clusters among bacteria and fungi, mostly of unknown function [5–7]. Nevertheless, the structural understanding of the machineries that produce these metabolites has long remained limited to studies of isolated domains [8] until relatively recently, and has evolved dramatically in the last few years.

NRPSs are classified into two categories, type I and type II. In type II NRPSs, the incorporation of an amino acid into the metabolite necessitates the involvement of several domains carried by distinct proteins [9]. By contrast, the type I NRPS megaenzymes use an assembly line strategy (figure 1) with modules that act sequentially, each being responsible for the incorporation of an amino acid into the final metabolite. A NRPS assembly line is composed of an initiation module (module 1 in figure 1*a*), a termination module

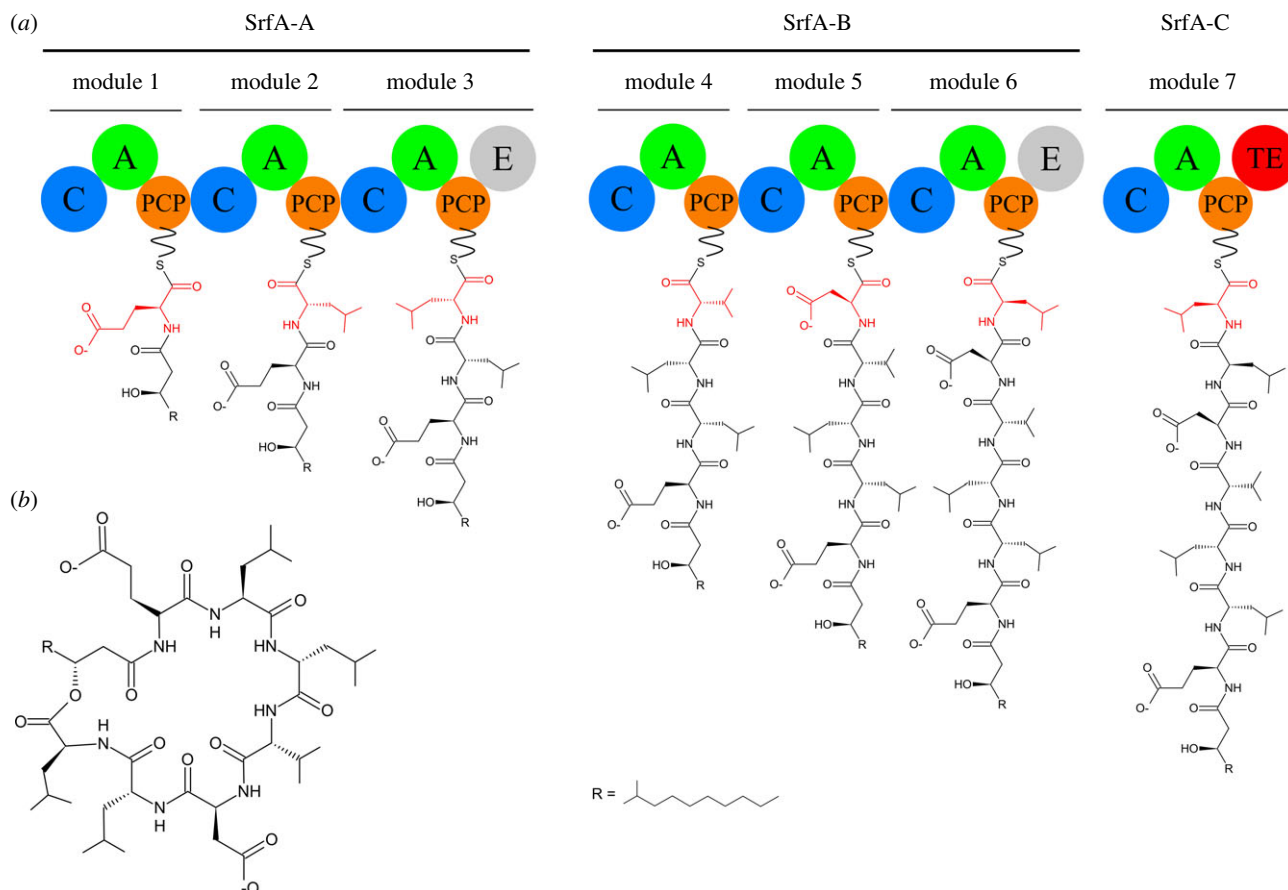


Figure 1. Organization of the surfactin NRPS. (a) The surfactin NRPS is composed of three polypeptides: SrfA-A, SrfA-B and SrfA-C. Each polypeptide is composed of one or several modules, each one being responsible for the incorporation of an amino acid, highlighted in red in the growing metabolite. Module 1 is the initiation module, module 7 is the termination module and those in between are elongation modules. The addition of an amino acid into the metabolite requires the cooperation between domains, represented as coloured spheres. All surfactin NRPS modules possess a unique non-catalytic domain, the PCP (in orange) that tethers the growing metabolite. Each module also contains at least two catalytic domains: a condensation domain (C, in blue) and an adenylation domain (A, in green). Additionally, modules 3 and 6 possess an optional epimerization domain (E, in grey). Finally, the termination module ends with a thioesterase (TE, in red) domain that releases and cyclizes the surfactin molecule. (b) Chemical structure of the surfactin.

(module 7 in figure 1a) and one or several elongation modules (2 to 6 in figure 1a). In general, several modules are fused into a single polypeptide. Polyketide synthases (PKSs) that are also modular megaenzymes, adopt the same assembly line logic as NRPSs but they use small carbon chain substrates instead of amino acid substrates [10]. This similar strategy explains the existence of numerous hybrid NRPS/PKS assembly lines that produce hybrid peptide-polyketide metabolites [11].

Most modules are composed of two catalytic domains, the adenylation (A) and the condensation (C) domains (figure 1a); each NRPS module also incorporates the non-catalytic, but essential, peptidyl carrier protein (PCP) domain [12]. In this review, we will consider that a classical module starts with a C domain and ends with a PCP domain (figure 1). The PCP domain functions as an anchoring platform for shuttling substrates to the different catalytic domains within a module; it also allows the transport of the modified substrate from the upstream to the downstream module (figure 1a). PCP domains, whose masses are in the range of 10 kDa, have been mostly studied in isolated form by NMR [13–15]. They fold as a right-handed four helix bundle, with the four helices (I, II, III and IV) being connected by loops. At the N-terminus of helix II, all PCP domains possess a conserved serine residue that serves as an attachment

point for a phosphopantetheine arm (PPant arm). This post translational modification is catalysed by phosphopantetheinyl transferases (PPtases) that convert apo-PCPs into holo-PCPs [16]. The 20 Å long PPant arm displays a free thiol at its extremity that allows the loading of various substrates. In this review, the term ‘loaded-PCP’ will be used to refer to PCPs modified with a PPant arm and loaded with a substrate. Indeed, a number of NRPSs have been structurally characterized in various productive conformations by employing a promiscuous PPtase, such as Sfp from *Bacillus subtilis* [17], for the loading of substrates, mimics or dead-end inhibitors, onto the megaenzymes [18–22].

The PCP domain delivers substrates to the catalytic domains. First, the adenylation (A) domain activates the incoming acid monomer. A domains select very diverse monomers including α -L- or α -D-amino acids, β -amino acids or aryl acids [23]. Subsequently, the condensation (C) domain catalyses peptide bond formation between two PCP-tethered monomers. Many optional tailoring domains that further chemically modify the metabolite under construction can also be found in NRPSs [24]. For example, the epimerization (E) domain converts natural L-amino acids into D-amino acids (modules 3 and 6 in figure 1a). Finally, each assembly line ends with a domain, such as a thioesterase

(TE) or a reductase (Re) (module 7 in figure 1*a*), that releases the final product. This final stage can introduce further diversity in the peptide as the release can occur either by hydrolysis or cyclization. Interestingly, since surfactin release occurs via macrolactonization [25,26], surfactin metabolites contain both amide and ester bonds, thus deserving the designation of 'depsipeptide' (figure 1*b*) [27].

NRPS flexibility allows the conformational changes required for the interactions between the PCP and the catalytic domains. However, this flexibility is a drawback to structural characterizations of large NRPS fragments and, accordingly, successful structural studies have often required the employment of chemical tools to reduce conformational heterogeneity [19–22]. Nevertheless, characterizing the movements of NRPS multienzymes is a requirement for the detailed understanding of these fascinating machineries. In this review, we focus on recent aspects of NRPS flexibility that allow PCP movements during a catalytic cycle, by describing both the successive conformations adopted by these enzymes during a cycle as well as movements that engender passage from one conformation to the next.

2. Known non-ribosomal peptide synthetase structures

The field of NRPS structural biology achieved major breakthroughs in the last five years with the publication of several crystal structures of multi-domain NRPSs, largely due to the work of the Schmeing and Gulick groups (figure 2) [19–22,28,29]. These results add to a high number of structures of individual catalytic domains and PCP-containing didomains, solved by NMR, X-ray crystallography or a combination of both techniques [8]. Each domain adopts the same fold in different structures, independent of the number of domains present in the protein. Nevertheless, understanding the organization of full modules as well as module–module interactions is essential to provide a better insight into these assembly lines. For a long time, however, the intrinsic flexibility of NRPSs prevented the structural characterization of full modules at high resolution. The first structure of a full module, that of the termination module of surfactin, SrfA-C, was solved by X-ray crystallography in 2008 (figures 1*a* and 2*a,b*) [30]. Since then, crystal structures of three other termination modules have been solved: AB3403 from an uncharacterized pathway of *Acinetobacter baumannii*, the enterobactin termination module EntF and ObiF1 from the obafluorin assembly line (figure 2*c–e*) [20,29,30]. In 2017, a combined effort in X-ray crystallography and negative staining electron microscopy (EM) provided insights into the structures of the last two modules of DhbF involved in the synthesis of bacillibactin (figure 2*f,g*) [21]. Lastly, the structural elucidation of the dimodular protein LgrA, from the gramicidin synthetase complex that contains an initiation and an elongation module, was a breakthrough in the comprehension of supra-modular NRPS organization. Indeed, 12 crystal structures of LgrA fragments complexed with substrates, substrate analogues and dead-end inhibitors were solved and provided a full picture of the catalytic cycle of a dimodular NRPS (figures 2*h–j* and 3) [19,22].

3. Loading of the amino acid onto the peptidyl carrier protein domain

The adenylation (A) domain is divided into two subdomains: the N-terminal subdomain, A_{core} , consists of around 400 amino acids, while the C-terminal subdomain, A_{sub} , comprises around 100 amino acids [31]. The A domain catalyses two reactions: the activation of the acid monomer using ATP (adenylation) and its subsequent transfer to the PCP domain (thiolation or thioesterification). The A domain is able to adopt several conformations that have been described as 'the domain alternation cycle' and which are supported by several structures of complete NRPS modules (figure 3) [19,20,22,31]. Remarkably, the thiolation state has been characterized multiple times using non-hydrolysable analogues (figure 3*b*, thiolation conformation) [18–20,22]. For example, the structure of PA1221, a natural A-PCP didomain NRPS from *Pseudomonas aeruginosa*, was obtained both in its apo form and in a loaded form locked in the thiolation conformation through the use of the inhibitor valyl-adenosine vinylsulfonamide (AVS) [32]. In the apo form, the electron density for the PCP domain was absent, suggesting the domain was flexible, whereas the whole didomain was visible in the presence of the AVS inhibitor, indicating that it stabilized the A-PCP interface. This suggests that the numerous crystal structures of NRPSs in the thiolation conformation do not reflect a preferentially adopted conformation *in vivo* but, more likely, a conformation that favours crystallization.

The cycle starts when the A domain, in an open conformation, is available for substrate binding (figure 3*b*, open conformation) [31]. The A_{core} subdomain contains the monomer binding pocket that accommodates ATP and the acid monomer. Upon substrate binding, a 30° rotation of the A_{sub} subdomain leads to the closed conformation, which is then suitable for adenylation (figure 3*b*, closed conformation) [19]. This conformation allows the entry of an A_{sub} loop into the A_{core} subdomain; this loop contains a conserved catalytic lysine that stabilizes the acid substrate and ATP [33]. Subsequently, after adenylation and pyrophosphate release, a 140° rotation of the A_{sub} subdomain allows the conversion between the closed and the thiolation conformations, the latter being able to catalyse thioesterification (figure 3*b*, thiolation conformation) [31].

The rotations of the A_{sub} subdomain are facilitated by a flexible hinge region containing a conserved aspartic acid or a lysine, located in the $A_{\text{core}}-A_{\text{sub}}$ linker [34]. The structure of an adenylate-forming enzyme with the hinge residue mutated into proline revealed an enzyme blocked in the adenylation conformation. Consistent with the structure, the mutant enzyme was still capable of adenylation, but not of thiolation [34]. The importance of the hinge residue has also been demonstrated in the context of the multi-domain NRPS EntF, in which the same hinge residue mutation abolished enterobactin production [35]. Therefore, the flexibility of the hinge residue in the $A_{\text{core}}-A_{\text{sub}}$ linker is essential to allow the movement of the A_{sub} subdomain relative to the A_{core} , which is necessary to allow conformational changes of the whole module. Indeed, rotations of the A_{sub} subdomain drive movement of the PCP owing to the linker connecting the two domains. Analysis of the A-PCP linker region reveals that it contains multiple prolines, absent in standalone A domains [35]. These prolines might rigidify

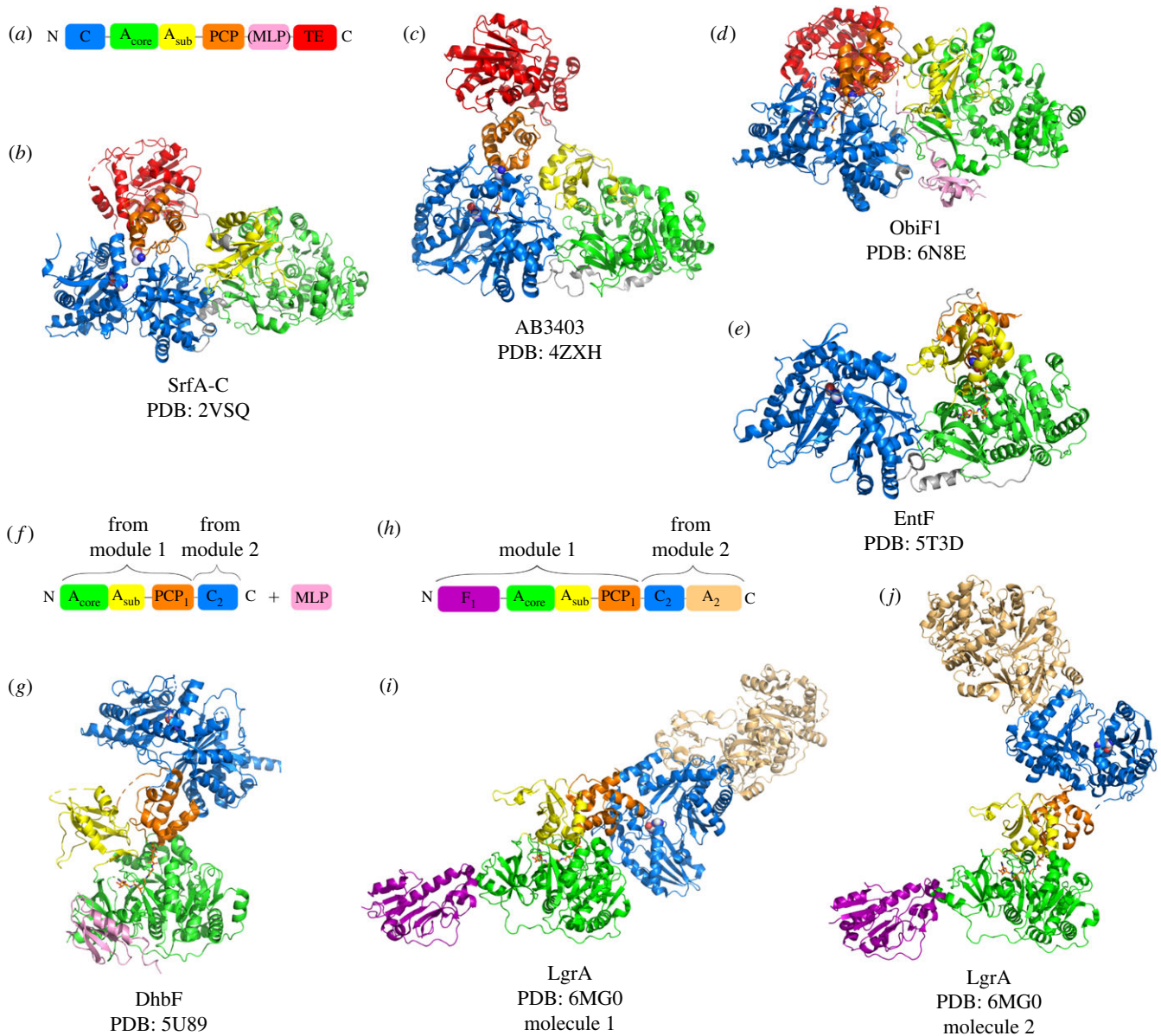


Figure 2. Structural gallery of NRPS modules. If present, the PPant arm attached to the PCP domain is represented as sticks. (a,b) Domain organization and crystal structures of the termination modules SrfA-C (PDB code: 2VSQ) [19], (c) AB-3403 (PDB code: 4ZXH) [21], (d) ObiF1 (PDB code: 6N8E) [22] and (e) EntF (PDB code: 5T3D) [21]. In addition to the PCP domain, these modules are composed of condensation (c), adenylation (a) and thioesterase (TE) domains. No electron density was detected for the TE domain of EntF. The catalytic His of the C domain is represented in spheres. (f,g) Domain organization and crystal structure of DhbF, a cross-module construct (PDB code: 5U89) [16]. (h–j) Domain organization and crystal structures of a five-domain construct of LgrA (PDB code: 6MG0) [24]. LgrA starts with a formylation (f) domain. The domain in pink in ObiF1 (d) and DhbF (g) represents MLP, an A domain activator.

the A-PCP linker, thus facilitating movements of the PCP domain in concert with the movements of the A_{sub} subdomain [35].

4. Modification of the peptidyl carrier protein-tethered amino acid

The vast diversity of NRPs arises in part from the action of tailoring domains, such as cyclization (Cy), epimerization (E), formylation (F), ketoreductase (KR), methyltransferase (Met) and oxidase (Ox) domains, that modify the peptide under construction [19,36–40]. It is not unusual for tailoring domains to be inserted within A domains, that are then called interrupted A domains [39,41]. The structure of LgrA in the formylation conformation provides insight into the mechanism of amino acid modification by a tailoring

domain, in this case the formylation of PCP-bound L-valine by an F domain using a formyltetrahydrofolate cofactor (figure 3b, formylation conformation) [19]. Since the F and PCP domains are separated by the large A domain (figure 3a), a substantial conformational change must occur to allow the PCP domain to position its PPant arm in the F active site. The A_{sub} subdomain rotates 180° from its position in the thiolation conformation while the PCP domain rotates 75°, thus moving 60 Å away from its thiolation position (compare figure 3b, thiolation and formylation conformations). The interaction surface between the F and PCP domains is very limited. Unfortunately, there is no structure of the full dimodular LgrA in the formylation conformation. It is worth noting that the structure of dimodular LgrA (figure 3b, condensation conformation) reveals that the position occupied by the PCP₁ domain in the formylation conformation is occupied by the C domain at a later stage of the catalytic cycle [22].

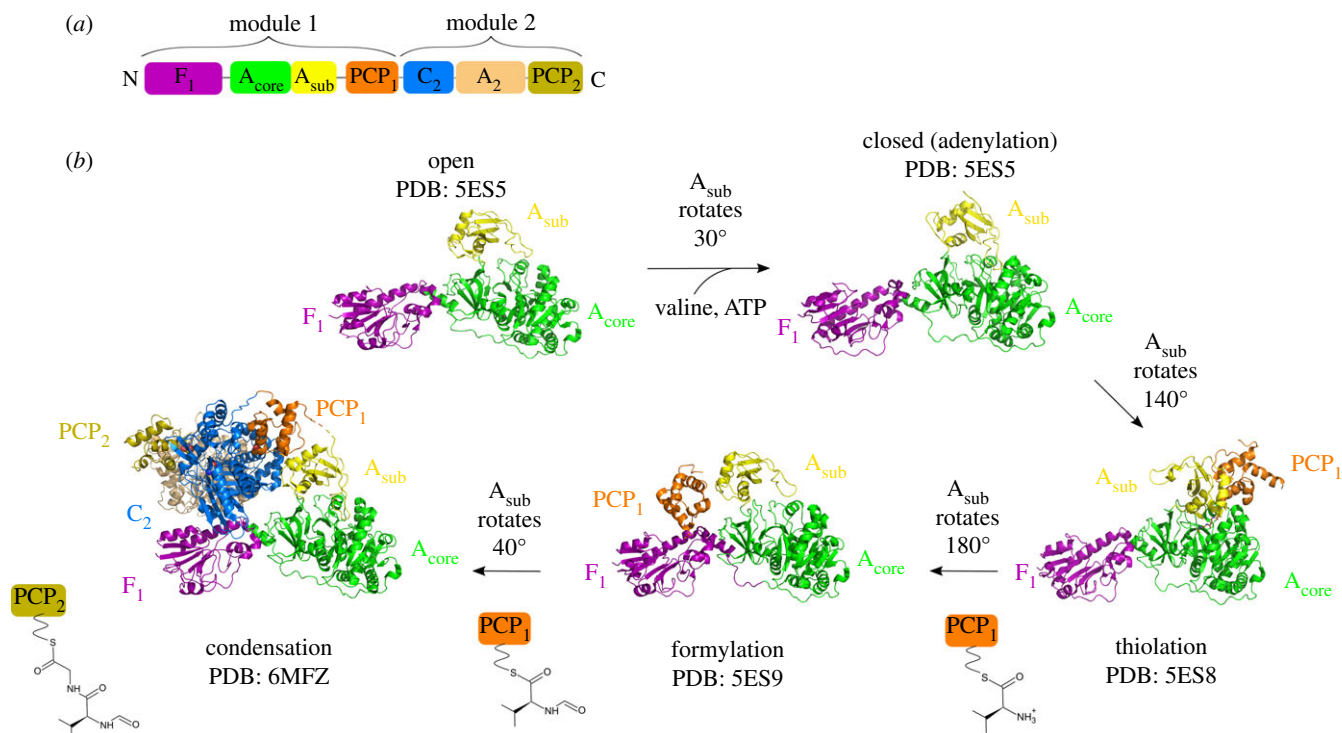


Figure 3. Catalytic cycle of LgrA, a dimodular NRPS. (a) Domain organization of LgrA. For clarity, the inactive epimerization domain of LgrA, that follows the PCP₂ domain, is not shown. (b) Catalytic cycle of LgrA illustrated by five crystal structures (PDB codes: 5ES5, 5ES8, 5ES9 and 6MFZ) [23,24]. Four different structures of module 1 reveal details about the catalytic cycle of an initiation module. The PCP₁ domain is disordered in the open and closed states. First, the open state allows binding of valine and ATP. The closed state is the conformation which is relevant for the activation of valine by ATP (i.e. the adenylation state). The thiolation conformation captures the transfer of valine from the A₁ domain to the PPant arm of the PCP₁ domain. In the formylation conformation, the F₁ domain adds a formyl group to valine, still attached to the PCP₁ domain. The structure of the full dimodular NRPS allows the visualization of the condensation conformation. After condensation, Gly activated by module 2 is covalently bound to formyl-Val by a peptide bond. In all presented structures, the F₁-A_{core} bidomain adopts similar conformations whereas the A_{sub} subdomain and the PCP₁ domain are positioned differently. Rotations of the A_{sub} subdomain induce movements of the PCP₁ domain due to the linker between them.

5. Elongation of the donor peptide chain with an acceptor amino acid

Elongation is the only reaction that necessitates interactions between domains belonging to different modules. This reaction is catalysed by a condensation (C) domain or, more rarely, by a cyclization (Cy) also called heterocyclization (HC) domain, that catalyses cyclization after condensation [42]. If the two modules belong to the same polypeptide, PCP and C domains are directly connected by a linker (figure 3*b*, condensation conformation). If they belong to different polypeptides, docking domains facilitate the interaction between the upstream PCP and the downstream C domain [43]. The C domain of an elongation module n catalyses peptide bond formation between the growing chain carried by the upstream PCP domain (donor PCP), that belongs to module $n-1$, and the activated amino acid carried by the downstream PCP domain (acceptor PCP) located on the same module n (figure 3*b*, condensation conformation) [42]. The growing peptide chain is directly transferred from one PCP domain to the next, without attachment to the C domain. Thus, the two PCP domains must bind simultaneously to two different binding sites on the C domain; these are referred to as donor and acceptor binding sites. However, the C domain must discriminate between the two PCP domains to maintain the directionality of the assembly line. Therefore, the condensation reaction results from the interaction between three domains (the C, the donor PCP and the acceptor PCP domains).

The C domain adopts the V-shaped pseudo-dimeric fold seen in the chloramphenicol acyltransferase family [42]. It is divided into N-terminal and C-terminal lobes, and the active site is located inside the N-terminal lobe, at the centre of a tunnel formed by the interface between the two lobes (figure 2). The structures of three termination modules, SrfA-C, AB3403 and ObiF1, show the acceptor PCP domain docked onto the C domain acceptor site (figure 2*a-d*) [20,29,30]. In SrfA-C, although the PCP domain is in its apo form due to a Ser-Ala mutation, the Ala is located 16 Å away from the catalytic His of the C domain, suggesting that this structure is compatible with a productive condensation reaction (figure 2*b*) [30]. Both holo-AB3403 and holo-ObiF1 show the PCP domain docked onto the C acceptor site with the PPant arm inserted through a tunnel that allows the positioning of the final thiol in proximity to the catalytic His of the C domain (figure 2*c,d*) [20,30]. It is worth noting that, as opposed to what was observed for the PCP interaction with the A domain, no substrate or inhibitor was needed to favour the interaction between the acceptor PCP and C domains.

Several dimodular LgrA structures have revealed for the first time the productive interaction between a donor PCP domain and its corresponding C domain (figure 3*b*, condensation conformation) [22]. Indeed, in four structures of LgrA, the donor PCP₁ domain is docked in the C₂ donor site, presenting its conserved Ser towards the catalytic His of the C domain, the two residues being separated by less than 20 Å. The donor PCP₁ domain globally has the same

orientation in these structures and three of them show electron density for the PPant arm in the donor tunnel of the C domain. However, the structure of F_1 - A_1 -PCP₁-C₂ with f-Val loaded onto the PPant arm of PCP₁ reveals that the reactive thioester group must slightly modify its position in order to be properly positioned for attack by the acceptor amino acid. The authors hypothesized that the donor substrate could only be correctly positioned in the presence of the PCP-bound-acceptor substrate in the active site of the C domain [22]. The importance of the loading status of the carrier protein domain for megaenzyme conformation has already been demonstrated for the modular PKS megaenzymes that incorporate domains analogous to the C, A and PCP of the NRPS systems [44,45]. Indeed, structures of the PikAIII PKS module loaded with various substrates obtained by cryo-electron microscopy revealed that the acyl carrier protein (ACP) domain adopts dramatically different positions according to the nature of the substrate loaded onto the ACP.

A detailed comprehension of the condensation reaction requires structural information on a NRPS including at least a donor PCP, an acceptor PCP and a condensation domain. The structure of holo-LgrA F_1 - A_1 -PCP₁-C₂-A₂-PCP₂ reveals both donor and acceptor PCPs docked onto a single C domain (figure 3*b*, condensation conformation) [22]. 29 Å separate the two PCP Ser residues loaded with their 20 Å-long PPant arms. The Ser from the acceptor PCP₂ is located 15 Å away from the catalytic His of the C domain while the Ser from the donor PCP₁ is 18 Å away from it. These distances are compatible with the proximity between substrates required for nucleophilic attack. Unfortunately, there is no electron density for the PCP PPant arms so the detailed interaction of donor and acceptor substrates cannot be deduced from this structure.

The detailed view of PCP-bound-substrates in the condensation conformation can be obtained using a mechanism-based probe, recently designed by the Gulick and Aldrich groups [46]. Although the structure of LgrA F_1 - A_1 -PCP₁-C₂-A₂-PCP₂ shows that an inhibitor is not necessary to lock a dimodular NRPS in the condensation conformation, this new chemical probe stabilized the interaction between the donor PCP, the acceptor PCP and the C domains. The enterobactin assembly line served as a model to prove the functionality of this probe [46]. The authors were able to mimic the PPant arm loaded with the natural substrate of the donor PCP by replacing the whole acyl-thioester portion of the substrate by a non-hydrolysable analogue incorporating a ketone functionality, thus preventing the release of the loaded substrate. The resulting crypto-PCP was shown to bind to the donor site of the C domain and these results allowed the construction of a model where the crypto-PCP inserts its unnatural PPant arm into the donor tunnel. The authors assumed that the pantetheine probe would then react with the natural acceptor substrate loaded on the acceptor PCP, forming an imine bond instead of the natural peptide bond formed between substrates. In this configuration, both PCP domains should be docked onto the C domain and linked together via their PPant arms connected through an imine bond. Therefore, this probe should help stabilize the interaction between the C and PCP domains during condensation and could lead to more crystal structures of bimodular NRPS locked in the condensation conformation.

The structure of the LgrA PCP₁ domain has been solved in association with its three catalytic partners (A_1 , F_1 and C_2 domains, figure 3*b*); therefore, the comparison of the three crystal structures provides insights into the conformational changes that allow the PCP₁ domain to shuttle between its three partners [19,22]. As described above, the large movements required for the PCP domain to reach its different catalytic partners are mainly driven by conformational changes of the A_{sub} subdomain that are transferred to the PCP domain by the A_{sub} -PCP rigid linker. Shifting from the formylation to the condensation conformation, the PCP₁ domain must cross 30 Å, achieved through a rotation of 40° of the A_{sub} subdomain (compare figure 3*b*, formylation and condensation conformations). Similarly, after condensation, the PCP₁ domain must detach from the C domain and travel back 50 Å to return to the A_1 active site (compare figure 3*b*, condensation and thiolation conformations), achieved by a rotation of 150° of the A_{sub} subdomain. Even in the absence of structures that show the second LgrA module in the thiolation conformation, we can easily extrapolate that the movements seen in module 1 could be similar in module 2.

Interestingly, a NRPS module can start a second catalytic cycle before the first one is complete [20]. For example, the structure of LgrA in the condensation conformation (figure 3*b*) shows the PCP₁ domain in the peptide donation conformation while the A_1 domain is in the closed conformation and can thus catalyse adenylation [22]. After adenylation, the aminoacyl-AMP is tightly sequestered in the A domain active site in the absence of the available PCP [47,48]. Subsequently, the A_1 domain can catalyse thiolation as soon as the C_2 domain has catalysed condensation which will liberate the PCP₁ domain. This decoupling between different domain activities likely increases the synthesis rate of NRPSs.

6. Release of the peptidyl carrier protein-tethered peptide

The structures of four termination modules harbouring a TE domain (C-A-PCP-TE) are now available, i.e. SrfA-C, AB3403, EntF and ObiF1 (figure 2*a-e*) [20,28–30]. In all four crystal structures, the TE positions are dramatically different (figure 2*a-d*), suggesting that the TE domain is most probably a mobile element. Negative staining EM images of EntF showed the TE domain in various positions compared to the other domains and no density was observable for the TE domain in the corresponding crystal structure (figure 2*e*), confirming that the EntF TE domain can adopt multiple conformations [20]. The ObiF1 module has an unusual domain organization, since the TE domain is followed by a MbtH-like protein (MLP) [30]. Interestingly, in these conditions, the MLP domain anchors the TE domain to the module (figure 2*d*). These elements suggest that the high mobility of the TE domain is due to the flexibility of the short PCP-TE linker and to the fact that, in general, no successive domain imposes structural restraints on the final TE domain. As none of the structures of these four termination modules revealed the interaction between the PCP and the TE domains, it was characterized through the crystal structure of the EntF PCP-TE didomain [49]. The authors used a phosphopantetheinyl-based inhibitor loaded onto the PCP domain that stabilized the transient interaction between the

PCP and TE domains [49], thus providing details of a productive PCP–TE interaction.

7. Non-ribosomal peptide synthetases flexibility at the supramodular scale, unrelated to the catalytic cycle

As described in the previous sections, NRPS flexibility allows the conformational rearrangements that are required for the PCP domain to interact with its catalytic partners. It is then legitimate to wonder whether NRPS flexibility is only restricted to the movements that shuttle PCP-tethered substrates to the different NRPS active sites or if there is flexibility at the supramodular scale, unrelated to the catalytic cycle. In other words, are there architectural rules governing the relationships between successive modules, or is their relationship random? In addition to the crystal structure of bimodular LgrA, structural models of multimodular NRPSs derive from low resolution techniques or from the combination of crystal structures. Indeed, in 2016, Marahiel and co-workers proposed a helical model for a hypothetical 7-module NRPS assembly by combining the C-A-PCP structure of the SrfA-C termination module (figure 2*b*) with the PCP-C cross-module structure of TycC from the tyrocidine synthetase [30,50,51]. The helical axis was occupied by the PCP domains and each module was rotated by 120° relative to the previous one.

Several EM observations indicated that NRPSs probably adopt a more flexible architecture than the helical model mentioned above. An early negative staining EM observation of a fungal 11-module NRPS, responsible for the synthesis of cyclosporin, pictured this 1.7 MDa machinery as an assembly of globular moieties, most likely modules, that could adopt either very compact or elongated structures [52]. It led to the hypothesis that NRPS modules are arranged as ‘beads on a chain’, suggesting that an NRPS assembly line would not adopt any specific architecture.

More recently, the dimodular NRPS DhbF (C₁-A₁-PCP₁-C₂-A₂-PCP₂+MLP) was also observed by negative staining EM [21]. Despite the presence of AVS inhibitors that limited its conformational heterogeneity, DhbF adopted a continuum of conformations as diverse as an elongated shape, an L shape or a very compact shape. Most particles could be sorted into five classes that differed by the relative positions of the first module in relation to the second one. Therefore, although the flexibility inside a module is limited due to the stable conformation adopted by the C-A didomain, it seems that there are few limitations to the position one module can adopt relatively to the adjacent one. These data favour an irregular architecture for NRPSs; however, the fact that the number of classes is limited to five suggests that the supramodular architecture of NRPSs is not completely random. In the crystal structure of the A₁-PCP₁-C₂ cross-module (figure 2*f,g*), there is no density for the PCP₁-C₂ linker, suggesting that it could be flexible [21]. Therefore, the movements of the PCP₁-C₂ linker combined with the absence of strong intermodule interactions could explain the various conformations adopted by the dimodular DhbF.

The six recent dimodular LgrA crystal structures provided further evidence that NRPSs do not adopt a unique stable architecture but rather a few conformations among a myriad of possibilities [22]. One striking example confirmed

the flexibility of the PCP₁-C₂ intermodule linker. Indeed, the F₁-A₁-PCP₁-C₂-A₂ variant was crystallized in the thiolation conformation for module 1 using a Val-AVS inhibitor and two molecules were found in the asymmetric unit (figure 2*h-j*). Within these two molecules, module 1 is identical but module 2 adopts two radically different positions. This behaviour results in two strikingly different LgrA shapes, reminiscent of the two DhbF structures observed by EM, one that is elongated and the other L-shaped. Therefore, it seems that locking one module in a specific conformation does not impose a unique conformation on the adjacent module. The most convincing evidence that NRPSs adopt a flexible architecture was obtained from the SAXS analyses of the LgrA F₁-A₁-PCP₁-C₂-A₂ construct [22]. They indicated that the conformations adopted in the crystal structures do not exactly reflect the conformations adopted in solution. To better estimate these, Reimer and co-workers used the ensemble optimization method to generate different models that took into account flexibility parameters [53]. The ensemble generated fit very well with the experimental data, thus confirming the flexibility of LgrA. However, it cannot be excluded that LgrA flexibility is only apparent and is an effect of the absence of the other components of the assembly line. Indeed, in addition to LgrA, the linear gramicidin NRPS is composed of three other proteins [54] that could restrain the conformations that LgrA can adopt.

8. Concluding remarks

The structural and functional studies of NRPS multienzymes are not limited to providing details regarding the production of complex metabolites but can also be applied to the discovery of new antibiotics. Indeed, the products of these machineries are often essential for bacterial virulence, hence targeting their biosynthesis is a promising strategy to fight microbial pathogens [6]. For instance, the multi-drug resistant *Klebsiella pneumoniae* uses several siderophores for iron acquisition, including the NRPs enterobactin and yersiniabactin [55–57]. Strains deficient for yersiniabactin production are much less virulent than the wild-type strains [58], suggesting that the yersiniabactin NRPS machinery could be potentially be explored as an antibacterial development target.

Moreover, the re-engineering of NRPS megaenzymes in order to produce new medically relevant molecules is of particular interest [59,60]. This prospect exists since the discovery of the modular organization of NRPSs [61]. To date, a straightforward strategy to re-engineer NRPS assembly lines to produce artificial peptides has been difficult to establish, although some successful reports of re-engineering were published [62–65]. Classical strategies using substitutions of A, C-A, PCP-C-A units or entire modules yielded only a small amount of synthesized peptide [59,60]. Recently, the Bode group successfully exploited a novel exchange strategy, using A-PCP-C exchange units (XUs) by fixing the borders of the XU within the flexible C and A domain linker [66]. They subsequently improved their strategy by dividing the C domain, placing the borders of the XU within the flexible linker that connects the N-terminal acceptor and C-terminal donor subdomains of C, yielding C_{Acc}-A-PCP-C_{Don} (XUs) [67]. This strategy allowed the authors to produce very high yields of novel NRPS peptides, paving the way for new biotechnological approaches that could optimize the production

of novel bioactive compounds through NRPS engineering. Therefore, an increased knowledge on the supramodular architecture of NRPSs, especially regarding the linker regions that allow enzyme flexibility, raises interesting perspectives for natural product re-engineering.

Data accessibility. This article has no additional data.

Authors' contributions. S.B., A.D. and P.M. wrote the paper together.

Competing interests. We declare we have no competing interests.

Funding. This study was funded by Fondation Innovations en Infectiologie (grant no. AO12-20), Ecole nationale Supérieure de Lyon and Ligue Contre le Cancer (grant no. M32917). Work on NRPS/PKS enzymes in the Bacterial Pathogenesis Laboratory is supported by grants from Le Comité de Savoie de la Ligue Contre le Cancer (to P.M.) and the FINOVI foundation (to A.D.). S.B. is the recipient of a PhD fellowship from the Ecole Nationale Supérieure (ENS) de Lyon.

Acknowledgements. The IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG, CEA).

References

- Cragg GM, Newman DJ. 2013 Natural products: a continuing source of novel drug leads. *Biochim. Biophys. Acta.* **1830**, 3670–3695. (doi:10.1016/j.bbagen.2013.02.008)
- Felnagle EA, Jackson EE, Chan YA, Podevels AM, Berti AD, McMahon MD, Thomas MG. 2008 Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **5**, 191–211. (doi:10.1021/mp700137g)
- Galm U, Wendt-Pienkowski E, Wang L, Huang S-X, Unsinn C, Tao M, Coughlin JM, Shen B. 2011 Comparative analysis of the biosynthetic gene clusters and pathways for three structurally related antitumor antibiotics: bleomycin, tallysomycin, and zorbamycin. *J. Nat. Prod.* **74**, 526–536. (doi:10.1021/np1008152)
- Menkhaus M, Ullrich C, Kluge B, Vater J, Vollenbroich D, Kamp RM. 1993 Structural and functional organization of the surfactin synthetase multienzyme system. *J. Biol. Chem.* **268**, 7678–7684. (doi:10.1016/S0021-9258(18)53010-6)
- Singh M, Chaudhary S, Sareen D. 2017 Non-ribosomal peptide synthetases: identifying the cryptic gene clusters and decoding the natural product. *J. Biosci.* **42**, 175–187. (doi:10.1007/s12038-017-9663-z)
- Gulick AM. 2017 Nonribosomal peptide synthetase biosynthetic clusters of ESKAPE pathogens. *Nat. Prod. Rep.* **34**, 981–1009. (doi:10.1039/C7NP00029D)
- Vassaux A, Meunier L, Vandenbol M, Baurain D, Fickers P, Jacques P, Leclère V. 2019 Nonribosomal peptides in fungal cell factories: from genome mining to optimized heterologous production. *Biotechnol. Adv.* **37**, 107449. (doi:10.1016/j.biotechadv.2019.107449)
- Izoré T, Cryle MJ. 2018 The many faces and important roles of protein-protein interactions during non-ribosomal peptide synthesis. *Nat. Prod. Rep.* **35**, 1120–1139. (doi:10.1039/C8NP00038G)
- Jaremko MJ, Davis TD, Corpuz JC, Burkart MD. 2020 Type II non-ribosomal peptide synthetase proteins: structure, mechanism, and protein-protein interactions. *Nat. Prod. Rep.* **37**, 355–379. (doi:10.1039/C9NP00047J)
- Hertweck C. 2009 The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688–4716. (doi:10.1002/anie.200806121)
- Miyana A, Kudo F, Eguchi T. 2018 Protein-protein interactions in polyketide synthase-nonribosomal peptide synthetase hybrid assembly lines. *Nat. Prod. Rep.* **35**, 1185–1209. (doi:10.1039/C8NP00022K)
- Crosby J, Crump MP. 2012 The structural role of the carrier protein—active controller or passive carrier. *Nat. Prod. Rep.* **29**, 1111–1137. (doi:10.1039/C2np20062g)
- Koglin A *et al.* 2006 Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science* **312**, 273–276. (doi:10.1126/science.1122928)
- Frueh DP, Arthanari H, Koglin A, Vosburg DA, Bennett AE, Walsh CT, Wagner G. 2008 Dynamic thiolation-thioesterase structure of a non-ribosomal peptide synthetase. *Nature* **454**, 903–906. (doi:10.1038/nature07162)
- Jaremko MJ, Lee DJ, Opella SJ, Burkart MD. 2015 Structure and substrate sequestration in the Pyoluteorin type II peptidyl carrier protein PtlL. *J. Am. Chem. Soc.* **137**, 11 546–11 549. (doi:10.1021/jacs.5b04525)
- Beld J, Sonnenschein EC, Vickery CR, Noel JP, Burkart MD. 2014 The phosphopantetheinyl transferases: catalysis of a post-translational modification crucial for life. *Nat. Prod. Rep.* **31**, 61–108. (doi:10.1039/C3NP70054B)
- Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C. 2001 Biosynthesis of complex polyketides in a metabolically engineered strain of *E. coli*. *Science* **291**, 1790–1792. (doi:10.1126/science.1058092)
- Gulick AM, Aldrich CC. 2018 Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. *Nat. Prod. Rep.* **35**, 1156–1184. (doi:10.1039/C8NP00044A)
- Reimer JM, Aloise MN, Harrison PM, Schmeing TM. 2016 Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. *Nature* **529**, 239–242. (doi:10.1038/nature16503)
- Drake EJ, Miller BR, Shi C, Tarrasch JT, Sundlov JA, Allen CL, Skiniotis G, Aldrich CC, Gulick AM. 2016 Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* **529**, 235–238. (doi:10.1038/nature16163)
- Tarry MJ, Haque AS, Bui KH, Schmeing TM. 2017 X-Ray crystallography and electron microscopy of cross- and multi-module nonribosomal peptide synthetase proteins reveal a flexible architecture. *Structure* **25**, 783–793.e4. (doi:10.1016/j.str.2017.03.014)
- Reimer JM, Eivaskhani M, Harb I, Guarné A, Weigt M, Schmeing TM. 2019 Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, eaaw4388. (doi:10.1126/science.aaw4388)
- Kudo F, Miyana A, Eguchi T. 2019 Structural basis of the nonribosomal codes for nonproteinogenic amino acid selective adenylation enzymes in the biosynthesis of natural products. *J. Ind. Microbiol. Biotechnol.* **46**, 515–536. (doi:10.1007/s10295-018-2084-7)
- Walsh CT, Chen H, Keating TA, Hubbard BK, Losey HC, Luo L, Marshall CG, Miller DA, Patel HM. 2001 Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. *Curr. Opin. Chem. Biol.* **5**, 525–534. (doi:10.1016/S1367-5931(00)00235-0)
- Tsang CC, Bruner SD, Kohli RM, Marahiel MA, Walsh CT, Sieber SA. 2002 Characterization of the surfactin synthetase C-terminal thioesterase domain as a cyclic depsipeptide synthase. *Biochemistry* **41**, 13 350–13 359. (doi:10.1021/bi026592a)
- Bruner SD, Weber T, Kohli RM, Schwarzer D, Marahiel MA, Walsh CT, Stubbs MT. 2002 Structural basis for the cyclization of the lipopeptide antibiotic surfactin by the thioesterase domain SrtE. *Structure* **10**, 301–310. (doi:10.1016/S0969-2126(02)00716-5)
- Alonzo DA, Schmeing TM. 2020 Biosynthesis of depsipeptides, or Depsi: the peptides with varied generations. *Protein Sci.* **29**, 2316–2347. (doi:10.1002/pro.3979)
- Miller BR, Drake EJ, Shi C, Aldrich CC, Gulick AM. 2016 Structures of a nonribosomal peptide synthetase module bound to Mbth-like proteins support a highly dynamic domain architecture. *J. Biol. Chem.* **291**, 22 559–22 571. (doi:10.1074/jbc.M116.746297)
- Kreitler DF, Gemmell EM, Schaffer JE, Wenciewicz TA, Gulick AM. 2019 The structural basis of N-acyl- α -amino- β -lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat. Commun.* **10**, 3432. (doi:10.1038/s41467-019-11383-7)
- Tanovic A, Samel SA, Essen L-O, Marahiel MA. 2008 Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* **321**, 659–663. (doi:10.1126/science.1159850)

31. Gulick AM. 2009 Conformational dynamics in the Acyl-CoA synthetases, adenylation domains of non-ribosomal peptide synthetases, and firefly luciferase. *ACS Chem. Biol.* **4**, 811–827. (doi:10.1021/cb900156h)
32. Mitchell CA, Shi C, Aldrich CC, Gulick AM. 2012 Structure of PA 1221, a nonribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry* **51**, 3252–3263. (doi:10.1021/bi300112e)
33. Branchini BR, Murtiashaw MH, Magyar RA, Anderson SM. 2000 The role of lysine 529, a conserved residue of the acyl-adenylate-forming enzyme superfamily, in firefly luciferase. *Biochemistry* **39**, 5433–5440. (doi:10.1021/bi9928804)
34. Wu R, Reger AS, Lu X, Gulick AM, Dunaway-Mariano D. 2009 The mechanism of domain alternation in the acyl-adenylate forming ligase superfamily member 4-chlorobenzoate: coenzyme A ligase. *Biochemistry* **48**, 4115–4125. (doi:10.1021/bi9002327)
35. Miller BR, Sundlov JA, Drake EJ, Makin TA, Gulick AM. 2014 Analysis of the linker region joining the adenylation and carrier protein domains of the modular nonribosomal peptide synthetases. *Proteins* **82**, 2691–2702. (doi:10.1002/prot.24635)
36. Dowling DP, Kung Y, Croft AK, Taghizadeh K, Kelly WL, Walsh CT, Drennan CL. 2016 Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl Acad. Sci. USA* **113**, 12 432–12 437. (doi:10.1073/pnas.1608615113)
37. Chen W-H, Li K, Guntaka NS, Bruner SD. 2016 Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases. *ACS Chem. Biol.* **11**, 2293–2303. (doi:10.1021/acscchembio.6b00332)
38. Alonzo DA, Chiche-Lapierre C, Tarry MJ, Wang J, Schmeing TM. 2020 Structural basis of keto acid utilization in nonribosomal depsipeptide synthesis. *Nat. Chem. Biol.* **16**, 493–496. (doi:10.1038/s41589-020-0481-5)
39. Mori S, Pang AH, Lundy TA, Garzan A, Tsodikov OV, Garneau-Tsodikova S. 2018 Structural basis for backbone N-methylation by an interrupted adenylation domain. *Nat. Chem. Biol.* **14**, 428–430. (doi:10.1038/s41589-018-0014-7)
40. Pang B, Chen Y, Gan F, Yan C, Jin L, Gin JW, Petzold CJ, Keasling JD. 2020 Investigation of indigoidine synthetase reveals a conserved active-site base residue of nonribosomal peptide synthetase oxidases. *J. Am. Chem. Soc.* **142**, 10 931–10 935. (doi:10.1021/jacs.0c04328)
41. Labby KJ, Watsula SG, Garneau-Tsodikova S. 2015 Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. *Nat. Prod. Rep.* **32**, 641–653. (doi:10.1039/C4NP00120F)
42. Bloudoff K, Schmeing TM. 2017 Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta Proteins Proteom.* **1865**, 1587–1604. (doi:10.1016/j.bbapap.2017.05.010)
43. Smith HG, Beech MJ, Lewandowski JR, Challis GL, Jenner M. In press. Docking domain-mediated subunit interactions in natural product megasynth(et)ases. *J. Ind. Microbiol. Biotechnol.*
44. Dutta S *et al.* 2014 Structure of a modular polyketide synthase. *Nature* **510**, 512–517. (doi:10.1038/nature13423)
45. Whicher JR *et al.* 2014 Structural rearrangements of a polyketide synthase module during its catalytic cycle. *Nature* **510**, 560–564. (doi:10.1038/nature13409)
46. Shi C, Miller BR, Alexander EM, Gulick AM, Aldrich CC. 2020 Design, synthesis, and biophysical evaluation of mechanism-based probes for condensation domains of nonribosomal peptide synthetases. *ACS Chem. Biol.* **15**, 1813–1819. (doi:10.1021/acscchembio.0c00411)
47. Keating TA, Suo Z, Ehmann DE, Walsh CT. 2000 Selectivity of the yersiniabactin synthetase adenylation domain in the two-step process of amino acid activation and transfer to a holo-carrier protein domain. *Biochemistry* **39**, 2297–2306. (doi:10.1021/bi992341z)
48. Ehmann DE, Shaw-Reid CA, Losey HC, Walsh CT. 2000 The EntF and EntE adenylation domains of *Escherichia coli* enterobactin synthetase: sequestration and selectivity in acyl-AMP transfers to thiolation domain cosubstrates. *Proc. Natl Acad. Sci. USA* **97**, 2509–2514. (doi:10.1073/pnas.040572897)
49. Liu Y, Zheng T, Bruner SD. 2011 Structural basis for phosphopantetheinyl carrier domain interactions in the terminal module of nonribosomal peptide synthetases. *Chem. Biol.* **18**, 1482–1488. (doi:10.1016/j.chembiol.2011.09.018)
50. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen L-O. 2007 Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* **15**, 781–792. (doi:10.1016/j.str.2007.05.008)
51. Marahiel MA. 2016 A structural model for multimodular NRPS assembly lines. *Nat. Prod. Rep.* **33**, 136–140. (doi:10.1039/C5NP00082C)
52. Hoppert M, Gentsch C, Schörgendorfer K. 2001 Structure and localization of cyclosporin synthetase, the key enzyme of cyclosporin biosynthesis in *Tolypocladium inflatum*. *Arch. Microbiol.* **176**, 285–293. (doi:10.1007/s002030100324)
53. Tria G, Mertens HDT, Kachala M, Svergun DI. 2015 Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCr* **2**, 207–217. (doi:10.1107/S205225251500202X)
54. Kessler N, Schuhmann H, Morneweg S, Linne U, Marahiel MA. 2004 The linear pentadecapeptide gramicidin is assembled by four multimodular nonribosomal peptide synthetases that comprise 16 modules with 56 catalytic domains. *J. Biol. Chem.* **279**, 7413–7419. (doi:10.1074/jbc.M309658200)
55. Shankar C, Veeraraghavan B, Nabarro LEB, Ravi R, Ragupathi NKD, Rupali P. 2018 Whole genome analysis of hypervirulent *Klebsiella pneumoniae* isolates from community and hospital acquired bloodstream infection. *BMC Microbiol.* **18**, 6. (doi:10.1186/s12866-017-1148-6)
56. Peltier F, Choquet M, Decroix V, Adjidi CC, Castelain S, Guiheneuf R, Pluquet E. 2019 Characterization of a multidrug-resistant *Klebsiella pneumoniae* ST607-K25 clone responsible for a nosocomial outbreak in a neonatal intensive care unit. *J. Med. Microbiol.* **68**, 67–76. (doi:10.1099/jmm.0.000884)
57. Remya PA, Shanthi M, Sekar U. 2019 Characterisation of virulence genes associated with pathogenicity in *Klebsiella pneumoniae*. *Indian J. Med. Microbiol.* **37**, 210–218. (doi:10.4103/ijmm.IJMM_19_157)
58. Lawlor MS, O'connor C, Miller VL. 2007 Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect. Immun.* **75**, 1463–1472. (doi:10.1128/IAI.00372-06)
59. Brown AS, Calcott MJ, Owen JG, Ackerley DF. 2018 Structural, functional and evolutionary perspectives on effective re-engineering of non-ribosomal peptide synthetase assembly lines. *Nat. Prod. Rep.* **35**, 1210–1228. (doi:10.1039/C8NP00036K)
60. Bozhüyük KA, Micklefield J, Wilkinson B. 2019 Engineering enzymatic assembly lines to produce new antibiotics. *Curr. Opin. Microbiol.* **51**, 88–96. (doi:10.1016/j.mib.2019.10.007)
61. Cane DE, Walsh CT, Khosla C. 1998 Harnessing the biosynthetic code: combinations, permutations, and mutations. *Science* **282**, 63–68. (doi:10.1126/science.282.5386.63)
62. Kegler C, Bode HB. 2020 Artificial splitting of a non-ribosomal peptide synthetase by inserting natural docking domains. *Angew. Chem. Int. Ed. Engl.* **59**, 13 463–13 467. (doi:10.1002/anie.201915989)
63. Calcott MJ, Owen JG, Ackerley DF. 2020 Efficient rational modification of non-ribosomal peptides by adenylation domain substitution. *Nat. Commun.* **11**, 4554. (doi:10.1038/s41467-020-18365-0)
64. Tietze A, Shi Y-N, Kronenwerth M, Bode HB. 2020 Nonribosomal peptides produced by minimal and engineered synthetases with terminal reductase domains. *Chembiochem* **21**, 2750–2754. (doi:10.1002/cbic.202000176)
65. Ruijne F, Kuipers OP. 2021 Combinatorial biosynthesis for the generation of new-to-nature peptide antimicrobials. *Biochem. Soc. Trans.* **49**, 203–215. (doi:10.1042/BST20200425)
66. Bozhüyük KAJ, Fleischhacker F, Linck A, Wesche F, Tietze A, Niesert C-P, Bode HB. 2018 De novo design and engineering of non-ribosomal peptide synthetases. *Nat. Chem.* **10**, 275–281. (doi:10.1038/nchem.2890)
67. Bozhüyük KAJ *et al.* 2019 Modification and de novo design of non-ribosomal peptide synthetases using specific assembly points within condensation domains. *Nat. Chem.* **11**, 653–661. (doi:10.1038/s41557-019-0276-z)