A natural prodrug activation mechanism in the biosynthesis of nonribosomal peptides

Daniela Reimer and Helge B. Bode*

Received 29th August 2013
DOI: 10.1039/c3np70081j
www.rsc.org/npr

Cite this: DOI: 10.1039/c3np70081j

Highlight

Introduction

Many secondary metabolites with biologically interesting activities are produced by multienzyme thiotemplate mechanisms using nonribosomal peptide synthetases (NRPS) and the fatty acid synthase (FAS)-related polyketide synthases (PKS) or by a hybrid biosynthesis thereof, which have been extensively reviewed.1 On the basis of their assembly line organization, NRPSs and type I PKSs are large multienzyme complexes that refer to reactions for peptide bond formations and Claisen-type condensations, which are organized in defined highly conserved catalytically active modules and domains.2,3 They are responsible for the steps required for the recognition and activation of the substrates, covalent binding of the building blocks to the catalytic domains, peptide/polyketide formation of intermediates but also for the modification and release of the final product. The structure of natural products derived from these assembly lines does not always follow the domain composition of the enzymes involved4 and thus structure predictions based on the assembly lines encoded by the biosynthesis gene cluster can be misleading. Examples are PKS and NRPS derived secondary metabolites not following the collinearity rule (that the order of biochemical steps follows the chromosomal order of the underlying genes in the biosynthesis gene cluster)5,6,7 or biosyntheses that include an iterative usage or the skipping of discrete modules.8–10 This Highlight gives an overview of the recently identified natural prodrug activation mechanism found in the biosynthesis of NRPS and/or PKS derived natural products showing some similarity to mechanisms formerly only known from ribosomally synthesized peptides.11 This activation mechanism demonstrates not only another origin for a lack of correlation between domain composition of the multienzyme complex and the final product structure but furthermore adds another layer of complexity to natural product biosynthesis, also of thiotemplate derived natural products.

Xenocoumacin: the first example of natural prodrug structures

The most abundant secondary metabolites in Xenorhabdus nematophila, an entomopathogenic bacterium symbiotically associated with Steinernema nematodes,12 are the antibiotically active xenocoumacins. They belong to the class of dihydroisocoumarin-derived compounds13 and are thought to be involved in killing bacteria in the insect gut,14 and soil living bacteria like the Gram-positive Bacillus subtilis, thereby eliminating other food competitors for the Xenorhabdus bacteria and its nematode host. The gene cluster encoding xenocoumacin biosynthesis was identified in X. nematophila ATCC 19061 (ref. 15) and its detailed annotation led to the prediction of a non-collinear hybrid PKS/NRPS pathway, which was first predicted to be partially inactive as no clear correlation between the xenocoumacin structure and the enzymes encoded in the biosynthesis gene cluster could be made.16,17 Surprisingly, the results of a deletion of the peptidase encoding gene xcnG bridged the gap between the size of the natural product and the biosynthetic pathway as larger xenocoumacin derivatives named prexenocoumacins were produced as first biosynthetic derivatives.

The current hypothesis is that the inactive prexenocoumacin B (1) and four additional derivatives, differing only in the fatty acid moiety, are produced by the biosynthetic enzymes inside the cytoplasm (Fig. 1a, exemplified for prexenocoumacin B). XcnG, a bifunctional protein with a periplasmic peptidase domain and three additional transmembrane helices, cleaves the acylated ω-asparagine residue (2) from all prexenocoumacin derivatives, resulting in the production of only one compound. The resulting bioactive xenocoumacin 1 (XCN 1) (3), which could be produced in E. coli supplemented with 1 and expressing xcnG,18 is transported by an as yet unidentified ABC transporter into the periplasm and presumably by a TolC-like...
protein through the outer membrane (Fig. 1a). As homologues of the peptidase (XcnG) and the encoding NRPS C-Ax-T-E (for condensation, adenylation specific for Asx, thiolation and epimerization) starting module (XcnA) were identified in several different bacterial genera, a widespread and important
mechanism for the activation of related natural products was suggested.

**Amicoumacin: the missing link between structure and strategy**

Two homologues of XcnG and the corresponding NRPS for the starting module were identified in different *Bacillus* spp.\(^{15,21}\) In the case of the structurally related amicoumacin biosynthetic pathway, a peptidase (Bpum_0630) with the same domain architecture as XcnG, so called type I domain architecture (peptidase domain and three transmembrane helices) is present.\(^{15}\) Recently, the predicted amicoumacin prodrugs were identified in a marine *B. subtilis*\(^{22}\) indicating a similar prodrug activation mechanism as found in *X. nematophilus*. The identified prodrugs are structurally similar to amicoumacin C (4) as they harbor an N-terminal asparagine or glutamine residue that is extended by a 9-methylundecanoic or 9-methyldecanoic acid. As in the case of prexenocoumacin/xenocoumacin, these so called lipoamicoumacins exhibit no antibacterial activity, whereas amicoumacin is a good antibiotic.\(^{23}\) Unfortunately, the authors failed to make a link between their isolated structures and the prodrug mechanism.

**Zwittermicin: two different types of prodrug peptidases**

Several *Bacillus* species produce zwittermicin, which shows a broad spectrum of activities including antibiotic activity against Gram-positive and -negative microorganisms.\(^{23}\) The biosynthetic machinery includes for the production of zwittermicin A (ZmA, 5), a peptidase ZmaM with type II architecture\(^{15}\) (Fig. 1b). Although to date, no pre-structure has been detected, it is assumed that the prezwittermicin acyl-α-Asn-ZmA (6) contains an N-terminal fatty acid and a α-asparagine moiety comparable to the prexenocoumacins as it can be deduced from the NRPS encoded in the gene cluster.\(^{15,24}\) In contrast to XcnG, ZmaM harbors the peptidase domain fused to the ABC-transporter domain with the nine transmembrane helices (termed type II domain architecture) required for secretion as described above. Similarly, in *Xenorhabdus bovienii* a cryptic biosynthetic gene cluster with a ZmaM-like peptidase (XbJ1_2693) was identified. The XcnG homologue XbJ1_2693 was able to cleave prexenocoumacin, indicating that the so far unknown substrate shows the same or at least a strong structural similarity to the N-terminus of prexenocoumacin. In contrast, ZmaM and Bpum_0630 were not able to cleave prexenocoumacin.\(^{15}\) These results indicated that not only the α-asparagine is responsible for the specificity but also the adjacent amino acid is necessary and therefore different classes of recognizing specific substrates might exist.

**Colibactin: detailed characterization of the involved peptidase**

Colibactin is probably the most interesting compound produced by a NRPS/PKS hybrid harboring a XcnG peptidase homologue named ClbP.\(^{25}\) The colibactin genomic island is distributed across isolates of commensal and extraintestinal pathogenic *Escherichia coli* strains (ExPEC) but is restricted to the phylogenetic groups B1 and B2.\(^{26}\) Moreover, there is evidence that the colibactin genomic island plays a role in colitis-associated colorectal cancer and inflammatory bowel disease.\(^{27,28}\) Surprisingly, it is also found in *E. coli* Nissle 1917,\(^{29}\) sold as probiotic under the brand Mutaflu® for almost 100 years. Highly conserved regions encoding colibactin biosynthesis were also found in the enterobacterial species *Klebsiella pneumoniae*, *Enterobacter aerogenes* and *Citrobacter koseri* and, interestingly, they are associated with versiniabactin coding regions.\(^{30}\) Colibactin causes DNA double strand breaks in mammalian cells and activates the DNA damage G2 checkpoint in the cell cycle leading to a cell cycle arrest in the G2/M transition, resulting in a nucleus enlargement called megalocytosis.\(^{26}\)

Detailed characterization of ClbP and XcnG revealed a strong structural homology to class C β-lactamases (AmpC). Peptidases of this class are periplasmic inner membrane proteins harboring an N-terminal signal sequence, a peptidase catalytic domain and three C-terminal transmembrane helices.\(^{15,25}\) The N-terminal signal sequence is responsible for the translocation of the peptidase domain into the periplasm and all transmembrane helices are necessary for catalytic activity, although only one transmembrane is needed for anchoring in the inner membrane.\(^{15,25}\) Mutational studies of the actin-like cytoskeleton MreB and the major proteins (SRP/SecY/YidC) of the translocation/insertion pathway normally used by inner membrane proteins have revealed an impaired or absent translocation process. This system recognizes the N-terminal signal sequence of ClbP and cleaves the protein at the time of translocation into the periplasmic site and inserts the enzyme into the inner membrane.\(^{21,22}\) The peptidase domain of ClbP harbors two structural domains, an α/β-region with seven stranded antiparallel β-sheets surrounded on both sides with six α-helices and three β-strands and an all α-region with four helices. Its catalytic pocket is located between the two structural domains and possesses the conserved motifs SxxK and YxS, typical for serine-type α-Ala–α-Ala carboxypeptidases of the MEROPS S12 enzyme family.\(^{31}\) A striking feature in contrast to AmpC, is the unusually large catalytic pocket, which could be interpreted as an adaption to a specific type of substrates. Furthermore, the ClbP catalytic pocket exhibits a highly negatively-charged surface especially at the two ends of the pocket. AmpC and other related S12 enzymes showing a positive potential, which play an important role in recognizing the
negatively charged substrates, e.g., cephalothin. In ClbP, D334 and D336, part of the C-terminal helix distant from the catalytic triade, might be the key residues responsible for the negative potential and are important for the peptidase bioactivity. Positively-charged natural products like prezwittermicin (6) match perfectly with the negative potential residues in natural prodrug peptidases. For ClbP, the catalytic triad comprises serine, lysine and tyrosine and six residues [E159, S188, H257, F316, G328, N331], which might be involved in substrate binding. These residues were identified on the basis of docking studies with imipenem, a substrate of the β-lactamase class. Nevertheless, involvement of these residues has to be demonstrated directly in vitro with the natural compound precolibactin. Although the complete structure of colibactin is still unknown, recently N-myristoyl-α-asparagine (7), as a free N-terminal moiety of precolibactin, could be characterized in vivo and isolated from E. coli Nissle 1917 cultures as well as from heterologous expression of the colibactin biosynthetic gene cluster. This is in agreement with an in vitro characterization of the prodrug scaffold as a C12 or C14 N-acyl α-asparagine linked to L-alanine or L-valine.

Didemnin: flexibility of N-terminal moiety

In the Gram-negative marine bacterium Tistrella mobilis, a symbiont of tunicates, another related but not analogous mechanism was found in the NRPS/PKS hybrid biosynthesis of the didemnins. Didemmins X and Y (8) are produced as acylglutamine ester derivatives by the NRPS/PKS enzyme complex and are cleaved in the extracellular space by so far unknown proteins into the active didemnin B (9). Domain names: A, adenylation; PCP, peptidyl carrier protein; C, condensation; E, epimerization.

Other related deacylation and activation mechanisms

To date, further strategies of deacylation of natural compound precursors differing from the xenocoumacin mechanism have been found in pyoverdine, zeamine and saframycin biosynthesis. During pyoverdine biosynthesis by Pseudomonas aeruginosa, a myristate moiety is removed by a periplasmic hydrolase (PvdQ) in the maturation pathway. For the PKS/NRPS/FAS hybrid derived zeamine from Serratia plymuthica, a post-assembly activation is postulated via an acyl-aminoacyl peptidase, cleaving short N-acylated peptides. In the biosynthesis of saframycin and other tetrahydroisoquinoline (THIQ) antibiotics (e.g. euteinascidin, quinocarcin) the acyl chain is used for a condensation (C) domain mediated Pictet-Spengler like reaction in forming the THIQ moiety, followed by deacylation as the last important step to generating the bioactive compound.

Formerly, the activation of compounds was only known as a common feature in the biosynthesis of ribosomal peptides, where the bioactive peptide is often derived by proteolytic cleavage, but was not expected to play a role in nonribosomally synthesized peptides. The NRPS associated activation mechanism exhibits a remarkable similarity to the maturation process in ribosomal peptides. Microcin J25, for example, is produced as a propeptide in E. coli and a cleavage of the N-terminal leader sequence by ATP-catalyzed proteases and successive maturation steps such as cyclization, result in the active peptide, which is then exported by ABC transporters and an outer membrane TolC protein.

Conclusion

In summary, all of the examples mentioned here highlight the importance of different strategies for activation of NRPS and/or PKS derived natural products to avoid self-destruction of the producer strain during antibiotic production. Thus, the mechanism described in this Highlight adds to the numerous other strategies of bacteria to cope with the production of antibiotics.
This mechanism, which uses a D-asparagine specific carboxypeptidase, is widespread among different bacterial taxa and thus adds a new layer of complexity to natural products biosynthesis and related mechanisms must be taken into account for genome mining approaches.\(^{41}\) When compound identification is based only on analysis of the biosynthetic gene cluster, mistakes in the prediction of the actually produced natural product might occur. Interestingly, such activation mechanisms seemed to have evolved independently, as different activation strategies have been found across different genera.

**Acknowledgements**

The authors are grateful to the DFG that funded the work leading to this highlight. H. B. B. is especially grateful to the European Research Council for a starting grant (grant agreement no. 311477).

**Reference**