



Cite this: *Nat. Prod. Rep.*, 2025, **42**, 739

## Fatty acyl-AMP ligases in bacterial natural product biosynthesis

Anne Liong <sup>ab</sup> and Pedro N. Leão <sup>\*a</sup>

Covering: covering up to 2024

Fatty Acyl-AMP Ligases (FAALs) belong to the family of adenylate-forming enzymes and activate fatty acyl substrates through adenylation. FAALs were discovered as key players in various natural product biosynthetic pathways, particularly in the assembly of polyketides and non-ribosomal peptides. These enzymes exhibit a conserved structural architecture that distinguishes them from their close relatives, the Fatty Acyl-CoA Ligases. FAALs display the starter unit in the biosynthesis of diverse natural products where they shuttle fatty acyl substrates into secondary metabolism for further chain elongation and/or modification. In this review, we cover the discovery, distribution and structure of FAALs as well as their role in natural product biosynthesis. In addition, we provide an overview about their genomic and biosynthetic contexts and summarize approaches used to analyze FAAL activity, predict their substrate specificity and to discover new compounds whose biosyntheses involve these enzymes.

Received 7th December 2024

DOI: 10.1039/d4np00073k

rsc.li/npr

- 1 **Introduction – adenylate-forming enzymes**
- 2 **Discovery, distribution, and structure of FAALs**
- 2.1 **Discovery of fatty acyl-AMP ligases**
- 2.2 **Distribution of FAALs in bacteria and eukaryotes**
- 2.3 **Structural analysis of FAALs compared to FACLs**
- 3 **Role of FAALs in natural products biosynthesis**
- 3.1 **Genomic context of FAALs and their ACPs in bacteria**
- 3.2 **Stand-alone FAALs**
- 3.3 **Didomain FAALs**
- 3.4 **FAALs as part of multidomain enzymes**
- 3.5 **Biochemical characterization of FAAL domains**
- 3.5.1 **Phosphate release assays**
- 3.5.2 **Radiolabelling of substrates**
- 3.5.3 **Mass spectrometry-based methods**
- 4 **Deorphanization of FAAL-containing BGCs and future applications**
- 4.1 **Diversity and discovery of FAAL-containing orphan BGCs**
- 4.2 **Substrate specificity and prediction**
- 5 **Conclusions**
- 6 **Data availability**
- 7 **Author contributions**
- 8 **Conflicts of interest**
- 9 **Acknowledgements**
- 10 **References**

### 1 Introduction – adenylate-forming enzymes

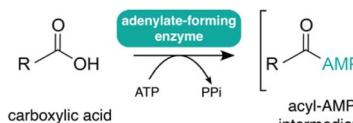
In nature, fatty acids are usually activated before being assimilated into various metabolic pathways. Adenylation represents one of the most important reactions in biology, initiating the activation of fatty acids and other substrates in diverse metabolic pathways. The catalytic mechanism of the adenylation reaction was discovered over 60 years ago. Within this mechanism, carboxylate substrates are activated through condensation with adenosine triphosphate (ATP) setting free a highly reactive acyl adenylate (acyl-AMP) species. After reaction of a nucleophile (amide, alcohol or thiol), the final product is formed with release of AMP (Fig. 1).<sup>1</sup>

Across all three domains of life, adenylation is fulfilled by a superfamily of widely spread adenylate-forming enzymes (ANL). These are involved in diverse biochemical pathways such as protein synthesis, posttranslational modifications, amino acid- or fatty acid metabolism. A global analysis of protein family domains showed that ANL enzymes are the third most abundant domain in known natural products biosynthetic pathways.<sup>2</sup> ANL enzymes can be divided into different classes based on their structural and catalytic properties.<sup>3</sup> Class I includes acyl/aryl-CoA ligases (ACSSs and ACLs), adenylation domains (A-domains) as part of non-ribosomal peptide synthetases (NRPSs), luciferases, aryl polyene adenylation enzymes,<sup>4</sup>  $\beta$ -lactone synthetases<sup>5</sup> and the most recently discovered group: fatty-acyl AMP ligases (FAALs).<sup>6</sup> Class II comprises

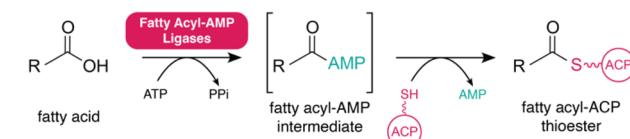
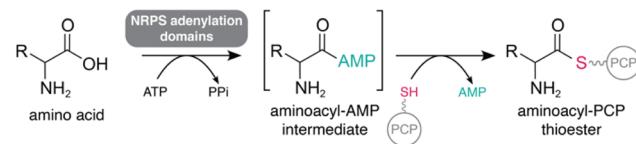
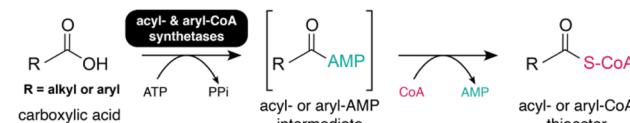
<sup>a</sup>CIIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Matosinhos, Portugal. E-mail: pleao@ciimar.up.pt

<sup>b</sup>ICBAS – School of Medicine and Biomedical Sciences, University of Porto, Porto, Portugal





**Fig. 1** General reaction scheme for adenylate-forming enzymes. A carboxylic acid is activated by an adenylate-forming enzyme to form an acyl-AMP intermediate. After nucleophilic attack, an ester, thioester or amide is formed with the release of AMP.



aminoacyl-tRNA synthetases. Class III consists of adenylating enzymes involved in siderophore synthesis.<sup>7</sup> All three classes of enzymes are dependent on  $Mg^{2+}$  for their biological activity with the number of ions varying among them.<sup>8,9</sup> Compared to class II (acting on amino acids) and class III (acting on dicarboxylic acid substrates), class I enzymes show promiscuity for a wider range of substrates such as fatty acids, amino acids, benzoic acids and their derivatives. The latter also do not show structural homology to class II or III enzymes.<sup>3</sup> All members of class I (except luciferases) are characterized by the transfer of an adenylate to a 4'-phosphopantetheine (4'-Ppant) moiety, which happens through thioester bond formation with CoA or holo-acyl carrier protein (holo-ACP).<sup>10</sup>

In this review, we focus on FAALs, enzymes (or domains) found in secondary metabolite biosynthetic gene clusters (BGCs), as an atypical enzyme system within the ANL superfamily. FAALs transfer activated fatty acyl substrates to a holo-ACP (Fig. 2), and prime the biosynthesis for a diversity of natural product families, including lipopeptides, different polyketide classes, as well as other lipids. We cover current knowledge about FAAL distribution among the tree of life, following their discovery in *Mycobacterium tuberculosis* (Mtb) and summarize recent research on FAAL structure and how this relates to their unique function. We then describe examples of

**Fig. 2** Reaction mechanism of ANL enzymes (acyl- and aryl-CoA synthetases, NRPS adenylation domains and fatty acyl-AMP ligases). All reactions include the formation of an acyl-AMP intermediate, followed by nucleophilic attack and formation of a thioester with release of AMP.

bacterial biosynthetic pathways that involve FAALs, of the breadth of structures that can be generated in this way, and also of how FAAL activity can be ascertained. Finally, we showcase methods that have and can be used to uncover metabolites produced by FAAL-containing pathways.

## 2 Discovery, distribution, and structure of FAALs

### 2.1 Discovery of fatty acyl-AMP ligases

FAALs were characterized for the first time in Mtb by Arora and colleagues<sup>6</sup> through bioinformatics approaches that included



Anne Liong

at ICBAS, University of Porto. Her research project includes the deorphanization of biosynthetic gene clusters (BGCs) from cyanobacteria and focuses on the discovery of new fatty acid-harboring natural products and their underlying biosynthesis and enzymology.



Pedro N. Leão

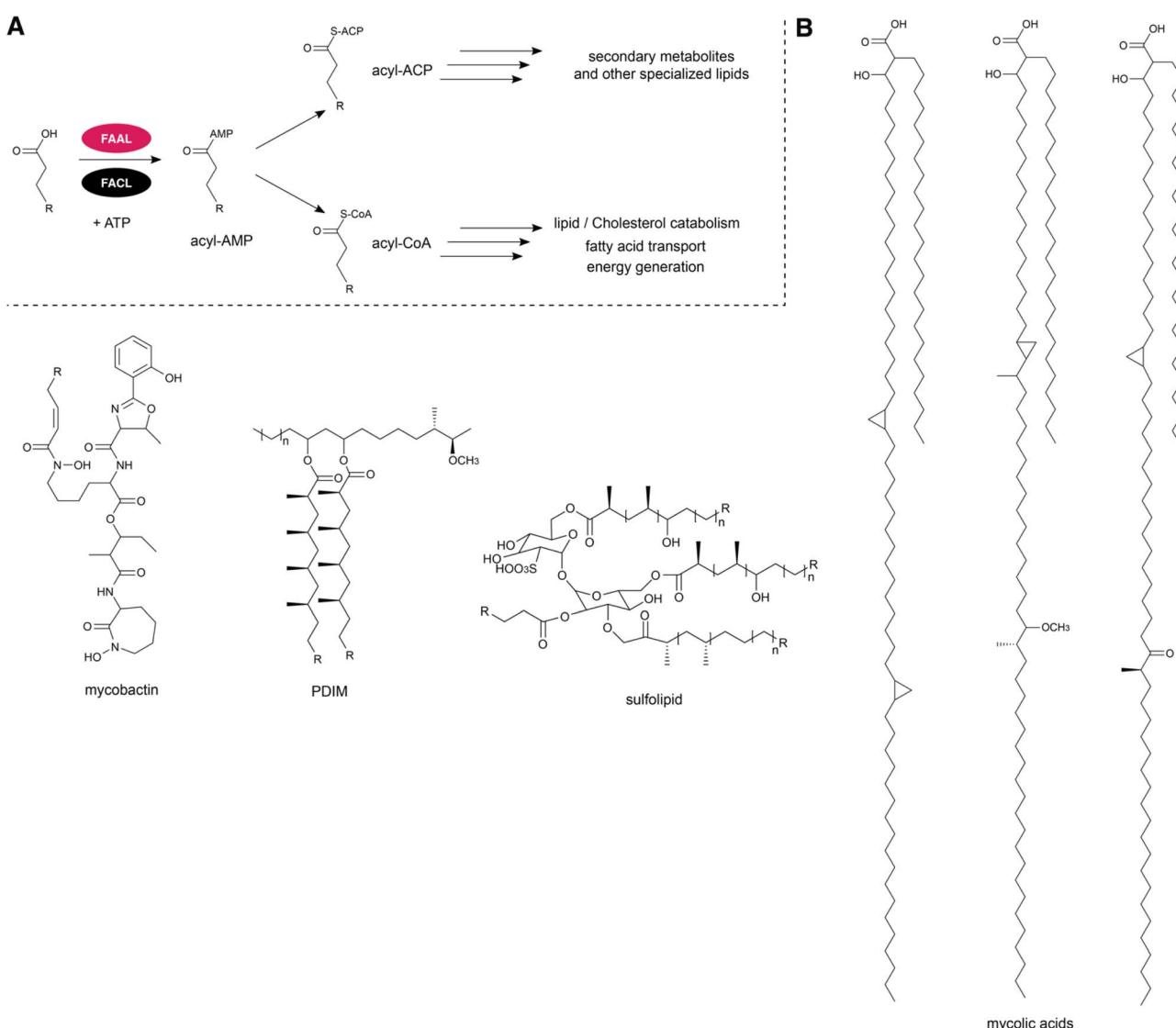
Pedro Leão obtained his PhD from the University of Porto in 2010 under the supervision of Prof. Vitor Vasconcelos. For his postdoctoral studies at CIIMAR (University of Porto), Pedro focused on marine natural product chemistry and biosynthesis with research stays at the Gerwick lab (Scripps Institution of Oceanography) and Balskus group (Harvard University). In 2018 he started his group at CIIMAR, focusing on the chemistry, biosynthesis and chemical ecology of cyanobacterial secondary metabolites.



analysis of the *Mtb* genome. In *Mtb*, Fatty Acyl-CoA Ligases (FAALs) participate in lipid and cholesterol catabolism, fatty acid transport and energy generation while FAALs are involved in the biosynthesis of various essential complex virulence-conferring lipids such as sulfolipids, mycobactin and mycolic acids (Fig. 3A). Those metabolites are essential for the survival and virulence of the bacterium under harsh environmental conditions.<sup>11</sup> For example, in the cell envelope of mycobacteria, the biosynthesis of mycolic acids is dependent on the presence of FAALs linking the incorporation of fatty acids with polyketide synthesis.

Mycolic acids are produced by a two-step catalytic reaction that begins with the activation of fatty acid carboxyl residues by adenosine triphosphate (ATP) hydrolysis, followed by a transfer to the acceptor molecule holo-ACP.<sup>12</sup> In *Mtb*, fatty acids are transferred by a group of FAALs onto ACP domains belonging to

type I polyketide synthases (PKSs) to produce lipidic metabolites.<sup>13</sup> For example, FAAL32 (also known as FadD32) is required for the activation of a meromycolic acid chain in mycolic acid biosynthesis and thus essential for mycobacterial growth.<sup>14</sup> Other FAALs such as FAAL26, FAAL28 and FAAL29 are involved in the biosynthesis of complex polyketides and glycolipids such as phthiocerol dimycocerosates (PDIMs) and phenolic glycolipids (PGLs). PDIMs are major virulence lipids in the cell wall of *Mtb* and are involved in the immune evasion strategies of *Mtb* helping the pathogen with survival within host macrophages.<sup>15</sup> More precisely, the biosynthesis of PDIM is encoded by a 44 kb cluster that consists of five modular PKSs where FadD26 activates long chain fatty acids and transfers them onto the first ACP domain of the adjacent PKS enzyme PpsA. Thus, biosynthesis of the phthiocerol moiety is fulfilled by crosstalk between a FAAL (FadD26) and the *pps* modular PKS cluster.<sup>16,17</sup> To



**Fig. 3** FAALs and FACLs function in *Mtb*. (A) FAALs use fatty acids to generate mycobacterial secondary metabolites while FACLs use them to generate fatty acyl-CoA followed by utilization in lipid- and cholesterol catabolism, for fatty acid transport and energy generation. (B) Structures of mycobacterial metabolites that involve FAALs in their biosynthesis: mycobactin, PDIM, sulfolipid and mycolic acids.

generate PDIM, mycocerosic acid moieties are installed by FAAL28. Interestingly, the PGL cluster in *Mtb* contains two FAALs, FAAL29 and FAAL22. While FAAL29 shows overlapping substrate profile for long chain fatty acids, FAAL22 has unusual specificity for *p*-hydroxybenzoic acid. The biosynthesis of phthiocerol or the phenolphthiocerol chains for DIMS and PGLs includes the loading of *p*-hydroxybenzoic acid and synthesis of *p*-hydroxyphenylalkanoic acid that is used as a substrate for FAAL29 and loaded onto PpsA-E for elongation.<sup>6,18,19</sup> Besides FAAL22 and FAAL29 that assemble the phenolphthiocerol lipid part, FAAL28 is needed for the installation of the two mycocerosic acid units to form PGLs (Fig. 3B).<sup>20</sup>

Hence, the discovery and characterization of FAALs in *Mtb* significantly advanced the understanding of complex lipid biosynthesis pathways crucial for the virulence of this human pathogen and underlines FAALs as main target for antitubercular drug development. Concomitantly, it set the stage for understanding the diversity of these enzymes in other organisms.

## 2.2 Distribution of FAALs in bacteria and eukaryotes

FAALs are present in both bacteria and eukaryotes, where they play essential roles in lipid metabolism.<sup>21</sup> In bacteria, FAAL members are not only encoded near PKS I gene clusters as stand-alone proteins but also at the N-terminus of NRPSs.<sup>22</sup> These two classes of large multimodular enzymes can together catalyze condensation reactions of more than 500 different monomers including  $\beta$ -keto acids, proteinogenic and non-proteinogenic amino acids, fatty acids and  $\alpha$ -hydroxy acids leading to a variety of diverse bioactive molecules.<sup>23</sup> Genome predictions on the genome of *Mtb* led to the identification of over 18 FAAL domains, several of them with so-far unexplored functions.<sup>24</sup> FAAL-encoding genes have been identified in bacteria, notably in actinobacteria, cyanobacteria and proteobacteria as well as in eukaryotes, most being genetically located near ACPs.<sup>25</sup>

Looking at the distribution in bacteria, FAALs have been most extensively studied in *Mtb* but are also present in other members of the actinobacteria phylum. Homologues could be identified in mycolic acid-producing actinobacteria such as *Nocardia farcinica* and *Rhodococcus* sp. RHA1,<sup>6</sup> as well as in *Corynebacterium glutamicum*,<sup>26</sup> but also lipopeptide-producing bacteria such as *Bacillus subtilis* and several cyanobacteria encode FAALs or proteins with FAAL domains.<sup>27,28</sup>

When compared to bacteria, FAALs are relatively rare in eukaryotes. Still, a set of FAAL sequences could be identified in eukaryotes such as fungi, algae, plants, stramenopiles and alveolates.<sup>25</sup> Interestingly, in the genome of *Homo sapiens*, proteins named disco-interacting protein 2 homologue C (DIP2C) harbouring two FAAL-like domains have been identified. Still, these protein domains showed only 25.0% and 26.75% of sequence identity with well characterized FAALs from *Legionella pneumophila*.<sup>29</sup> FAAL sequences have also been identified in metagenome assembled genomes (MAGs) obtained from coastal permeable sediment samples – FAALs were present in the MAGs of the archaeal taxa of *Halobacteriarchaeon* (ncbi

accession numbers NNJ93261.1 and NNJ93262.1).<sup>21</sup> The presence of FAALs in archaea might be interesting out of several reasons such as evolutionary significance, metabolic diversity and adaptability of these microorganisms. Archaea often inhabit extreme environments such as those with high temperatures, high salt concentrations or acidic conditions and this has been linked to their differentiated lipid profile, when compared to bacterial or eukaryotic lipids (e.g. ether-linked vs. ester-linked lipids).<sup>30</sup> The presence of FAAL in all kingdoms of life may suggest an early evolution and presence in LUCA (Last Universal Common Ancestor). The existing literature does not investigate further how FAAL genes evolved.

In summary, FAALs are primarily found in bacteria, especially within the phylum actinobacteria. Although their presence in eukaryotes and archaea is more restricted, the FAAL domain seems to be utilized across the tree of life in lipid metabolism.

## 2.3 Structural analysis of FAALs compared to FACLs

Before the discovery of key differences in amino acid sequence and structure, several FAALs have been wrongly annotated as FACLs. For example, this was the case for FadD10, which is involved in the biosynthesis of a virulence-related lipopeptide in *Mtb*.<sup>31</sup> Among 34 FadD proteins identified in the *Mtb* genome, bioinformatics analysis followed by biochemical characterization suggested that 12 of these proteins could be classified into the group of FAALs.<sup>12</sup> Analysis of the 3D-structures of the enzymes proved useful for classification. The first insights into FAALs 3D-structure were obtained by solving a crystal structure of the N-terminal domain of FAAL28 from *Mtb*. Due to difficulties in crystallization, only the N-terminal structure of the protein could be obtained *via* X-ray crystallography.<sup>6</sup> The first full-sequence crystal structures of FAALs were obtained from *Escherichia coli*, EcFAAL and *L. pneumophila*, LpFAAL, and included the acyl adenylates bound in their active sites (lauroyl-adenylate and myristoyl-adenylate in EcFAAL; lauroyl-adenylate in LpFAAL).<sup>32</sup> Ever since, various 3D-structures of FAALs from different bacteria were made available to the public (Table 1).

Overall, FAALs share a common domain organization, namely an N-terminal domain that contains an adenylation domain and a C-terminal domain. The N-terminal domain is responsible for binding and recognizing fatty acid substrates and catalyzes the acyl transfer to the ACP. More precisely, the adenylation domain catalyzes the conversion of fatty acids into acyl-AMPs by binding ATP and facilitating the formation of the acyl-AMP intermediate while the C-terminal domain is involved in enzyme stabilization and interacts with the ACP to mediate docking and alignment for the acyl-transfer.<sup>12,33</sup>

FAALs share structural features with FACLs that ligate fatty acids to Coenzyme A (CoA) *via* a thioester bond while FAALs load fatty acids onto their cognate ACP. Surprisingly, studies showed the preference of FAALs for holo-ACP but not for CoA although both contain the protruding Ppant group. It was proposed that the FAAL specific insertion (FSI), a stretch of several amino acids in the N-terminus of the FAALs but not in the FACLs, might be preventing domain rotations resulting in

Table 1 List of FAALs that have been structurally characterized via X-ray crystallography

Protein	Organism	PDB-ID	Substrate	Reference
FadD28 N-terminal	<i>M. tuberculosis</i>	3E53	Long-chain fatty acids	6
FadD28 N-terminal G330W	<i>M. tuberculosis</i>	3T5A	Long-chain fatty acids	33
FadD13 N-terminal	<i>M. tuberculosis</i>	3T5B, 3T5C, 5ZRN	Long-chain fatty acids	33
FadD23 N-terminal	<i>M. tuberculosis</i>	8HCZ	Mainly C16:0	34
FadD10	<i>M. tuberculosis</i>	4ISB	Mainly C14:0	31
FadD23	<i>M. tuberculosis</i>	8IQU	Mainly C16:0	34
FadD32	<i>M. tuberculosis</i>	5HM3	Mainly C14:0	35
EcFAAL	<i>E. coli</i>	3PBK	C12:0, C14:0	32
LpFAAL	<i>L. pneumophila</i>	3KXW, 3LNV	C12:0	32
FAAL domain R336A of PKS	<i>Picosynechococcus</i> sp. PCC 7002	7R7F	Unknown	To be published
FAAL domain A229/R336A of PKS	<i>Picosynechococcus</i> sp. PCC 7002	7R7G	Unknown	To be published
FAAL domain of PKS	<i>Picosynechococcus</i> sp. PCC 7002	7R7E	Unknown	To be published
FadD32	<i>Mycobacterium marinum</i>	5EY9	C12:0, C14:0	36
FadD32	<i>Mycobacterium smegmatis</i>	5EY8	C12:0, C14:0	36
FAAL	<i>Mycobacterium smegmatis</i>	5D6N	Unknown	37

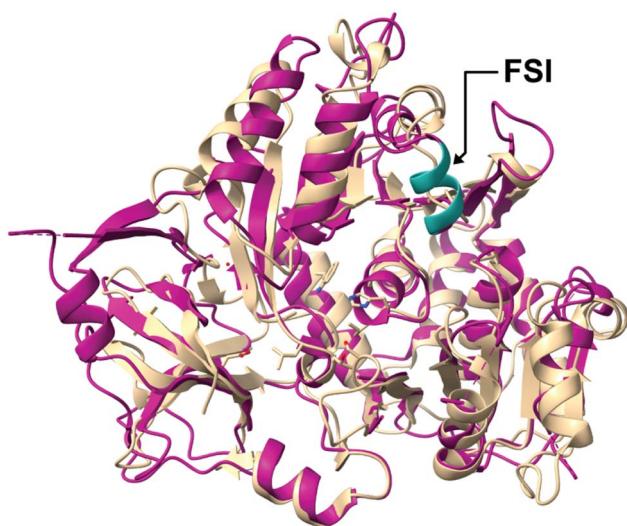


Fig. 4 Structural alignment of N-terminal domain of *M. tuberculosis* FAAL28 (magenta, PDB-ID: 3E53) with N-terminal domain of FACL13 (beige, PDB-ID: 5ZRN) highlighting the FAAL specific insertion sequence (FSI, turquoise).

the formation of acyl-CoA derivatives (Fig. 4).<sup>6</sup> Studies demonstrated that the introduction of the FSI from Mtb into a FACL results in decreased acyl-CoA production. To confirm these results and to understand if acyl-CoA formation is dependent on the FSI, Arora *et al.* and Goyal *et al.*<sup>6,33</sup> tested the deletion of the sequence in Mtb-FAAL which resulted in a weak acyl-AMP reaction with CoA, leaving the inner workings of this mechanism still not fully understood.

In another recent study it was found that FAALs use highly selective alternative binding sites that can distinguish 3',5'-bisphosphate-containing CoA from holo-ACP, thus ensuring CoA-independent activation and redirection of fatty acids towards the biosynthesis of lipidic metabolites. Still, the exact mechanism leading to the observed selectivity is not yet fully understood.<sup>25</sup> The presence of FAAL-like proteins in bacteria, archaea and eukaryote genomes led to the conclusion that formation of acyl-

ACPs by these enzymes is relevant in lipid metabolite biosynthesis across the tree of life, and was evolutionary mediated by the addition of a FAAL specific insertion sequence into FACLs.<sup>33</sup>

In both FAALs and FACLs, the larger N-terminal domain is connected to the smaller C-terminal domain through a flexible loop. This flexibility enables the C-terminal domain to adopt different orientations, facilitating substrate binding and, in FAALs, interactions with the ACP or, in FACLs, with CoA for fatty acyl group transfer. Additionally, the N-terminal domain is divided into three subdomains that form a pocket for substrate and product binding.<sup>32</sup> Structural features of ANL family members were compared by Patil *et al.*<sup>25</sup> and the study showed that FAALs interact with their ligands through (i) hydrogen bonds formed by residues of the C-terminal protein domain, (ii) water-mediated contacts formed by N-terminal helices and (iii) interactions with phosphates through positively charged residues (Arg/Lys). Structure-alignments of different FAALs, namely FadD32 from *Mycobacterium marinum*, FadD32 from *Mycobacterium smegmatis*, FadD28, FadD13 and FadD10 from Mtb, FAALs from *L. pneumophila* and *E. coli*, were generated by Guillet *et al.*<sup>36</sup> and compared with FACLs. All analyzed FAALs contained a conserved Ile-363-Val-386 SI4 segment that builds the bridge between N- and C-terminal domains.<sup>32</sup> In previous studies, it was shown that this motif is a specific trait of FAAL homologs, sufficient to prevent the formation of acyl-CoA derivatives.<sup>6</sup>

Because crystallization of EcFAAL and LpFAAL led to co-crystallization of the acyl adenylate bound to the enzyme, 3D-structures could be analyzed regarding substrate and product binding (Fig. 5A). According to structural analysis, three loops form the entrance to the active site, while residues 545–553 (annotation according to EcFAAL) from the C-terminal domain may belong to the channel reaching the active site of the enzyme. During substrate binding and product dissociation, the highly flexible hinge region enables C-terminal domain movements. In both analyzed structures, the AMP moieties seem to interact with conserved residues, while the adenyl rings are bound in a hydrophobic pocket supported by hydrogen bonds formed with main and side chain atoms (Fig. 5B).<sup>32</sup>



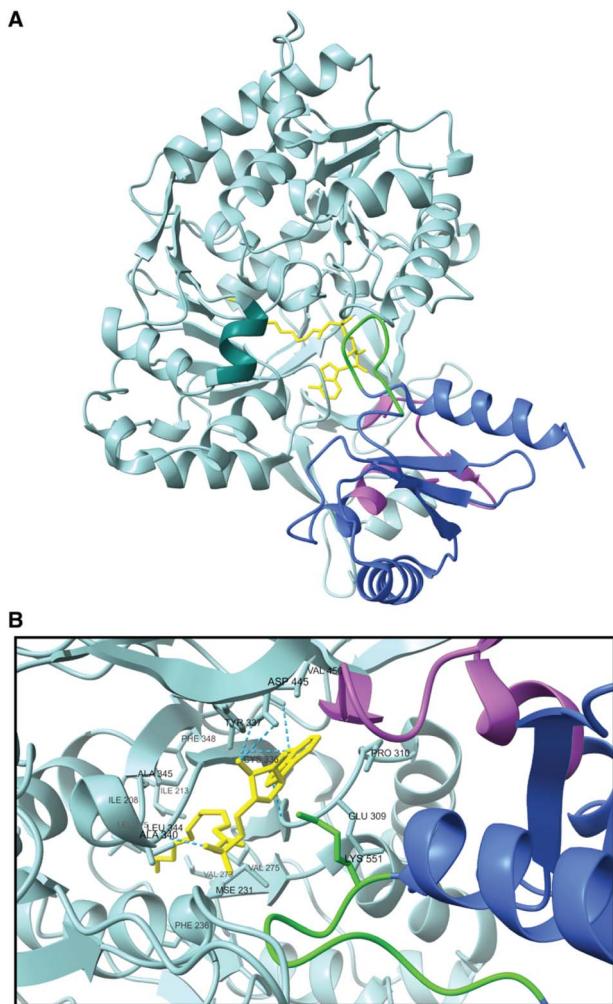


Fig. 5 (A) Structure of chain A of EcFAAL (PDB-ID: 3PBK), ligand 5'-O-[(S)-(dodecanoyloxy)(hydroxy)phosphoryl]adenosine in yellow, N-terminal domain in light blue, hinge region in pink, C-terminal domain in blue, highly conserved loop region reaching into the active site in green and FSI in turquoise. (B) Zoom in into active site of EcFAAL.

The hydrophobic pocket in EcFAAL is primarily formed by Pro310, Tyr337 and Val456. Further substrate positioning is achieved through hydrogen bonds between the adenyl ring and Cys336 in the N-terminal domain as well as hydrogen bonding of the ribose moiety to Asp445. Extensive van der Waals interactions also play a crucial role in stabilizing the substrate within the enzyme's pocket. Additionally, the phosphate oxygens are bound to Ala340 *via* hydrogen bonds, supporting the overall stability of the substrate binding. These interactions are conserved among FAAL enzymes and ensure efficient substrate binding and positioning.<sup>32</sup>

### 3 Role of FAALs in natural products biosynthesis

#### 3.1 Genomic context of FAALs and their ACPs in bacteria

In bacteria, FAAL-encoding genes are often found within BGCs associated with the multimodular PKS and NRPS machineries.

These encode for the biosynthesis of complex lipidic and peptidic natural products, respectively, and are often found in both Gram-positive and Gram-negative bacteria.<sup>12</sup> FAALs play a key role in initiating biosynthesis by activating fatty acids and loading them onto ACPs; the resulting thioester is then primed for nucleophilic attack by the activated monomer in the ensuing module. In most cases, the FAAL-encoding gene is located near genes coding for ACPs or is directly associated with an ACP-domain in a didomain protein. This ensures the efficient transfer of activated fatty acid to the rest of the biosynthetic machinery. But FAAL domains can be also part of multidomain proteins as recently described for the biosynthesis of occidiofungins, antifungal natural products of hybrid PKS III/NRPS origin in *Burkholderia* species.<sup>39</sup> In Cyanobacteria and Pseudomonadota, FAALs fused to either PKS (*e.g.* in desmamides,<sup>40</sup> micacocidin<sup>41</sup>) or NRPS (*e.g.* vioprolides<sup>42</sup>) were also described. A more detailed discussion on this topic is provided in below, together with illustrative examples.

To date, a large variety of FAAL-encoding gene clusters have been identified and this number is increasing, as advances in bioinformatics lead to better annotation of FAAL domains.<sup>43</sup> Taking into account both published fatty acyl-chain containing natural products and the MIBiG database,<sup>44</sup> most of the so far characterized fatty acyl-containing natural products associated with FAAL-containing pathways originate from the phyla Cyanobacteria, Pseudomonadota, Actinobacteria, Firmicutes, Myxococcota and Bacteroidetes (Table 2).

#### 3.2 Stand-alone FAALs

FAAL-encoding genes are often stand-alone, despite usually being co-localized with their cognate ACP. This is the case in jamaicamide biosynthesis, which we use here as an example (Fig. 6A). Jamaicamides are a group of highly functionalized lipopeptides from cyanobacteria that contain alkynyl bromide, vinyl chloride, beta-methoxy enone, and pyrrolinone ring functionalities. The fatty acid-incorporating enzyme of this BGC is a stand-alone FAAL, JamA, which was initially annotated as an acyl-ACP synthetase.<sup>47</sup> In the biosynthesis of these compounds, a 5-hexenoic acid unit is activated by JamA and loaded onto the ACP JamB as an acyl-ACP thioester. A desaturase then acts on the alkyne moiety to generate a 5-hexynoyl-ACP and this unit is then further elongated and modified through PKS and NRPS machinery as well as tailoring enzymes.<sup>47,96</sup> Jamaicamide biosynthesis illustrates the richness of enzymatic transformations associated with fatty acid moieties in secondary metabolism.<sup>97</sup> The cylindrocyclophanes are another example of a natural product in which a stand-alone FAAL initiates biosynthesis, in this case interfacing with type III PKS machinery (Fig. 6A). Here, the FAAL loads decanoic acid onto its ACP which is then elongated by type I PKS machinery followed by elongation and cyclization by the encoded type III PKS enzyme.<sup>45,98</sup>

#### 3.3 Didomain FAALs

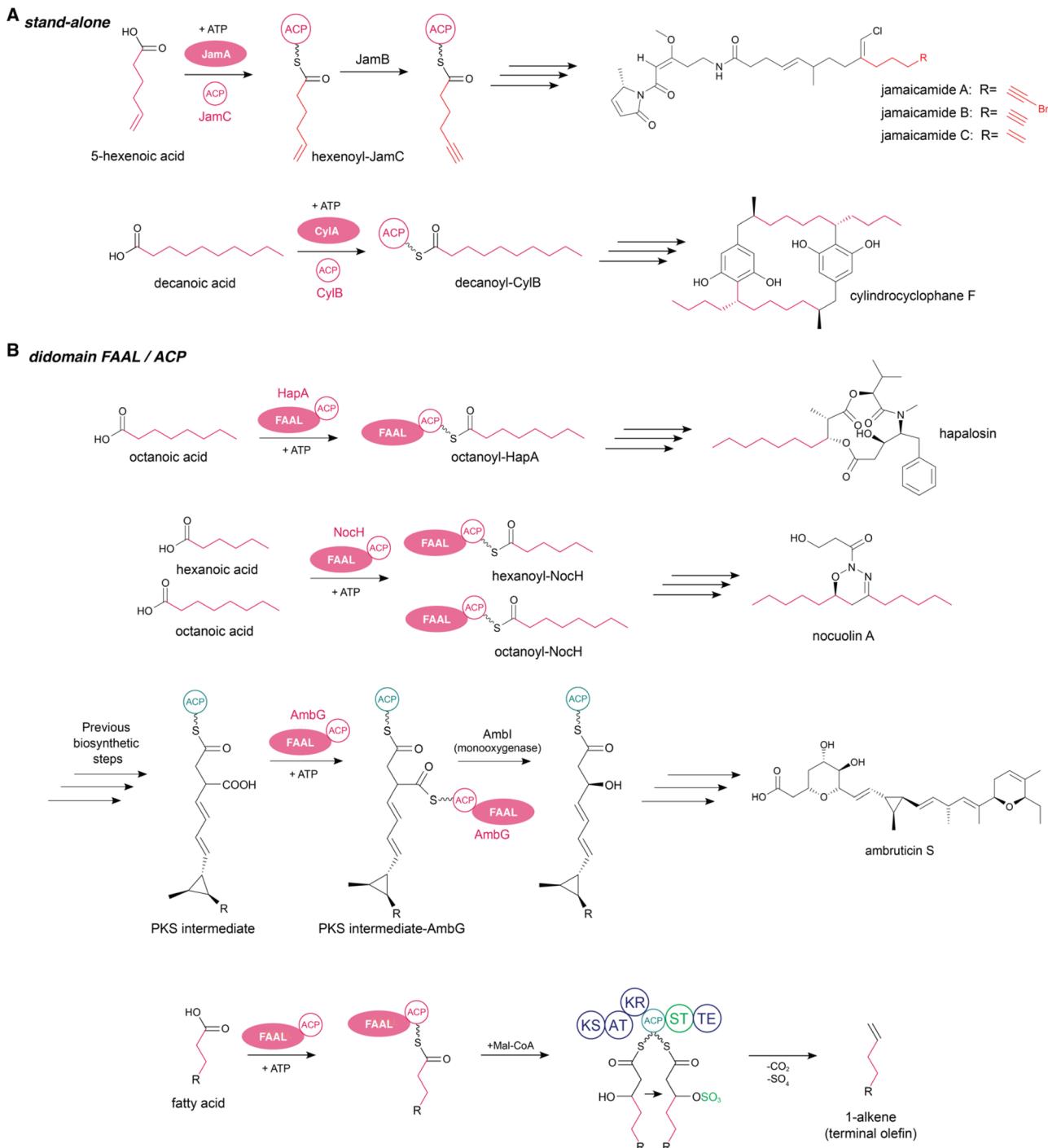
FAAL domains are also found fused with their cognate ACP in a didomain protein (Fig. 6B). A canonical example for such



**Table 2** Examples of published natural products associated with FAAL-encoding biosynthetic gene clusters. Each biosynthetic gene cluster encodes at least one FAAL enzyme/domain

Compound name	Substrate	Evidence level	Phylum	Reference
Cylindrocyclophanes	C10:0	<i>In vitro</i> data	Cyanobacteria	45
Hectochlorin	C6:0	Hypothesis	Cyanobacteria	46
Jamaicamides	C6:1, C6:2	<i>In vitro</i> data	Cyanobacteria	47
Malyngamides	C8:0	Hypothesis	Cyanobacteria	48
Nocuolin A	C6:0, C8:0	<i>In vitro</i> data	Cyanobacteria	28
Hapalosin	C8:0	Hypothesis	Cyanobacteria	49
Puwainaphycin	C12:0–C16:0	Hypothesis	Cyanobacteria	50
Microginins	C10:0	Hypothesis	Cyanobacteria	51
Columbamides	C12:0, C14:0	Hypothesis	Cyanobacteria	52
Desmamides	C6:0	Hypothesis	Cyanobacteria	40
Olefin	Variable	Hypothesis	Cyanobacteria	53
Hassalidins	C14:0, C16:0	Hypothesis	Cyanobacteria	54
Minutissamides	C12:0–C16:0	Hypothesis	Cyanobacteria	55
Laxaphycin	C6:0	Hypothesis	Cyanobacteria	56
Chlorinated lactylates	C12:0	Hypothesis	Cyanobacteria	57
Scytocyclamides	C6:0	Hypothesis	Cyanobacteria	58
Vatiamides	C6:0	Hypothesis	Cyanobacteria	59
Carbamidocyclophanes	C10:0	Hypothesis	Cyanobacteria	60
Merocyclophane C, D	C10:0	Hypothesis	Cyanobacteria	61
Chlorosphaerolactylates A–D	C12:0	Hypothesis	Cyanobacteria	57
Nocuolactylates	C12:0	<i>In vitro</i> data	Cyanobacteria	28
Fischerazoles A–C	C16:0	Hypothesis	Cyanobacteria	62
Nostovalerolactones	C8:0	Hypothesis	Cyanobacteria	63
Ralsolamycin	C16:0	Hypothesis	Pseudomonadota	64
Micacocidin	C6:0	<i>In vitro</i> data	Pseudomonadota	65
Ambruticin	C14:0, C16:0	<i>In vitro</i> data	Pseudomonadota	66
Tambjamine	C12:0	<i>In vitro</i> data	Pseudomonadota	67
Alkylresorcylic acid	C18:0	<i>In vitro</i> data	Pseudomonadota	68
Caryoynencin	C18:1	Hypothesis	Pseudomonadota	69
Vioprolide	C16:0, C17:0	<i>In vitro</i> data	Pseudomonadota	42
Gramibactin	C8:0	Hypothesis	Pseudomonadota	70
Trinickiabactin	C8:0	Hypothesis	Pseudomonadota	71
Histicorragatin	C8:0	Hypothesis	Pseudomonadota	72
Taiwachelins	C12:0	Hypothesis	Pseudomonadota	73
Fabrubactin A, B	C10:0	<i>In vitro</i> data	Pseudomonadota	74
Variochelin A, B	C12:0	Hypothesis	Pseudomonadota	75
Plantaribactin	C12:0	Hypothesis	Pseudomonadota	76
Ralsolamycin	C16:0	Hypothesis	Pseudomonadota	77
Occidiofungins	C18:0	Hypothesis	Pseudomonadota	39
Serobactin	C10:0–C14:0	Hypothesis	Pseudomonadota	78
Rakicidin D	C6:0	Hypothesis	Pseudomonadota	79
Massiliachelin	C6:0	Hypothesis	Pseudomonadota	80
Mycolic acid	C14:0	<i>In vitro</i> data	Actinobacteria	81
PDIM	C29:0, C30:0, C32:0 (branched)	Hypothesis	Actinobacteria	16
Glycopeptidolipid	C26:0–C34:0	Hypothesis	Actinobacteria	82
Daptomycin	C8:0–C14:0	<i>In vitro</i> data	Actinobacteria	83
Taromycin	C8:0	Hypothesis	Actinobacteria	84
Catenulispoloides	C5:0	Hypothesis	Actinobacteria	85
Telomycin	C8:0	Hypothesis	Actinobacteria	86
Rotihibin A	C10:0	Hypothesis	Actinobacteria	87
Sulfolipids	C24:0–C34:0	Hypothesis	Actinobacteria	6
Naphthyridinomycin	C14:0	<i>In vitro</i> data	Actinobacteria	88
Mycosubtilin	C16:0	Hypothesis	Firmicutes	89
Mycobactin	C14:0–C18:0	Hypothesis	Firmicutes	90
Bacillomycin D	C14–C17 $\beta$ -amino fatty acid	Hypothesis	Firmicutes	91
Iturin A	C14:0–C17:0	Hypothesis	Firmicutes	92
Chondrochloren A	C4:0	Hypothesis	Myxococcota	93
Ambruticins	C14:0, C16:0	<i>In vitro</i> data	Myxococcota	66
Alkylpyrone-407	C18:0	<i>In vitro</i> data	Myxococcota	94
Alkylpyrone-393				
Corramycin	C6:0–C12:0	<i>In vitro</i> data	Bacteroidetes	95





**Fig. 6** Examples of FAALs as stand-alone, didomain and part of multidomain enzymes. (A) Stand-alone FAALs JamA in jamaicamides and CylA in cylindrocyclophanes biosynthesis, (B) didomain FAALs HapA in hapalosin, NocH in nocuolin A and AmbG in ambruticin S biosynthesis, KS: ketosynthase, AT: acyltransferase, KR: ketoreductase, ST: sulfotransferase, TE: thioesterase.

a didomain arrangement can be found in the hapalosin BGC where FAAL initiates the biosynthesis through loading of an octanoic acid onto its ACP followed by transfer to a type I PKS and further elongations and modifications by several type I PKS and NRPS modules.<sup>49</sup> A non-canonical example is the nocuolin A BGC, which also encodes for a fused FAAL-ACP protein (NocH). In the biosynthesis of nocuolin A two fatty acyl-NocH

thioesters are simultaneously used by the downstream ketosynthase NocG to form a beta-keto acid featuring a 13-carbon chain. So far it is the only known biosynthesis where the same FAAL-ACP is used twice, loading two different substrates, namely hexanoic- and octanoic acid.<sup>49</sup>

Another example for the presence of a non-canonical FAAL-ACP didomain is AmbG in ambruticin biosynthesis.<sup>66</sup> AmbG



accepted a diverse range of substrates ranging from fatty acids to structurally diverse carboxylic acids including functionalized fatty acids and unsaturated and aromatic carboxylic acids. AmbG is the first example of a FAAL-ACP didomain that is centrally located in a type I PKS and able to activate polyketidic intermediates.<sup>66</sup>

Didomain FAAL-ACP enzymes in association with PKS enzymes are also involved in hydrocarbon and lipid metabolism, namely in the extensively studied terminal olefin biosynthesis. The olefin synthase (OLS) pathway is present in several different clades of cyanobacteria and usually consists of a FAAL and ACP, followed by a modular ketosynthase (KS), acyltransferase (AT), ketoreductase (KR) and ACP domains followed by a separate sulfotransferase (ST) and thioesterase (TE) module. This BGC architecture is like the FAAL-PKS architecture in *Mtb*. After loading of a fatty acid onto the ACP, the KS domain extends the fatty acyl-ACP *via* an acetate unit and reduces the beta-carbonyl to a hydroxy group, followed by decarboxylation and desulfation to create a terminal double bond resulting in the biosynthesis of odd-chain length hydrocarbons with a terminal olefin.<sup>100</sup> However, in some cyanobacterial strains such as *Cyanobacterium stanieri* PCC 7202, the FAAL involved in olefin biosynthesis is a stand-alone protein.<sup>53</sup>

### 3.4 FAALs as part of multidomain enzymes

Because FAALs eventually form acyl thioesters, these can recruit and load FA starter units to initiate biosynthesis in PKS or NRPS assembly lines, instead of the more typical initiation through acyltransferase (AT) or adenylation (A) domains, respectively.<sup>101</sup> As mentioned above, A domains (which usually activate and load amino acids) and FAALs are members of the same protein superfamily but they only share moderate sequence homology. Besides conserved motifs responsible for ATP-binding, divergence occurs in their substrate-binding regions.<sup>3</sup> Nevertheless, such embedded FAAL domains can be misannotated by bioinformatics pipelines. For example, in the biosynthesis of the desmamides (Fig. 7), a recently discovered group of lipoglycopeptides, a FAAL domain misannotated as an A domain by antiSMASH<sup>102</sup> is proposed to load a hexanoic acid started unit

onto an ACP, as part of a PKS module. This is followed by chain elongation using malonyl-CoA and, through the activity of a ketoreductase, the generation of a linear eight-carbon  $\beta$ -hydroxy acyl-ACP intermediate. Another C2 unit is condensed to obtain a 10-carbon unit, followed by a reduction of its  $\beta$ -keto group and methylation by the methyltransferase DsmH. In the biosynthetic hypothesis, DsmI and DsmJ as NRPS introduce the corresponding amino acids and the cyclized metabolite is finally released through activity of a thioesterase. Surprisingly, no ACP could be identified between the FAAL domain and the first PKS module.<sup>40</sup>

### 3.5 Biochemical characterization of FAAL domains

Several biochemical methods have been developed for the analysis of adenylylating enzyme activities that can be translated to FAAL activity, which is useful for determining substrate scope, binding and kinetics, to understand their reaction mechanism or to determine substrate preference in competition assays. Such methods include *in vitro* assays followed by chromatographic separations for product detection and liquid chromatography-mass spectrometry (LC-MS) analysis (e.g. CylA in cylindrocyclicophanes biosynthesis<sup>45</sup>), indirect measuring through release or exchange of pyrophosphates (PPi) (e.g. FbnG in fabrubactin biosynthesis<sup>74</sup>), radiolabelling followed by TLC analysis in the study of mycobacterial FAAL domains<sup>12</sup> or loading of acyl carrier proteins with the corresponding fatty acid substrate (e.g. NocH in nocuolin A biosynthesis<sup>99</sup>).

**3.5.1 Phosphate release assays.** Since FAALs and other adenylylating enzymes catalyze reactions where an AMP moiety is transferred from ATP to a substrate, PPi is released and can then be hydrolyzed to produce inorganic phosphate (Pi). Using phosphate release kits, the release of Pi correlates with the activity of the adenylylating enzyme.<sup>103,104</sup> The detection of Pi can be realized through various methods, but most release kits use a colorimetric method where Pi forms a complex with a reagent and absorbance is measured. In the malachite green assay, Pi reacts with molybdate under acidic conditions to form a phosphomolybdate complex which reacts with malachite green dye to form a green-coloured complex which increases absorption at

#### FAAL domain as part of multidomain enzyme

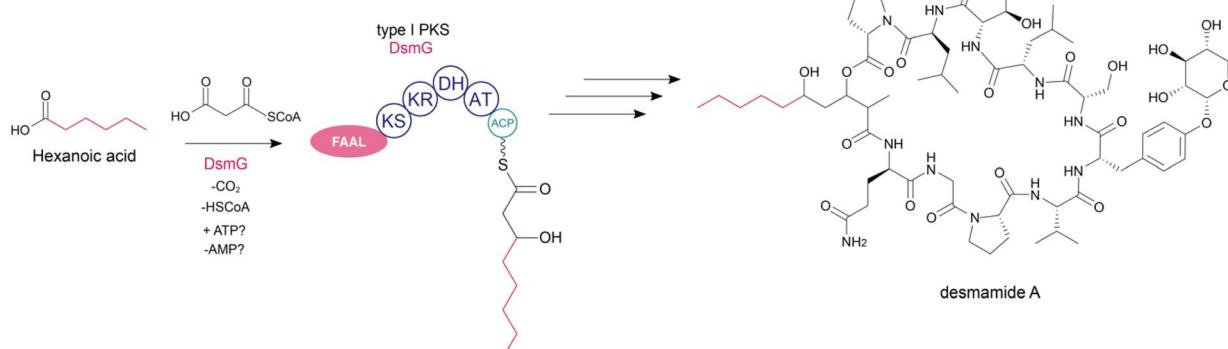


Fig. 7 FAAL as part of multidomain enzymes in desmamide A biosynthesis. KS: ketosynthase, KR: ketoreductase, DH: dehydrogenase, AT: acyltransferase.



630 nm. Pi release assays are highly sensitive and can detect low levels of Pi.<sup>105,106</sup> For example, in the study of the mycobacterial FadD32, the released PPi is hydrolyzed by a pyrophosphatase coupled reaction, followed by incubation with malachite green. The method enables a very high throughput which is why a large library of different enzyme inhibitors could be screened in the described work.<sup>107</sup>

**3.5.2 Radiolabelling of substrates.** Another method to measure FAAL activity is through the formation of radiolabelled ACP-substrate species. The assay contains an activated ACP-protein and radiolabelled substrates (<sup>3</sup>H or <sup>14</sup>C for fatty acids). Activated holo-ACP is obtained through incubation with the promiscuous Sfp (a phosphopantetheine transferase) which transfers a phosphopantetheine (Ppant) onto the conserved serine residue in the ACP. The adenylate is then covalently attached to the thiol group of the Ppant arm forming a thioester. After incorporation of the substrate, non-reacted substrate and other components can be separated from the radiolabelled acyl-ACP product that can then be detected and quantified, reflecting the activity of the FAAL enzyme.<sup>6</sup> In the analysis of mycobacterial FadD32, the FAAL involved in mycolic acids biosynthesis, radiolabelled fatty acids were incubated with the enzyme, followed by separation of reaction products by TLC and detection by autoradiography.<sup>81</sup>

**3.5.3 Mass spectrometry-based methods.** Several LC-MS based methods have been developed and optimized over the years to analyze ACPs and their loaded substrates. One method is the direct LC-MS analysis of intact ACPs which involves injection of intact ACP protein into a LC-MS system followed by analysis of the mass difference between the non-substrate-loaded apo-ACP and the loaded holo-ACP.<sup>108</sup> For example, in the characterization of FAAL enzyme involved in the biosynthesis of cylindrocyclophanes, direct analysis of the intact ACP was carried out to determine the loading capacity of the corresponding FAAL, CylA.<sup>45</sup> Standards were prepared for the apo-, holo- and decanoic acid-loaded ACP and compared to the assay sample *via* HPLC separation. An alternative is bottom-up proteomics *via* Ppant-ejection.<sup>109</sup> This method includes fragmentation of the ACP within the mass spectrometer and analysis of the substrate-bound ACP-fragment. In bottom-up LC-MS techniques, ACPs are digested either enzymatically with trypsin or chymotrypsin or chemically with for example cyanogen bromide. Digestion is followed by analysis in the LC-MS and fragmentation *via* MS/MS of the loaded ACP-fragment and search for the phosphopantetheine-containing fragment.

## 4 Deorphanization of FAAL-containing BGCs and future applications

### 4.1 Diversity and discovery of FAAL-containing orphan BGCs

Bacterial lipopeptides are a group of secondary metabolites with special interest due to their antibacterial, antifungal, anticancer and antiviral bioactivities.<sup>110</sup> The diversity of biological activities can be explained by their amphipathic molecular structure allowing micellar interaction with cell membranes of

microorganisms.<sup>111</sup> The first lipopeptide discovered and confirmed to have a FAAL involved in its biosynthesis was mycosubtilin produced by *Bacillus subtilis*.<sup>89</sup> Other well-studied examples of cyclic lipopeptides with FAAL-mediated lipoinitiation of their biosynthesis are iturin, bacillomycin and puwainaphycin (Table 2). Due to their prevalence in cyanobacteria, FAAL function and FAAL-containing BGCs in these organisms have received some attention. Galica *et al.*<sup>51</sup> searched publicly available genomes for gene clusters that encode FAALs and investigated the distribution of those gene clusters among cyanobacteria, considering phylogeny, ecology, and habitat. In their study, they found more than half (56%) of the lipopeptide BGCs analyzed to be present in heterocyst-forming cyanobacteria (Nostocales). In contrast, only 10% of Synechococcales strains showed positive for these BGCs. Galica *et al.*<sup>51</sup> also observed an occurrence of lipopeptide BGCs in around 16% of publicly available genomes from cyanobacteria while it was reported that NRPS and NRPS/PKS gene clusters would have an abundance of around 51–56% in their genomes.<sup>112,113</sup> Genome size correlated with the number of NRPS-containing gene clusters underlining the interest into cyanobacterial genomes of larger size. The same study showed the abundance of lipopeptide and herewith FAAL-containing BGCs especially in substrate-associated cyanobacteria with higher diversity in their domain organization than those coming from planktonic cyanobacteria.<sup>51</sup>

Several software programs for annotation of BGCs in genome data are currently available, for example antiSMASH,<sup>102</sup> PRISM<sup>114</sup> and ClusterFinder<sup>4</sup> to name a few. Among these, antiSMASH is the most widely used bioinformatics tool designed to identify and analyze BGCs in bacterial and fungal genomes.<sup>102</sup> It annotates biosynthetic signature amino acid sequences such as the “CAL” domain which in most cases corresponds to a FAAL-domain. While genome mining approaches to discover FAAL-containing BGCs are currently available, uncovering the associated natural product is seldom straightforward.<sup>51</sup> In many cases, the BGCs are not expressed under typical laboratory conditions, but also accurate structural predictions from the BGC composition and architecture are necessary to match it against an orphan natural product or even to discover it in, for example, metabolomics data.<sup>115</sup>

In summary, FAAL-containing orphan BGCs reflect an untapped reservoir of natural products diversity with significant potential for the discovery of novel bioactive compounds, such as lipopeptides, which may be accelerated by improving structural predictions and through a better knowledge of BGC regulation in the producing organisms.

### 4.2 Substrate specificity and prediction

A better understanding of FAAL specificities would improve predicting the final natural product structures and therefore their discovery. This would also be of special interest since fatty acid chain length influences the properties such as the cytotoxicity of natural products. For example, Saurav *et al.*<sup>116</sup> analyzed the effect of the fatty acid chain length on the cytotoxicity of puwainaphycins. It was observed that the lipophilicity



of the fatty acid residue is essential for the compound's cytotoxicity and thus for the interaction with the plasma membrane.<sup>117</sup>

Initial attempts to determine substrate specificities of the acyl-adenylate superfamily were based on active site residues profiles.<sup>118</sup> Using a computational approach, a total of 15 amino acid positions were defined as specificity-determining residues for each enzyme class. The sequential order of these 15 residues was defined as the active site profile in the training set and by using sequence alignments and specific scoring matrices, substrate preference was assigned based on the highest scoring. The authors were able to predict substrate preference for different subfamilies of proteins. For FAALs, these predictions included preference for medium ( $n = 2-10$ ) and long chain ( $n = 12-16$ ) fatty acid incorporation.<sup>118</sup>

Machine learning tools have been used to predict the function and substrate specificity of ANL enzymes more accurately.<sup>2</sup> These tools include the prediction of substrates for NRPS adenylation (A) domains which has unlocked the possibility of accessing the biosynthetic potential of uncultured microorganisms by connecting a predicted natural product with its BGC.<sup>119-121</sup> For example, bioinformatic predictions have led to the discovery of the BGC for leinamycin family compounds<sup>122</sup> and helped to predict the structure of several lipopeptides based on the A-domain sequence.<sup>123</sup> In another study, researchers developed an ensemble of substrate-specific Hidden Markov Models (HMMs) to classify and predict specificity of A-domains in NRPSs and acyltransferase domains in PKSs. Models were trained based on known domain sequences with annotated substrate specificities. The HMM approach showed superior performance compared to traditional methods. Predictions were evaluated against experimentally characterized datasets and demonstrated strong reliability.<sup>124</sup> Novel attempts of bioinformatics guided substrate prediction were based on latent semantic indexing (LSI) techniques that are used in text mining. These analyzed and predicted protein properties based on sequence data only by identifying patterns and correlations that could successfully predict the substrate specificity of A-domains with high accuracy.<sup>125</sup> Another bioinformatics approach for substrate prediction of NRPS domains was based on virtual screening of A-domains. Authors employed evaluation and ranking of potential substrates based on their binding affinity to the A-domains, created a library of potential substrates and used docking simulations to analyze for the fit of each substrate into the binding sites of different A-domains. This virtual screening approach predicted substrate preference with high accuracy and was matching experimental data.<sup>126</sup>

Existing tools still have limitations, such as the inability of predicting with confidence the chain length of lipid tails incorporated into lipopeptides by specifically FAAL enzymes. To some extent, this might be explained by the fact that FAALs have shown to have broad substrate acceptance. For example, the FAAL encoded on the corramycin BGC was predicted to load fatty acids ranging from pentanoic acid to dodecanoic acid. *In vitro* results confirmed the activation of numerous linear fatty acids with various lengths. Still, highest activity was observed for decanoic acid and decreased activity for hexanoic acid

underlining that a substrate preference of FAALs does exist.<sup>95</sup> The finding that lipopeptides often occur with different fatty acid chain length substitutions can be explained by the broad substrate tolerance of FAALs in general. Compared to FAALs, FACLs also exhibit substrate acceptance for different fatty acyl chain lengths, but they tend to have a more restricted substrate range due to their role in providing specific fatty acyl-CoA intermediates to primary metabolism.<sup>118</sup> A more recent prediction algorithm for adenylating enzymes, AdenylPred, was developed by Robinson *et al.*<sup>2</sup> This computational prediction algorithm was designed to predict the substrate specificity and function of class I adenylate-forming enzymes by using machine learning techniques (random forest approach). AdenylPred aids to advance the discovery of natural products by analyzing adenylation domains amino acid sequences and predicting the enzymes' role. Within-class accuracy was highest for FAALs and NRPS adenylation domains due to the larger amount of experimental data for those protein families. Still, the algorithm could only predict fatty acid substrate ranges rather than a precise fatty acid chain length.<sup>2</sup>

Since no  $\beta$ -oxidation pathway has been yet identified in cyanobacteria, a method for the discovery of fatty acid harbouring natural products that allows for direct labelling of the cyanobacterial metabolome followed by dereplication has recently been developed. This method is based on the incorporation of stable isotope-labelled fatty acids into the cyanobacterial lipidome and thus enabling the deorphanization of yet unknown BGCs. Using this novel strategy, analogues of haptolactin and the new family of compounds, nocuolactylates, were uncovered.<sup>28</sup>

The biotechnological potential of FAAL enzymes was recently demonstrated in the context of daptomycin biosynthesis.<sup>127</sup> Researchers swapped the original FAAL enzyme with non-native homologs, specifically CylA and HmqF, which improved fatty acyl specificity and resulted in higher daptomycin purity (70% and 90% compared to 40%), despite the FAALs originating from phylogenetically distant bacteria. Additionally, the authors fused FAALs to ACPs, leading to remarkable increases in daptomycin titers. This highlights the potential of FAAL enzymes to enhance production titers and to aid in the design of novel natural products.

Our current understanding of FAAL substrate specificity is still rudimentary. For example, the development of prediction algorithms based on machine learning could be enhanced by expanding the dataset of characterized FAALs. Additionally, the generation of comprehensive substrate libraries and conducting mutagenesis experiments would provide more precise specificity profiles.

Synthetic biology offers opportunities to design customized FAAL variants through directed evolution and modular assembly of biosynthetic pathways. Integrating FAAL into non-natural NRPS or PKS systems could enable the synthesis of novel and more bioactive products. However, challenges such as accurately mimicking native interactions between FAALs and their downstream ACPs as well as the limited understanding of post-translational modifications remain.



## 5 Conclusions

First identified in *Mycobacterium tuberculosis*, the discovery of FAALs marked a significant advancement in understanding lipid metabolism in bacteria. Over time, FAALs have been found across all three domains of life – bacteria, archaea, and eukaryotes – highlighting their essential role in lipid metabolism. A breakthrough in differentiating FAALs from the structurally similar FACLs was the identification of the FAAL specific insertion (FSI), a unique amino acid sequence found exclusively in FAALs.

The resolution of several FAAL crystal structures has further illuminated their reaction mechanisms and substrate-binding properties. FAALs are now known to participate in the biosynthesis of a variety of natural products, particularly within Cyanobacteria, Actinobacteria, Pseudomonadota, Firmicutes, and Myxococcota. As genome mining continues, it is expected that more FAAL-associated natural products will be discovered.

FAALs demonstrate functional versatility by existing either as stand-alone, didomain proteins or as components of multienzyme complexes. The availability of diverse methods to characterize FAAL proteins has accelerated the discovery and deorphanization of related BGCs. Additionally, the ability to predict FAAL substrates enhances the connection between BGCs and their corresponding natural products, further driving research into natural products.

In summary, FAALs, once overlooked, have now emerged as widely distributed enzymes that play a critical role in the biosynthesis of various natural products, particularly polyketides and lipopeptides. Their unique function in channelling fatty acids into secondary metabolism makes them a promising target for metabolic engineering and enzyme modification to create novel fatty-acylated natural products in bacteria.

## 6 Data availability

The structural alignments were performed using UCSF ChimeraX (version 1.3). All structural data used in this study was retrieved from the Protein Data Bank (PDB). No new experimental data was generated for this review.

## 7 Author contributions

A. L. and P. N. L. conceptualized and wrote the article.

## 8 Conflicts of interest

There are no conflicts to declare.

## 9 Acknowledgements

This work was developed under the European Research Council through a Starting Grant (759840), from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 952374 and Fundação para a Ciência e a Tecnologia (FCT) through the strategic funding grants UIDB/04423/

2020 and UIDP/04423/2020. Anne Liong is funded by the PhD scholarship 2020.08183.BD.

## 10 References

- 1 P. Berg, *J. Am. Chem. Soc.*, 1955, **77**, 3163–3164.
- 2 S. L. Robinson, B. R. Terlouw, M. D. Smith, S. J. Pidot, T. P. Stinear, M. H. Medema and L. P. Wackett, *J. Biol. Chem.*, 2020, **295**, 14826–14839.
- 3 S. Schmelz and J. H. Naismith, *Curr. Opin. Struct. Biol.*, 2009, **19**, 666–671.
- 4 P. Cimermancic, M. H. Medema, J. Claesen, K. Kurita, L. C. W. Brown, K. Mavrommatis, A. Pati, P. A. Godfrey, M. Koehrsen and J. Clardy, *Cell*, 2014, **158**, 412–421.
- 5 J. K. Christenson, J. E. Richman, M. R. Jensen, J. Y. Neufeld, C. M. Wilmot and L. P. Wackett, *Biochemistry*, 2017, **56**, 348–351.
- 6 P. Arora, A. Goyal, V. T. Natarajan, E. Rajakumara, P. Verma, R. Gupta, M. Yousuf, O. A. Trivedi, D. Mohanty and A. Tyagi, *Nat. Chem. Biol.*, 2009, **5**, 166–173.
- 7 M. Iffa and D. Söll, *Annu. Rev. Biochem.*, 2000, **69**, 617–650.
- 8 R. K. Airas, *Biophys. Chem.*, 2007, **131**, 29–35.
- 9 H. Yonus, P. Neumann, S. Zimmermann, J. r. J. May, M. A. Marahiel and M. T. Stubbs, *J. Biol. Chem.*, 2008, **283**, 32484–32491.
- 10 L. Clark, D. Leatherby, E. Krilich, A. J. Ropolewski and J. Perozich, *PLoS One*, 2018, **13**, e0203218.
- 11 C. Ghazaei, *J. Res. Med. Sci.*, 2018, **23**, 63.
- 12 O. A. Trivedi, P. Arora, V. Sridharan, R. Tickoo, D. Mohanty and R. S. Gokhale, *Nature*, 2004, **428**, 441–445.
- 13 R. S. Gokhale, P. Saxena, T. Chopra and D. Mohanty, *Nat. Prod. Rep.*, 2007, **24**, 267–277.
- 14 N.-H. Le, P. Constant, S. Tranier, V. Nahoum, V. Guillet, L. Maveyraud, M. Daffé, L. Mourey, P. Verhaeghe and H. Marrakchi, *Bioorg. Med. Chem.*, 2022, **71**, 116938.
- 15 C. Díaz, J. Pérez del Palacio, P. L. Valero-Guillén, P. Mena García, I. Pérez, F. Vicente, C. Martín, O. Genilloud, A. Sanchez Pozo and J. s. Gonzalo-Asensio, *ACS Infect. Dis.*, 2019, **5**, 1317–1326.
- 16 O. A. Trivedi, P. Arora, A. Vats, M. Z. Ansari, R. Tickoo, V. Sridharan, D. Mohanty and R. S. Gokhale, *Mol. Cell*, 2005, **17**, 631–643.
- 17 D. Mohanty, R. Sankaranarayanan and R. S. Gokhale, *Tuberculosis*, 2011, **91**, 448–455.
- 18 J. A. Ferreras, K. L. Stirrett, X. Lu, J.-S. Ryu, C. E. Soll, D. S. Tan and L. E. Quadri, *Chem. Biol.*, 2008, **15**, 51–61.
- 19 R. Siméone, M. Léger, P. Constant, W. Malaga, H. Marrakchi, M. Daffé, C. Guilhot and C. Chalut, *FEBS J.*, 2010, **277**, 2715–2725.
- 20 M. Baran, K. D. Grimes, P. A. Sibbald, P. Fu, H. I. Boshoff, D. J. Wilson and C. C. Aldrich, *Eur. J. Med. Chem.*, 2020, **201**, 112408.
- 21 Y.-J. Chen, P. M. Leung, P. L. Cook, W. W. Wong, T. Hutchinson, V. Eate, A. J. Kessler and C. Greening, *ISME J.*, 2022, **16**, 750–763.
- 22 B. R. Miller and A. M. Gulick, *Nonribosomal Peptide and Polyketide Biosynthesis: Methods and Protocols*, 2016, pp. 3–29.



23 M. Strioker, A. Tanović and M. A. Marahiel, *Curr. Opin. Struct. Biol.*, 2010, **20**, 234–240.

24 S. Cole, R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S. Gordon, K. Eglmeier, S. Gas and C. Barry Iii, *Nature*, 1998, **396**, 190.

25 G. S. Patil, P. Kinatukara, S. Mondal, S. Shambhavi, K. D. Patel, S. Pramanik, N. Dubey, S. Narasimhan, M. K. Madduri and B. Pal, *Elife*, 2021, **10**, e70067.

26 D. Portevin, C. de Sousa-D'Auria, H. Montrozier, C. Houssin, A. Stella, M.-A. Lanéelle, F. Bardou, C. Guilhot and M. Daffé, *J. Biol. Chem.*, 2005, **280**, 8862–8874.

27 D. B. Hansen, S. B. Bumpus, Z. D. Aron, N. L. Kelleher and C. T. Walsh, *J. Am. Chem. Soc.*, 2007, **129**, 6366–6367.

28 S. A. Figueiredo, M. Preto, G. Moreira, T. P. Martins, K. Abt, A. Melo, V. M. Vasconcelos and P. N. Leão, *Angew. Chem., Int. Ed.*, 2021, **60**, 10064–10072.

29 S. Mondal, B. Pal and R. Sankaranarayanan, *Proteins: Struct., Funct., Bioinf.*, 2025, **93**, 26–37.

30 S. Jain, A. Caforio and A. J. Driessen, *Front. Microbiol.*, 2014, **5**, 641.

31 Z. Liu, T. R. Ioerger, F. Wang and J. C. Sacchettini, *J. Biol. Chem.*, 2013, **288**, 18473–18483.

32 Z. Zhang, R. Zhou, J. M. Sauder, P. J. Tonge, S. K. Burley and S. Swaminathan, *J. Mol. Biol.*, 2011, **406**, 313–324.

33 A. Goyal, P. Verma, M. Anandhakrishnan, R. S. Gokhale and R. Sankaranarayanan, *J. Mol. Biol.*, 2012, **416**, 221–238.

34 M. Yan, L. Cao, L. Zhao, W. Zhou, X. Liu, W. Zhang and Z. Rao, *Front. Microbiol.*, 2023, **14**, 1090534.

35 M. L. Kuhn, E. Alexander, G. Minasov, H. J. Page, Z. Warwrzak, L. Shuvalova, K. J. Flores, D. J. Wilson, C. Shi and C. C. Aldrich, *ACS Infect. Dis.*, 2016, **2**, 579–591.

36 V. Guillet, S. Galandrin, L. Maveyraud, S. Ladevèze, V. Mariaule, C. Bon, N. Eynard, M. Daffé, H. Marrakchi and L. Mourey, *J. Biol. Chem.*, 2016, **291**, 7973–7989.

37 W. Li, S. Gu, J. Fleming and L. Bi, *Sci. Rep.*, 2015, **5**, 15493.

38 A. Goyal, M. Yousuf, E. Rajakumara, P. Arora, R. S. Gokhale and R. Sankaranarayanan, *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.*, 2006, **62**, 350–352.

39 E. Simonetti, F. Alvarez, N. Feldman, M. Vinacour, I. N. Roberts and J. A. Ruiz, *Biol. Control*, 2021, **155**, 104530.

40 S. Freitas, R. Castelo-Branco, A. Wenzel-Storjohann, V. M. Vasconcelos, D. Tasdemir and P. N. Leão, *J. Nat. Prod.*, 2022, **85**, 1704–1714.

41 H. Kage, M. F. Kreutzer, B. Wackler, D. Hoffmeister and M. Nett, *Chem. Biol.*, 2013, **20**, 764–771.

42 D. Auerbach, F. Yan, Y. Zhang and R. Müller, *ACS Chem. Biol.*, 2018, **13**, 3123–3130.

43 K. Blin, M. H. Medema, R. Kottmann, S. Y. Lee and T. Weber, *Nucleic Acids Res.*, 2016, gkw960.

44 B. R. Terlouw, K. Blin, J. C. Navarro-Munoz, N. E. Avalon, M. G. Chevrette, S. Egbert, S. Lee, D. Meijer, M. J. Recchia and Z. L. Reitz, *Nucleic Acids Res.*, 2023, **51**, D603–D610.

45 H. Nakamura, H. A. Hamer, G. Sirasani and E. P. Balskus, *J. Am. Chem. Soc.*, 2012, **134**, 18518–18521.

46 A. V. Ramaswamy, C. M. Sorrels and W. H. Gerwick, *J. Nat. Prod.*, 2007, **70**, 1977–1986.

47 D. J. Edwards, B. L. Marquez, L. M. Nogle, K. McPhail, D. E. Goeger, M. A. Roberts and W. H. Gerwick, *Chem. Biol.*, 2004, **11**, 817–833.

48 W. H. Gekwick, S. Reyes and B. Alvarado, *Phytochemistry*, 1987, **26**, 1701–1704.

49 K. Stratmann, D. L. Burgoyne, R. E. Moore, G. M. Patterson and C. D. Smith, *J. Org. Chem.*, 1994, **59**, 7219–7226.

50 J. Mareš, J. Hájek, P. Urajová, J. Kopecký and P. Hrouzek, *PLoS One*, 2014, **9**, e111904.

51 T. Galica, P. Hrouzek and J. Mareš, *J. Phycol.*, 2017, **53**, 985–998.

52 J. A. V. Lopez, J. G. Petitbois, C. S. Vairappan, T. Umezawa, F. Matsuda and T. Okino, *Org. Lett.*, 2017, **19**, 4231–4234.

53 T. Zhu, T. Scalvenzi, N. Sasso, X. Lu and M. Gugger, *Appl. Environ. Microbiol.*, 2018, **84**, e00425.

54 J. Vestola, T. K. Shishido, J. Jokela, D. P. Fewer, O. Aitio, P. Permi, M. Wahlsten, H. Wang, L. Rouhiainen and K. Sivonen, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E1909–E1917.

55 H.-S. Kang, M. Sturdy, A. Krunic, H. Kim, Q. Shen, S. M. Swanson and J. Orjala, *Bioorg. Med. Chem.*, 2012, **20**, 6134–6143.

56 L. Darcel, S. Das, I. Bonnard, B. Banaigs and N. Inguimbert, *Mar. Drugs*, 2021, **19**, 473.

57 K. Abt, R. Castelo-Branco and P. N. Leão, *J. Nat. Prod.*, 2021, **84**, 278–286.

58 L. M. P. Heinilä, D. P. Fewer, J. K. Jokela, M. Wahlsten, A. Jortikka and K. Sivonen, *Front. Microbiol.*, 2020, **11**, 578878.

59 N. A. Moss, G. Seiler, T. F. Leão, G. Castro-Falcón, L. Gerwick, C. C. Hughes and W. H. Gerwick, *Angew. Chem., Int. Ed.*, 2019, **58**, 9027–9031.

60 M. Preitsch, S. E. Heiden, M. Beerbaum, T. H. Niedermeyer, M. Schneefeld, J. Herrmann, J. Kumpfmüller, A. Thürmer, I. Neidhardt and C. Wiesner, *Mar. Drugs*, 2016, **14**, 21.

61 D. S. May, W.-L. Chen, D. D. Lantvit, X. Zhang, A. Krunic, J. E. Burdette, A. Eustaquio and J. Orjala, *J. Nat. Prod.*, 2017, **80**, 1073–1080.

62 S. Figueiredo, K. Abt, T. Martins, I. Lacomba, A. Forero, C. Jiménez, J. Rodríguez and P. Leão, *ChemRxiv*, 2022, DOI: [10.26434/chemrxiv-2022-v8zlp](https://doi.org/10.26434/chemrxiv-2022-v8zlp).

63 J. Krumbholz, K. Ishida, M. Baunach, J. E. Teikari, M. M. Rose, S. Sasso, C. Hertweck and E. Dittmann, *Angew. Chem., Int. Ed.*, 2022, **61**, e202204545.

64 J. E. Spraker, L. M. Sanchez, T. M. Lowe, P. C. Dorrestein and N. P. Keller, *ISME J.*, 2016, **10**, 2317–2330.

65 M. F. Kreutzer, H. Kage, P. Gebhardt, B. Wackler, H. P. Saluz, D. Hoffmeister and M. Nett, *Appl. Environ. Microbiol.*, 2011, **77**, 6117–6124.

66 F. Hemmerling, K. E. Lebe, J. Wunderlich and F. Hahn, *ChemBioChem*, 2018, **19**, 1006–1011.

67 P. M. Marchetti, V. Kelly, J. P. Simpson, M. Ward and D. J. Campopiano, *Org. Biomol. Chem.*, 2018, **16**, 2735–2740.

68 T. Hayashi, Y. Kitamura, N. Funa, Y. Ohnishi and S. Horinouchi, *ChemBioChem*, 2011, **12**, 2166–2176.



69 C. Ross, K. Scherlach, F. Kloss and C. Hertweck, *Angew. Chem., Int. Ed.*, 2014, **53**, 7794–7798.

70 R. Hermenau, K. Ishida, S. Gama, B. Hoffmann, M. Pfeifer-Leeg, W. Plass, J. F. Mohr, T. Wichard, H.-P. Saluz and C. Hertweck, *Nat. Chem. Biol.*, 2018, **14**, 841–843.

71 J. Jiao, J. Du, A. Frediansyah, G. Jahanshah and H. Gross, *J. Antibiot.*, 2020, **73**, 28–34.

72 S. Matthijs, N. Brandt, M. Ongena, W. Achouak, J.-M. Meyer and H. Budzikiewicz, *BioMetals*, 2016, **29**, 467–485.

73 M. F. Kreutzer and M. Nett, *Org. Biomol. Chem.*, 2012, **10**, 9338–9343.

74 V. Vinnik, F. Zhang, H. Park, T. B. Cook, K. Throckmorton, B. F. Pfleger, T. S. Bugni and M. G. Thomas, *ACS Chem. Biol.*, 2020, **16**, 125–135.

75 C. Kurth, S. Schieferdecker, K. Athanasopoulou, I. Seccareccia and M. Nett, *J. Nat. Prod.*, 2016, **79**, 865–872.

76 R. Hermenau, J. L. Mehl, K. Ishida, B. Dose, S. J. Pidot, T. P. Stinear and C. Hertweck, *Angew. Chem.*, 2019, **131**, 13158–13163.

77 N. Matsukawa, C. Tsumori, K. Ohnishi and K. Kai, *ACS Chem. Biol.*, 2023, **18**, 572–582.

78 F. Roseconi, D. Davyt, V. Martínez, M. Martínez, J. A. Abin-Carriquiry, H. Zane, A. Butler, E. M. de Souza and E. Fabiano, *Environ. Microbiol.*, 2013, **15**, 916–927.

79 H. Komaki, A. Ishikawa, N. Ichikawa, A. Hosoyama, M. Hamada, E. Harunari, T. Nihira, W. Panbangred and Y. Igarashi, *Stand. Genomic Sci.*, 2016, **11**, 1–9.

80 T. Steinmetz, B. K. Lombe and M. Nett, *Beilstein J. Org. Chem.*, 2023, **19**, 909–917.

81 M. Léger, S. Gavalda, V. Guillet, B. van Der Rest, N. Slama, H. Montrozier, L. Mourey, A. Quémard, M. Daffé and H. Marrakchi, *Chem. Biol.*, 2009, **16**, 510–519.

82 A. Vats, A. K. Singh, R. Mukherjee, T. Chopra, M. S. Ravindran, D. Mohanty, D. Chatterji, J.-M. Reyrat and R. S. Gokhale, *J. Biol. Chem.*, 2012, **287**, 30677–30687.

83 M. Wittmann, U. Linne, V. Pohlmann and M. A. Marahiel, *FEBS J.*, 2008, **275**, 5343–5354.

84 K. Yamanaka, K. A. Reynolds, R. D. Kersten, K. S. Ryan, D. J. Gonzalez, V. Nizet, P. C. Dorrestein and B. S. Moore, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 1957–1962.

85 S. Son, Y.-S. Hong, Y. Futamura, M. Jang, J. K. Lee, K. T. Heo, S.-K. Ko, J. S. Lee, S. Takahashi and H. Osada, *Org. Lett.*, 2018, **20**, 7234–7238.

86 C. W. Johnston, M. A. Skinnider, C. A. Dejong, P. N. Rees, G. M. Chen, C. G. Walker, S. French, E. D. Brown, J. Bérdy and D. Y. Liu, *Nat. Chem. Biol.*, 2016, **12**, 233–239.

87 S. Planckaert, B. Deflandre, A.-M. De Vries, M. Ameye, J. C. Martins, K. Audenaert, S. Rigali and B. Devreese, *Microbiol. Spectrum*, 2021, **9**, 00571, DOI: [10.1128/spectrum](https://doi.org/10.1128/spectrum).

88 J.-Y. Pu, C. Peng, M.-C. Tang, Y. Zhang, J.-P. Guo, L.-Q. Song, Q. Hua and G.-L. Tang, *Org. Lett.*, 2013, **15**, 3674–3677.

89 E. H. Duitman, L. W. Hamoen, M. Rembold, G. Venema, H. Seitz, W. Saenger, F. Bernhard, R. Reinhardt, M. Schmidt and C. Ullrich, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 13294–13299.

90 G. Snow, *Biochem. J.*, 1965, **97**, 166–175.

91 J. Y. Lu, K. Zhou, W. T. Huang, P. Zhou, S. Yang, X. Zhao, J. Xie, L. Xia and X. Ding, *Appl. Microbiol. Biotechnol.*, 2019, **103**, 7647–7662.

92 K. Tsuge, T. Akiyama and M. Shoda, *J. Bacteriol.*, 2001, **183**, 6265–6273.

93 S. Rachid, M. Scharfe, H. Blöcker, K. J. Weissman and R. Müller, *Chem. Biol.*, 2009, **16**, 70–81.

94 J. J. Hug, F. Panter, D. Krug and R. Müller, *J. Ind. Microbiol. Biotechnol.*, 2019, **46**, 319–334.

95 C. Couturier, S. Groß, A. von Tesmar, J. Hoffmann, S. Deckarm, A. Fievet, N. Dubarry, T. Taillier, C. Pöverlein, H. Stump, M. Kurz, L. Toti, S. Haag Richter, D. Schummer, P. Sizun, M. Hoffmann, R. Prasad Awal, N. Zaburannyi, K. Harmrolfs, J. Wink, E. Lessoud, T. Vermat, V. Cazals, S. Silve, A. Bauer, M. Mourez, L. Fraisse, C. Leroi-Geissler, A. Rey, S. Versluys, E. Bacqué, R. Müller and S. Renard, *Angew. Chem., Int. Ed. Engl.*, 2022, **61**(51), e202210747.

96 X. Zhu, J. Liu and W. Zhang, *Nat. Chem. Biol.*, 2015, **11**, 115–120.

97 P. N. Leão, T. P. Martins, K. Abt, J. P. Reis, S. Figueiredo, R. Castelo-Branco and S. Freitas, *Chem. Commun.*, 2023, **59**, 4436–4446.

98 H. Nakamura, E. E. Schultz and E. P. Balskus, *Nat. Chem. Biol.*, 2017, **13**, 916–921.

99 T. P. Martins, N. R. Glasser, D. J. Kountz, P. Oliveira, E. P. Balskus and P. N. Leão, *ACS Chem. Biol.*, 2022, **17**, 2528–2537.

100 N. A. Moss, M. J. Bertin, K. Kleigrewe, T. F. Leão, L. Gerwick and W. H. Gerwick, *J. Ind. Microbiol. Biotechnol.*, 2016, **43**, 313–324.

101 M. A. Fischbach and C. T. Walsh, *Chem. Rev.*, 2006, **106**, 3468–3496.

102 K. Blin, H. U. Kim, M. H. Medema and T. Weber, *Briefings Bioinf.*, 2019, **20**, 1103–1113.

103 S. D. Katewa and S. S. Katyare, *Anal. Biochem.*, 2003, **323**, 180–187.

104 M. R. Webb, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 4884–4887.

105 J. Feng, Y. Chen, J. Pu, X. Yang, C. Zhang, S. Zhu, Y. Zhao, Y. Yuan, H. Yuan and F. Liao, *Anal. Biochem.*, 2011, **409**, 144–149.

106 S. G. Carter and D. W. Karl, *J. Biochem. Biophys. Methods*, 1982, **7**, 7–13.

107 S. Galandrin, V. Guillet, R. S. Rane, M. Léger, N. Radha, N. Eynard, K. Das, T. S. Balganesh, L. Mourey and M. Daffé, *J. Biomol. Screening*, 2013, **18**, 576–587.

108 L. M. Jenkins, J.-W. Nam, B. S. Evans and D. K. Allen, in *Plant Lipids: Methods and Protocols*, Springer, 2021, pp. 219–247.

109 P. C. Dorrestein, S. B. Bumpus, C. T. Calderone, S. Garneau-Tsodikova, Z. D. Aron, P. D. Straight, R. Kolter, C. T. Walsh and N. L. Kelleher, *Biochemistry*, 2006, **45**, 12756–12766.

110 C. O. Adetunji, O. A. Anani, O. T. Olaniyan, A. Inobeme, J. K. Oloke, W. D. Palnam and S. Ali, in *Green Sustainable Process for Chemical and Environmental Engineering and Science*, Elsevier, 2022, pp. 189–204.



111 M. Inès and G. Dhouha, *Peptides*, 2015, **71**, 100–112.

112 P. M. Shih, D. Wu, A. Latifi, S. D. Axen, D. P. Fewer, E. Talla, A. Calteau, F. Cai, N. Tandeu de Marsac and R. Rippka, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 1053–1058.

113 H. Wang, D. P. Fewer, L. Holm, L. Rouhiainen and K. Sivonen, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 9259–9264.

114 M. A. Skinnider, N. J. Merwin, C. W. Johnston and N. A. Magarvey, *Nucleic Acids Res.*, 2017, **45**, W49–W54.

115 D. Mao, B. K. Okada, Y. Wu, F. Xu and M. R. Seyedsayamdst, *Curr. Opin. Microbiol.*, 2018, **45**, 156–163.

116 K. Saurav, A. Caso, P. Urajová, P. Hrouzek, G. Esposito, K. i. Delawská, M. Macho, J. Hájek, J. Cheel and S. Saha, *ACS Omega*, 2022, **7**, 11818–11828.

117 P. Hrouzek, M. Kuzma, J. Černý, P. Novák, R. Fišer, P. Šimek, A. Lukešová and J. i. Kopecký, *Chem. Res. Toxicol.*, 2012, **25**, 1203–1211.

118 P. Khurana, R. S. Gokhale and D. Mohanty, *BMC Bioinf.*, 2010, **11**, 1–17.

119 M. Röttig, M. H. Medema, K. Blin, T. Weber, C. Rausch and O. Kohlbacher, *Nucleic Acids Res.*, 2011, **39**, W362–W367.

120 M. G. Chevrette, F. Aicheler, O. Kohlbacher, C. R. Currie and M. H. Medema, *Bioinformatics*, 2017, **33**, 3202–3210.

121 C. Rausch, T. Weber, O. Kohlbacher, W. Wohlleben and D. H. Huson, *Nucleic Acids Res.*, 2005, **33**, 5799–5808.

122 G. Pan, Z. Xu, Z. Guo, Hindra, M. Ma, D. Yang, H. Zhou, Y. Gansemans, X. Zhu and Y. Huang, *Proc. Natl. Acad. Sci. U. S. A.*, 2017, **114**, E11131–E11140.

123 H. Zhao, Y.-P. Liu and L.-Q. Zhang, *Front. Microbiol.*, 2019, **10**, 544.

124 B. I. Khayatt, L. Overmars, R. J. Siezen and C. Francke, *PLoS One*, 2013, **8**, e62136.

125 D. Baranašić, J. Zucko, J. Diminic, R. Gacesa, P. F. Long, J. Cullum, D. Hranueli and A. Starcevic, *J. Ind. Microbiol. Biotechnol.*, 2014, **41**, 461–467.

126 T. V. Lee, R. D. Johnson, V. L. Arcus and J. S. Lott, *Proteins: Struct., Funct., Bioinf.*, 2015, **83**, 2052–2066.

127 C.-H. Ji, S. Park, K. Lee, H.-W. Je and H.-S. Kang, *J. Am. Chem. Soc.*, 2024, **146**, 30434–30442.

