Docking domain-mediated subunit interactions in natural product megasynth(etas)es

Helen G. Smith 1,2, Matthew J. Beech 2, Józef R. Lewandowski 2, Gregory L. Challis 2,3,4, Matthew Jenner 2,3

1Warwick Medical School, University of Warwick, Coventry CV4 7AL, UK
2Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK
3Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, UK
4Warwick Biochemistry and Molecular Biology Centre, University of Warwick, Coventry CV4 7AL, UK
5ARC Centre of Excellence for Innovations in Peptide and Protein Science, Monash University, Clayton, VIC 3800, Australia

Correspondence should be addressed to: Gregory L. Challis at g.l.challis@warwick.ac.uk and Matthew Jenner at m.jenner@warwick.ac.uk

Abstract: Polyketide synthase (PKS) and non-ribosomal peptide synthetase (NRPS) multienzymes produce numerous high value metabolites. The protein subunits which constitute these megasynth(etas)es must undergo ordered self-assembly to ensure correct organisation of catalytic domains for the biosynthesis of a given natural product. Short amino acid regions at the N- and C-termini of each subunit, termed docking domains (DDs), often occur in complementary pairs, which interact to facilitate substrate transfer and maintain pathway fidelity. This review details all structurally characterised examples of NRPS and PKS DDs to date and summarises efforts to utilise DDs for the engineering of biosynthetic pathways.

Keywords: Polyketide synthase, Non-ribosomal peptide synthetase, Biosynthesis

Graphical abstract

Docking domains mediate the ordered self-assembly of giant megasynth(etas)es such as PKSs and NRPSs, this article outlines current understanding of their structure, function and application in bioengineering.

Introduction

Modular polyketide synthases (PKSs) and non-ribosomal peptide synthetases (NRPSs) represent two classes of extraordinary molecular machines, responsible for the biosynthetic assembly of polyketide and non-ribosomal peptide natural products, respectively. Over the past 30 years, the chemical products of these ‘megaenzymes’ have been the focus of extensive research due to the potential of these molecules as sources of new pharmaceuticals and agrochemicals (Newman & Cragg, 2020). Frequently equated to molecular assembly lines, modular PKSs and NRPSs typically consist of multiple large protein subunits comprised of discretely-folded catalytic domains organised into modules with overall molecular weights in the MDa range (Weissman & Muller, 2008). Both PKS and NRPS paradigms employ carrier protein domains; an acyl carrier protein (ACP) domain in modular PKSs and a peptidyl carrier protein (PCP) domain in NRPSs, which are post-translationally modified via tethering of a 4′-phosphopantetheine (Ppant) group to a conserved serine residue (Crosby & Crump, 2012). This facilitates covalent attachment of the biosynthetic intermediates via a thioester linkage, whilst also providing the required flexibility to visit active sites of other enzymatic domains within the module (Fischbach & Walsh, 2006; Hertweck, 2009).

Polyketide biosynthesis involves the head-to-tail condensation of acyl and malonyl-derived thioester units, similar to the catalytic cycle of fatty acid biosynthesis (Smith & Tsai, 2007;
Fig. 1. (a) Partial domain organisation of a hypothetical PKS assembly line. Enzymatic domains are represented by spheres and biosynthetic intermediates are shown appended to each ACP domain. Subunit and module labelling conventions are highlighted above the PKS, and features relating to the main text are highlighted. (b) Domain organisation of cis-AT (left) and trans-AT (right) PKS modules. The stand-alone AT domain in trans-AT PKS loads extender units onto multiple ACP domains, whereas the AT domain in cis-AT PKSs loads the ACP domain within its module. (c) Partial domain organisation of a hypothetical NRPS assembly line. Enzymatic domains are represented by spheres and biosynthetic intermediates are shown appended to each PCP domain. Subunit and module labelling conventions are highlighted above the NRPS, and features relating to the main text are highlighted. Domain abbreviations are as follows: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; ACP, acyl carrier protein; C, condensation; A, adenylation; PCP, peptidyl carrier protein; CDD, C-terminal docking domain; NDD, N-terminal docking domain.

Staunton & Weissman, 2001). During polyketide biosynthesis the catalytic cycle of a module commences when an acyltransferase (AT) domain loads an (alkyl)malonyl extender unit onto the Ppant thiol of the ACP domain. The ketosynthase (KS) domain then catalyses a decarboxylative Claisen condensation between an incoming extender unit and the upstream acylthioester intermediate, yielding a β-keto-thioester. Optional catalytic domains within the module, such as ketoreductase (KR), dehydratase (DH), enoylreductase (ER) and methyltransferase (MT) domains, allow modification of the α- and β-carbons of the β-keto-thioester (Keatinge-Clay, 2012; Khosla et al., 2007) (Fig. 1a). Modular PKSs can be sub-divided into two phylogenetically distinct classes; the
cis-AT and trans-AT PKSs. In cis-AT PKS systems an AT domain is integrated into each module of the PKS. In trans-AT PKSs the modules lack AT domains and a stand-alone AT acts in trans to functionalise each ACP domain (Helfrich & Piel, 2016, Kosol et al., 2018) (Fig 1b).

NRPSs employ a similar biosynthetic logic to modular PKSs, where each module harbours the enzymatic domains required for selection, activation and incorporation of amino acid building blocks into the growing peptidyl chain (Reimer et al., 2018; Süssmuth & Mainz, 2017). In this case, an adenylation (A) domain specifically selects and activates an amino acid via adenylation of the carboxyl group, permitting subsequent tethering to the Pcpt thiol of the Pcp domain (Challis et al., 2000; Stachelhaus et al., 1999). A condensation (C) domain then catalyses peptide bond formation between the Pcp-bound aminoacetyl thioester and the peptidyl thioester intermediate attached to the upstream Pcp domain (Fig 1c). Common optional domains in NRPS machinery include an epimerisation (E) domain, which catalyses inversion of the stereochemistry at the α-carbon, and an N-methyltransferase (N-MT) domain that methylates the nitrogen of the peptide linkage (Miller & Gulick, 2016).

In both PKSs and NRPSs, the catalytic domains are often split across multiple large protein subunits. These subunits must undergo ordered self-assembly to ensure the directionality of the assembly line is maintained, and to guarantee biosynthetic fidelity. Short regions of amino acids situated at the N- and C-termini of multi-modular subunits have been shown to direct ordered assembly of both PKSs and NRPSs (Broadhurst et al., 2003; Kosol et al., 2018, Richter et al., 2008) (Fig 1a and c). These regions are known as docking domains (DDs), and tend to occur in complementary pairs that interact in the low μM range, but specifically enough to maintain biosynthetic fidelity (Dodge et al., 2018, Kosol et al., 2018; Miyanaga et al., 2018). Their discovery has presented a unique opportunity to expand the combinatorial potential of PKS and NRPS assembly lines via grafting of DD pairs onto heterologous modules to create novel chemical scaffolds. However, to achieve success in such bioengineering experiments, the nature of the docking interfaces needs to be understood in molecular detail.

Over the last 20 years, there have been several efforts to elucidate the structures and molecular interaction mechanisms involved in modular PKS and NRPS DD interfaces. Structural studies of DDs are often undertaken via NMR spectroscopy or X-ray crystallography and have required production of a covalently fused C-terminal DD (cis-DD), and N-terminal DD (trans-DD) complex, as one or both DDs within the pair may only fold correctly in the presence of its cognate partner (Broadhurst et al., 2003). More recently, newer techniques such as small angle X-ray Scattering (SAXS) (Dorival et al., 2016, Risser et al., 2020) and carbene footprinting mass spectrometry have been employed (Jenner et al., 2018; Kosol et al., 2019), alongside traditional techniques to elucidate both DD structure and molecular interactions acting across the interface.

This review details all characterised examples of NRPS and PKS DDs to date and summarises efforts to utilise these to engineer PKS and NRPS biosynthetic pathways. The potential of DDs to be exploited for assembly line engineering is considerable, and as the mechanisms underpinning DD interactions are further elucidated it is expected that their use in pathway engineering will be expanded to a wider variety of systems.

PKS DDs

PKS DDs can be categorised by the type of system they occur in: cis- or trans-AT PKSs. It is worth noting that the subunit junctions of cis-AT PKSs are predominantly intermodular, resulting in many examples of ACP/KS domain junctions. Trans-AT PKSs, on the other hand, tend to have intramodular subunit junctions giving rise to multiple ‘split module’ architectures (e.g. KS/KR, KS/DH and DH/KR domain junctions) (Helfrich & Piel, 2016). Perhaps because of this, there is more structural diversity in the type of DDs that occur in trans-AT PKS systems.

Cis-AT PKS DDs

Cis-AT PKS DDs are split into classes according to the type of organism that produces the metabolite; Class 1 DDs have been proposed to be present primarily in Actinobacterial modular PKSs, whereas Class 2 DDs are proposed to be found predominantly in modular PKSs from Cyanobacteria, Myxobacteria and other Gram-negative bacteria (Broadhurst et al., 2003; Buchholz et al., 2009; Whicher et al., 2013). Class 2 cis-AT PKS DDs are found both in purely PKS systems and in hybrid PKS–NRPSs, but will be discussed in this section (Whicher et al., 2013).

Class 1 cis-AT PKS DDs: Four α-helix bundles

Class 1 PKS DDs occur exclusively at ACP/KS subunit junctions, with complementary DD regions at the C-terminus of the ACP domain (cis-DD), and the N-terminus of the KS domain (trans-DD). Three α-helices (α1–α3) comprise the cis-DD region; α1 and α2 form a dimerisation element, and α3 is directly involved in contacts with the trans-DD (Broadhurst et al., 2003). It is worth noting that, unlike NRPS systems, PKSs are dimeric and consequently a total of six α-helices makes up the cis-DD region. The trans-DD connected to the downstream KS domain consists of a single α-helix (α4), which forms a coiled-coil with an identical trans-DD in the dimeric KS (Fig 2a and b). The shorter α3 helices from the cis-DD interact with the trans-DD by clamping to each side of the coiled-coil to form a four α-helix bundle complex (Broadhurst et al., 2003) (Fig 2b).

Structurally characterised examples of actinobacterial Class 1 PKS DDs include the DEBS2/DEBS3 interface from the erythromycin PKS (Broadhurst et al., 2003) and the PikAIII/PikAIV interface from the pikromycin PKS (Buchholz et al., 2009). Interestingly, a Class 1 DD has also been characterised from the Bamb_5920/Bamb_5919 interface from the encycloyxin Iia hybrid PKS–NRPS, in Gram-negative Burkholderia species (Risser et al., 2020) (Fig 2).

The DEBS2/3 DD pair was the first to be structurally characterised. Covalently tethered cis-DD(α1–α2) and trans-DD(α3) complexes were analysed by solution state NMR spectroscopy and both the structure and molecular interactions across the interface were elucidated. Interactions between the trans-DD α-helices forming the dimeric coiled-coil were found to be predominantly hydrophobic, though some salt bridges were present. Sequence alignments with other Class 1 DDs showed that the residues involved in salt bridge formation were conserved across systems, indicative of their importance in creation of the coiled-coil interaction interface. Important interactions between the cis-DD (α1) and trans-DD of this interface were also identified as involving a combination of hydrophobic contacts and salt bridges. Hydrophobic contacts highlighted in Fig 2b were found to be highly conserved across multiple Class 1 DDs. Examples include F67 on α3, Y96 on α4 and L97 on α4′ (Fig 2c and d). Further sequence alignments led to the hypothesis that the salt bridges, such as those formed between R73 of α1 and D105 of α4′, confer specificity to the interface. On interaction of non-cognate DDs, repulsive interactions would occur at these positions, this would prevent docking of non-cognate pairs and ensure biosynthetic fidelity is maintained (Broadhurst et al., 2003).
Fig. 2. Structural features, sequence alignments and use in biosynthetic engineering of Class 1 PKSs DDs. (a) Domain architecture of the DEBS2–DEBS3 intersubunit junction from the PKS involved in the biosynthesis of erythromycin A. (b) Solution state NMR structure of the 4-α-helix bundle docked complex formed by the covalently tethered C-terminal and DEBS3 N-terminal DDs (PDB accession code: 1PZR). A dimerisation motif is found upstream of the docking interface and comprises four additional helices (PDB accession code: 1PZQ). Inset (top): Key electrostatic interactions between helix α3 and helices α4 and α4′ that confer specificity to the docking interface. Inset (bottom): Hydrophobic interface formed between the α3, α4 and α4′ helices. (c) Sequence alignment of select 4-α-helix bundle C-terminal docking domains (DDs), including the dimerisation motif. (d) Sequence alignment of select 4-α-helix bundle N-terminal DDs. Asterisks (*) denote the positions of the interfacial residues highlighted in (b). Red chevrons (>) denote interacting electrostatic residues highlighted in (b) where charge is not conserved across DDs. Above the alignments, a schematic displaying the positions of the secondary structural elements from the solution state NMR structure of the DEBS2–DEBS3 DD complex is provided. Residue numbering provided in (b), (c) and (d) is relative to that from PDB accession code 1PZQ. (e) Artificial splitting of the PikAI subunit from the PKS responsible for pikromycin biosynthesis using four α-helix bundle DDs from the Plm1–Plm2 subunit junction in phoslactomycin biosynthesis. Quantification of pikromycin production was determined by HPLC analysis of culture extracts from a *Streptomyces venezuelae* ΔpikAI mutant complemented with the engineered PikAI proteins in trans.

Structural and interaction data for the PikAIII/PikAIV DD pair were obtained via crystallisation of a covalently tethered 7DD–8DD complex. The structure and key amino acid contacts, both between the 7DD and 8DD and between the two α-helices of the coiled–coil, were found to be similar to that characterised from the erythromycin system. However, the specificity conferring ionic contact between R73 and D105 in the DEBS2-3 system is not maintained. The PikAIII 7DD is shorter than the DEBS2 7DD, at only nine amino acids compared to fifteen. It is therefore understandable that electrostatic interactions across the interface may occur at alternate positions. Sequence alignments of DEBS2 and PikAIII DDs with other DDs from the erythromycin and pikromycin biosynthetic pathways found that charge reversal of a single amino acid residue was enough to
confer specificity to these interfaces. This highlights the importance of these electrostatic interactions in driving ordered self-assembly of the modular PKS subunits (Buchholz et al., 2009).

A covalently tethered DD-DDDD complex also enabled structure elucidation of the Bamb_5920/5919 DD pair from the eacycloxil Ila biosynthetic system. Solution-state NMR spectroscopy revealed a complex with a highly similar structure to those previously elucidated. However, the Bamb_5920 DD lacks the double helix dimerisation motif present in the erythromycin system (Fig. 2b), indicating that the dimerisation motif may not be essential for docking. The Bamb_5920/5919 docking interface was, as with other Class 1 DDs, found to interact via a combination of hydrophobic contacts and salt bridges. However, attempts to align these DD sequences with other characterised examples were problematic and highlighted their divergent nature (Fig. 2c and d). This could perhaps be a consequence of the DD originating from a hybrid PKS–NRPS, leading to evolutionary divergence at the sequence level despite high structural similarity. Nevertheless, as this DD is the first characterised example of a Class 1 DD pair identified in a hybrid PKS–NRPS it offers the potential to facilitate the production of hybrid biosynthetic pathways. As further examples of this class of DD are identified in hybrid systems, more will be understood about their modes of interaction and specificity-conferring mechanisms (Risser et al., 2020).

Class 1 PKS DDs were identified via sequence analysis several years prior to elucidation of the structure of the DEBS2-3 DD pair. Many early experiments were undertaken on the DEBS PKS both at the DEBS1-2 and the DEBS2-3 interaction interface to better understand the role of DDs. These experiments included overproduction of modules lacking their DDs. These remained functional, showing their non-essentiality for modular PKS function in vitro (Broadhurst et al., 2003). Replacement of native DD pairs with cognate pairs from other biosynthetic pathways allowed in vitro interaction of modules with a K_5 almost identical to the wild-type system (Broadhurst et al., 2003). Cognate DD pairs were found to enable in vitro interaction of modules from different biosynthetic pathways, as demonstrated by creation of a hybrid rifamycin-DEBS PKS (Gokhale et al., 1999). Introduction of a non-cognate partner DD to a modular PKS interface caused a large decrease in product titre, but surprisingly production was not completely abolished (Tsuji et al., 2001). This provided preliminary evidence that more than just interaction of DD pairs is necessary for productive protein–protein interaction and substrate transfer across subunit junctions. At the point the first structure of a DD pair was elucidated, much was already known about both the portability and necessity of these interacting regions.

More recently, in vivo pathway engineering was attempted on the pikromycin PKS. A cognate DD pair was found to facilitate interaction between modules that would normally be covalently tethered. However, this was accompanied by a dramatic decrease in product yield (Yan et al., 2009) (Fig. 2e). This may indicate the presence of secondary interactions at this interface between the ACP and KS domains, which are essential for productive interaction.

**Class 2 cis-AT PKS DDs: Eight α-helix bundles**

There are three characterised examples of Class 2 PKS DDs: two from the curacin modular PKS at the CurG/H and CurK/L subunit junctions (Whicher et al., 2013), and one from the enacycloxil hybrid PKS–NRPS at the Bamb_5925/5924 subunit junction (Risser et al., 2020). All characterised examples of Class 2 DDs form dimeric eight α-helix bundles, with each DD contributing two α-helices (Fig. 3a–c). Much like Class 1 PKS DDs, a parallel coiled–coil formed by the first DD α-helix interacting with its symmetry mate (α_4 and α_4') is at the centre of the bundle. The two DD α-helices both interact with the HI DD coiled–coil. This interaction interface is predominantly hydrophobic, with salt bridges hypothesised to confer specificity (Fig. 3d–f). However, the sites of these specific interactions appear to vary. For example, interactions for the CurK/L complex, as highlighted in Fig. 3e, are between E24/K70 and E24/R73 (Whicher et al., 2013). In Bamb_5925/24 K70 forms salt bridges with E5 and E8 on helix α_1 (Fig. 3f), rather than an amino acid on helix α_2 as in the CurK/L complex (Risser et al., 2020) (Fig. 3e). Furthermore, a productive electrostatic interaction does not exist at these sites for the CurG/H interface (Fig. 3d).

There is a much greater level of conservation at the positions of hydrophobic contact across all characterised examples of Class 2 DDs (Fig. 3b and c). For example, the hydrophobic amino acids highlighted in CurG/H in Fig. 3d; L6, I9, L17, L59, A66 and L67, are conserved in both CurK/L and Bamb_5925/24 (Fig. 3e and f). There is a high level of structural similarity between CurG/H and Bamb_5925/24. However, CurK/L has a much shorter primary DD α-helix (α_1) and lacks the primary γ-DD α-helix (α_3) observed for CurH and Bamb_5924 2 DDs. Sequence analysis shows that three N-terminal amino acids are missing from CurL, which may prevent the γ-DD from folding correctly. However, structurally the CurK/L complex appears closer to that of a Class 1 four α-helix bundle than that of a Class 2 system, despite its sequence aligning well to those of other Class 2 PKS DDs (Risser et al., 2020; Whicher et al., 2013). Once the structures of additional DD pairs have been elucidated, it will become clear whether Class 2 PKS DDs should be split into further subcategories according to their docked complex structure.

Despite the overall similarity of Bamb_5925/24 to CurG/H, there are several distinct differences between these two DD pairs (Risser et al., 2020). The second α-helix of the Bamb_5924 2 DD adopts a very different orientation on docking compared to the CurH 2 DD (Risser et al., 2020). Furthermore, all four 2 DD α-helices interact at the coiled–coiled interface (Fig. 3d, compared to only those involved in coil formation in the CurK/L complex). These factors all contribute to a more compact eight helix bundle at the Bamb_5925/24 interface compared to CurG/H (Risser et al., 2020). This may be a feature common to eight α-helix bundles in hybrid PKS–NRPSs.

Interestingly, the CurK 2 DD was found to be promiscuous, interacting with both CurH and CurM 2 DDs, albeit with a reduced affinity (Whicher et al., 2013). This observation is difficult to rationalise based on inspection of the amino acid sequences alone. However, it represents the sole example of PKS DDs working outside the confines of their cognate pairs and suggests that productive DD interactions may not be solely responsible for maintenance of biosynthetic fidelity. Substrate specificity and gating, or domain–domain interactions across subunit junctions are also likely to be important in productive protein–protein interactions.

Engineering of the pikromycin system was undertaken using curacin DD pairs, the CurG/H and CurK/L pairs were used to replace native DDs at the FikAIII/IV interface (Fig. 3g) (Whicher et al., 2013). The pikromycin PKS produces two products: 10-deoxymethynolide (10-dml), a methymycin precursor, and narbonolide (nbl), a pikromycin precursor. While overall product yield remains constant on exchange of Class 1 PKS DDs with their Class 2 counterparts, the product profile changes from 50% nbl to predominantly 10-dml (Whicher et al., 2013). This indicates that Class 2 DDs may promote different subunit interactions to Class 1 DDs,
Fig. 3. Structural features, sequence alignments and use in biosynthetic engineering of PKS Class 2 DDs. (a) Domain architecture of the CurG-CurH, CurK-CurL and Bamb_5925-Bamb_5924 intersubunit junctions from the hybrid PKS-NRPSs responsible for the biosynthesis of curacin A and encacyloxin IIa. Cognate DD pairs are depicted using complimentary fitting shapes. (b) Sequence alignment of select eight α-helix bundle C-terminal DDs. (c) Sequence alignment of select eight α-helix bundle N-terminal DDs. Asterisks (*) denote the positions of the conserved hydrophobic interfacial residues highlighted in (d). Above the alignments, schematics displaying the positions of the secondary structural elements observed in the structures of each of the DD complexes is provided. (d) X-ray crystal structure of the eight α-helix bundle docked complex formed by the covalently tethered CurG C-terminal and CurH N-terminal DDs (PDB accession code: 4MYY). Inset: Hydrophobic residues across α₁, α₂, α₄, α₃′ and α₄′ implicated in the formation of the CurG–CurH DD interface. (e) X-ray crystal structure of the docked complex formed by the covalently tethered CurK C-terminal and CurL N-terminal DDs (PDB accession code: 4MYZ). Inset: Electrostatic interactions between the α₂′ and α₄ helices. The same interactions are not conserved in the CurG–CurH interface. (f) Solution state NMR structure of the docked complex formed by the covalently tethered Bamb_5925 C-terminal and Bamb_5924 N-terminal DDs (PDB accession code: 6TDN). Inset: Electrostatic interactions between α₁ and α₄ (top) and α₁ and α₃ (bottom). The former is also observed in the CurG–CurH interface. Residue numbering in (b)–(e) is relative to that of PDB entry 4MYY. (g). Engineering of the PikAIII–IV intersubunit junction from the pikromycin PKS. This PKS produces two products, pikromycin, of which narbolide (nbl) is the precursor, and methymycin, for which 10-deoxymethynolide (10-dml) is the precursor, resulting from a module-skipping mechanism. Exchanging the WT four α-helix bundle DD pair at this junction with the eight α-helix bundle CurG-H or CurK-L DD pair was demonstrated to maintain productive interaction between proteins by in vitro assays. Additionally, introducing eight α-helix bundle DDs was found to change the product profile from 50% nbl, to predominantly 10-dml, suggesting more effective delivery of the PikAIII ACP-tethered substrate to the PikAIV TE domain. Percentage production values are given with respect to the WT DD pair.

perhaps demonstrating that they permit increased sampling of the terminal end of the downstream module.

Trans-AT PKS DDs

Unlike cis-AT systems, trans-AT PKSs possess both inter- and intramodular subunit junctions. Consequently, these subunit junctions are located between a variety of domains, giving rise to ‘split module’ architectures such as KS/KR, KR/DH and KS/DH domain interfaces, amongst others (Dorival et al., 2016; Jenner et al., 2018; Zeng et al., 2016). There are two key classes of DDs reported in trans-AT PKSs which account for a large number of the subunit junctions: the four α-helix bundle DDs and dehydratase docking (DHD) domains.

Four α-helix bundles

Four α-helix bundle DDs were identified as regions of ∼25 amino acids at the N- and C-termini of protein subunits in trans-AT PKSs. While these DD regions occur at traditional intermodular ACP/KS subunit junctions, they are also found at many of the unusual
Intramodular subunit junctions responsible for ‘split module’ domain architectures (Dorival et al., 2016; Zeng et al., 2016).

The VirA–FG ACP/KS interface in the virginiamycin PKS was the first characterised example of a four-helix bundle DD (Dorival et al., 2016). Solution state NMR spectroscopy of the covalently tethered CDD–NDD complex elucidated both the docked structure and molecular interaction mechanisms involved in this interface. In the docked complex, both the NDD and CDD regions adopt two α-helices, forming a helix-turn-helix motif on each side of the four α-helix bundle. Two sets of α-helices are offset by 127° in the complex, producing an interaction interface that encompasses all four α-helices (Fig. 4c and d) (Dorival et al., 2016). While trans-AT PKSs are indeed dimeric, SAXS data suggests that this does not result in dimerisation of the C-DD and N-DD regions. Instead, there are two separate copies of the VirA–FG DD pair. Consequently, we propose that, rather than describing this interface as a further example of a four α-helical bundle, it should be termed a ‘double-helix pair’. This allows it to be categorised separately from Class 1 cis-AT PKS DDs, accounting for differences in both structure and interaction interface (Dorival et al., 2016). The core of the VirA–FG interface is hydrophobic with C-DD residues L16, L20, I30 and V34 interacting with A53, L57, F60 and L74 of the NDD, as highlighted in Fig. 4d. Key salt bridges thought to confer specificity are D17/R66, N26/K61 and E31/K54.

Further work investigating intersubunit interactions in the macrolactin trans-AT PKS yielded an X-ray crystal structure of the N-terminal KR domain from the MlnE subunit with the NDD appended, allowing visualisation of where the DD sits with respect to the KR domain (Zeng et al., 2016). In this instance the upstream subunit, MlnD, has a KS domain at its C-terminus, an example of the KS/KR ‘split module’ architecture. Other ‘split module’ interfaces in the macrolactin system harbouring DD pairs were examined by analytical gel filtration and isothermal titration calorimetry (ITC) for their ability to form complexes. This showed that non-cognate C-DD–N-DD pairs were unable to form stable complexes (Fig. 4a and b). However, DD swapping experiments showed that cognate C-DD–N-DD pairs could be appended to different catalytic domains and still form a functional complex (Zeng et al., 2016). This highlights the portability of these short DD pairs and their potential utility for biosynthetic engineering efforts (Meinke et al., 2019).

**DHD domains**

Recent work on the gladiolin trans-AT PKS identified a new class of DD which occurs solely at KS/DH domain junctions (Jenner et al., 2018). Sequence-level inspection of C- and N-termini at these interfaces revealed a ~70 amino acid region was appended to the C-terminus of the KS domains. However, no additional region could be identified at the N-terminus of the DH domains. In vitro acyl transfer and mechanism-based crosslinking experiments showed that the region appended to the C-terminal of the KS domain was critical for functional complex formation and was therefore termed a dehydratase docking (DHD) domain. Application of solution state NMR and circular dichroism spectroscopy to the isolated DHD domain revealed a limited propensity to form secondary structure elements in solution, and inherent intrinsic disorder NMR titrations of the 15N-labelled DHD domain with the DH domain identified two interacting regions, indicating that the former interacts directly to the exterior surface of the latter. This was confirmed by carbene footprinting mass spectrometry analysis, resulting in masking of regions on the DH domain surface upon incubation with the DHD domain (Jenner et al., 2018) (Fig. 5a, c and d).
**Fig. 5.** Structural features and sequence alignments of PKS DHD domains and corresponding DH domains. (a) Domain architecture (top) and structural model (bottom) of GbnD4–GbnD5 KS-DH inter-subunit junction from the gladiolin trans-AT PKS. Regions highlighted in red on the DHD domain and DH domain have been shown to interact. (b) Domain architecture (top) and X-ray crystal structure (bottom) of RhiE KS-B di-domain from the rhizoxin trans-AT PKS (PDB accession code: 4KC5). The B domain is structurally homologous to a DH domain, and the region connecting the KS to the B domain is highlighted in red. The flanking subdomain (FSD) is highlighted, which is absent from KS domains at KS–DH junctions. (c) Sequence alignment of selected DHD domains from the C-termini of KS domains. (d) Sequence alignment of selected N-terminal DH domains corresponding to the DHD domains in (c). Asterisks (*) denote the positions of the interfacial residues highlighted in (a), as observed experimentally by NMR spectroscopy and carbene footprinting mass spectrometry. Domain abbreviation: B, branching domain.
Although precise docking orientation and specific amino acid pairwise interactions for this interface are yet to be elucidated, some global insights into the organisation of the KS/DH interface can be inferred from the X-ray crystal structure of the RhIE KS-B di-domain (Bretschnieder et al., 2013). The B domain, which in this instance forms a structural scaffold for a chain branching reaction in rhizobin biosynthesis (Partida-Martinez & Hertweck, 2007), has the same characteristic double-helical fold as a DH domain (Keatinge-Clay, 2012). The structure of the di-domain highlights that the KS and the B domains are connected via a long stretch of residues that lack secondary structure elements prior to the start of the B domain, where a set of antiparallel β-sheets form in addition to a short helix (Fig. 5b). Furthermore, the contacts that these parts of the KS-B linker make with the B domain itself correlate well with the interacting regions on the surface of the DH domain identified by carbene footprinting (Fig. 5a).

Interestingly, the KS domain at DHD domain-containing interfaces lacks a flanking subdomain (FSD); a region of protein that is structurally important at the KS–AT interface in trans-AT PKSs and is believed to be an evolutionary relic in trans-AT PKS systems (Gay et al., 2014, 2016). The significance of this is currently unclear but given that the FSD would naturally precede the DHD domain, it may have been lost to permit close association of the KS and DH domains during complex formation.

NRPS DDs

NRPSs are responsible for the biosynthesis of non-ribosomal peptide natural products. A minimal NRPS module comprises an adenylation (A) domain, a condensation (C) domain and a PCP domain. A domains select and activate specific amino acids as aminoacyl thioesters. C domains then catalyse amide bond formation between successive PCP-bound aminoacyl thioesters resulting in assembly of a peptide chain. There are two known classes of DD that have been reported to occur exclusively in NRPS systems; communication (COM) domains (Hahn & Stachelhaus, 2004) and peptide-antimicrobial-Xenorhabdus (PAX) domains (Watzel et al., 2020). A third class, the β-hairpin docking (βD) domains have been observed in both NRPSs and hybrid PKS-NRPSs and will be discussed under the hybrid PKS–NRPS section (Dowling et al., 2016; Richter et al., 2008).

COM Domains: Helix-Hand Motif

Communication-mediating or COM domains have been identified at multiple E/C domain junctions in NRPSs (Chiocchini et al., 2006; Hahn & Stachelhaus, 2004). As was observed for the DEBS DDs in cis-AT PKSs, many experiments were undertaken to increase understanding of the role of COM domains prior to elucidation of their structure (Hahn & Stachelhaus, 2006; Siewers et al., 2010). A possible donor COM (COMd) domain structure was obtained relatively serendipitously. Upon crystallisation of the final module of the surfactin NRPS (SrFAC; COM4–A–PCP–TE) the structure of the N-terminal acceptor COM (COMa) domain was identified (Tanovic et al., 2008). This was somewhat unexpected, as DDs are often highly flexible, unstructured proteins which may only interact, this led to production of dipeptide and tripeptide products (Hahn & Stachelhaus, 2006) (Fig. 6e). However, this work was carried out prior to elucidation of the structure of COM domains. Therefore, only the N-terminal portion of the hand motif ($\alpha_1$ and $\beta_1$) was resolved onto BacB. This innate promiscuity of COM domains is further evidenced by crossstalk experiments employing the surfactin, tyrocidine and gramicidin S NRPS pathways. The non-cognate TyCA COM4 and TyCC COM4 domains from the tyrocidine NRPS were used to replace the cognate SrFA-A/SrFB-B COM and COM domain pair in the surfactin NRPS in B. subtilis CC112 (Chiocchini et al., 2006). As the TyCA COM4 domain interacts with the surfactin SrFA-C COM4 domain, this led to skipping of SrFA-B and production of a lipotetrapeptide product, rather than the full-length lipopeptapeptide (Chiocchini et al., 2006). Furthermore, in vitro work on the gramicidin S NRPS showed GrsA was able to interact directly with TyCB from the tyrocidine NRPS resulting in production of cyclo-d-Phe-l-Pro-diketopiperazine (Torsten Stachelhaus et al., 1998). The alignments provided in Fig. 6c and d show a high level of sequence conservation across all COMd and COMd domain pairs. Therefore their inherent promiscuity is perhaps unsurprising.

As the COM4 domain is not wholly independent of its tethered catalytic domain (i.e. structural elements are contributed from other sections of the polypeptide chain), the question of whether it would be possible to cleanly ‘cleave’ COMa regions for NRPS pathway engineering remains unclear. For engineering purposes, it may be more productive to focus on utilising the inherent promiscuity of COM domains and on COMd domain-based grafting studies, before attempting COMa domain-based engineering.

PAX DDs: Three α-Helix Bundles

A new class of NRPS DD was recently identified via analysis of the PaxB-C PCP/C domain junction from the PAX assembly line
Fig. 6. Structural features, sequence alignments and use in biosynthetic engineering of NRPS COM domains. (a) Domain organisation of the SrfAB–SrfAC intersubunit junction. (b) X-ray crystal structure of the SrfAC condensation domain (PDB accession code: 2VSQ). Inset: the helix-hand motif is formed of a helix and a beta sheet comprising three non-contiguous strands. A portion of the C-terminal protein tag of SrfAC, shown in yellow, was found to interact with the helix-hand motif and is proposed to mimic the NDD helix. (c) Sequence alignment of selected N-terminal COMA domains, encompassing the two regions of protein comprising the helix-hand motif. Residue numbering throughout is relative to that of PDB entry 2VSQ. (d) Sequence alignment of selected C-terminal COMD domains. The sequence of the interacting region of the SrfAC C-terminal tag is provided and aligned as described by Tanovic et al. (e) Engineering of intermodular interfaces using the TycA–TycB COM domain pairs to mediate productive crosstalk between non-cognate NRPS modules from the tyrocidine, bacitracin and surfactin A assembly lines. Successive interactions of TycAΔ–BacB2 and BacB2–SrfAC indicated by the blue arrows lead to formation of the tripeptide shown in the blue box, while the direct interaction between TycAΔ–SrfAC indicated by the grey arrow leads to the dipeptide product shown in the grey box.

(Watzel et al., 2020) (Fig. 7a). Solution state NMR spectroscopy of a covalently tethered DD–NDD complex illuminated both the structure of the DDs and the molecular interactions across the interface. The NDD comprises a single a-helix which docks centrally within the two a-helix DD to give a three a-helix bundle with a V-shaped overall structure (Fig. 7b) (Watzel et al., 2020). As for most other DDs, salt bridges confer specificity, in this case K19/E75 and R23/D68. The remaining interactions are hydrophobic. Key hydrophobic residues include L8, L11, V15 and L16 on the NDD and L72, L76, L84 and L88 on the CDD (Fig. 7b–d). Bioinformatics analyses predicted further examples of this class of DD bound to C, PCP, ACP and oxidoreductase domains, showing their portability and potential utility for NRPS pathway engineering (Watzel et al., 2020).

Hybrid PKS–NRPS DDs

Hybrid PKS–NRPSs assemble mixed polyketide/non-ribosomal peptide natural products composed of both amino acids and (alkyl)malonyl-derived building blocks. These systems illustrate that, with the right tools, it may be possible to engineer pathways involving both NRPS and PKS biosynthetic machinery, greatly increasing the diversity of the products.

β-Hairpin DDs

β-Hairpin DDs (βhD domains) have been identified in both purely NRPS and hybrid PKS–NRPS pathways, at PCP/C or PCP/Cy domain interfaces (Dowling et al., 2016; Hacker et al., 2018; Kosol et al., 2019; Richter et al., 2008). An intrinsically disordered short linear motif (SLIM) DD of less than 15 amino acids, interacts with the much larger N-terminal βhD domain (Hacker et al., 2018; Kosol et al., 2019). This class of DD is perhaps the most well characterised with structures of six different βhD domains deposited in the PDB; three from the rhabdopeptide NRPS (solution state NMR structures) (Hacker et al., 2018), one from the tubulysin PKS–NRPS (solution state NMR structure) (Richter et al., 2008), one from the epothilone PKS–NRPS (X-ray crystal structure) (Dowling et al., 2016) and one from the encacyloxyx Ila PKS–NRPS (X-ray crystal structure) (Kosol et al., 2019) (Fig. 8a and b). The structure of the N-terminal βhD domain was found to be highly conserved across all systems adopting an αββαα fold with a central β-hairpin. βhD domains are connected to the downstream domain via flexible linker regions of approximately 20 amino acids (Dowling et al., 2016; Kosol et al., 2019). The EpoB crystal structure shows the βhD domain in three distinct conformations relative to the downstream Cy domain, with no contacts observed between the DD and the catalytic domain (Dowling et al., 2016). The Bamb_5915 crystal structure shows a further unique conformation of the βhD domain relative to its downstream domain. This suggests βhD domains are inherently mobile with respect to the tethered catalytic domain. However, the role of this mobility remains unclear. Sampling of multiple conformations may aid recruitment of a SLIM binding partner, with the flexible linker then allowing delivery of the carrier protein-bound substrate to the active site of the catalytic domain (Kosol et al., 2019).
Studies of the rhabdopeptide NRPS defined the first structure of a SLiM–\(\beta\)hD domain complex. Solution state NMR spectroscopy of a covalently tethered Kj12B-C \(\beta\)DD-\(\beta\)DD complex found that only the final five amino acids of the SLiM interact with the \(\beta\)hD domain (Hacker et al., 2018). These amino acids form a short \(\beta\)-strand lying antiparallel to, and interacting with, the \(\beta\)hD domain via two key salt bridges, E28/R96 and R24/E98 (Hacker et al., 2018) (Fig. 8a). Mutation of the charged \(\beta\)hD domain interfaces has also been demonstrated to undergo crosstalk in vitro. The Sven\_0512 PCP domain from the watasemycin NRPS, which has a SLiM appended to its C-terminus, has been demonstrated to interact productively with the Bamb\_5915 C domain via its \(\beta\)hD domain (Fig. 8e). This is despite the fact that the native partner \(\beta\)hD domain from the watasemycin system has only 35% identity to that of the corresponding domain from Bamb\_5915. Addition of excess Sven\_0512 SLiM peptide inhibits turnover, demonstrating the SLiM–\(\beta\)hD domain interaction plays an important role in product formation (Kosol et al., 2019). The inherent promiscuity of SLiM–\(\beta\)hD domain interfaces has the potential to be exploited for biosynthetic engineering. However, direct interactions between the carrier protein and the catalytic domain must also be considered to build efficient hybrid pathways.

Several engineering experiments involving \(\beta\)hD domains and associated interfaces have already been undertaken. Engineering of the rhabdopeptide producing NRPS to alter its product profile involved replacement of Kj12A/B and Kj12B/C with Xenorhabdus-derived SLiM–\(\beta\)hD domain pairs (XabA/B and XabB/C) from xenoamicin producing pathways in an attempt to prevent module iteration (Cai et al., 2019). Despite replacement with cognate \(\beta\)hD pairs, each successive modification was found to decrease the product yield (Cai et al., 2019). This may indicate the importance of maintaining the overall charge of the \(\beta\)hD-2 domain in the SLiM-\(\beta\)hD complex.

More recently DDs have been used to split the single trimodular xefoamidepeptide NRPS protein subunit into three separate proteins (Kegler & Bode, 2020) (Fig. 9). Two junctions were engineered: an E/C domain junction, where the SLiM–\(\beta\)hD domain pair from the TaxlllaA/B interface of the taxlllaNRPS was grafted, and a PCP/C domain junction where the PaxB/C DDs were tethered. Insertion of Pax DDs greatly decreased product titres, but use of the SLiM–\(\beta\)hD domain pair was found to increase product yield (Kegler & Bode, 2020). It has been acknowledged that protein solubility could contribute to these results. Regardless of whether this is the case, the use of a \(\beta\)hD domain at this cut site to engineer NRPS systems appears promising.

Work on the enacyloxin \(\beta\)hD domain led to identification of more than 1400 examples of SLiM–\(\beta\)hD domain pairs across various domain interfaces (Kosol et al., 2019). SLiMs were identified at the C-terminus of oxidases, heterocyclisation domains, PCP and ACP domains, while \(\beta\)hD domains were identified at the N-terminus of C, E, C, TE and MT domains, thioester reductases and halogenases, indicating the prevalence of this type of docking motif. Furthermore, additional examples of one SLiM peptide recruiting more than one downstream subunit to the same PCP domain, as in the rhabdopeptide NRPS, were identified (Kosol et al., 2019). This indicates the potential of SLiM–\(\beta\)hD domain pairs to be employed in genetic engineering of multiple types of system to produce analogues of polyketides, non-ribosomal peptides or hybrids thereof.

**Other Docking Tools**

Work in other areas of biochemistry and biotechnology has led to the development of novel protein docking tools that show
Fig. 8. Structural features, sequence alignments and use in biosynthetic engineering of \( \beta hD \) domains. (a) Solution state NMR structure of the docked complex formed by the covalently tethered Kj12B C-terminal SLiM and Kj12C N-terminal \( \beta hD \) domain (PDB accession code: 6EWV). Inset (top): Residues involved in salt bridge-forming interactions between the SLiM (\( \beta 3 \)) and \( \beta 2 \) strand of the \( \beta hD \) domain. Inset (bottom): Hydrophobic interface formed between \( \beta 2 \), \( \beta 3 \), \( \alpha 2 \) and \( \alpha 3 \). (b) Domain architecture of the rhabdopeptide-producing NRPS. Note that this iterative system contains three \( \beta hD \) domains and two SLiMs, which can all interact with varying affinities. (c) Sequence alignment of select DD SLiMs. Highlighted in red are SLiMs appended to the carrier protein domain of a PKS module, those in black are appended to the carrier protein domain of NRPS modules. TxlA, blue, contains a SLiM downstream of an E domain. (d) Sequence alignment of select N-terminal \( \beta hD \) domains. To the right of the alignment, the domain directly downstream of the \( \beta hD \) domain in each protein is shown. In (c) and (d), a schematic indicating the positions of the secondary structural elements from the solution state NMR structure is provided. Residue numbering in (a)–(d) is relative to that of PDB entry 6EWV. In (c) and (d), asterisks (*) denote the positions of the key interfacial residues highlighted in (a). (e) SLiM–\( \beta hD \) domain junction in the enacyloxin NRPS–PKS pathway. Crosstalk of the Sven_0512 PCP domain, a SLiM-bearing carrier protein domain from the watasemycin NRPS, with the enacyloxin Bamb_5915 \( \beta hD-C \) domain from the enacyloxin PKS-NRPS was able to produce \( N \)-acetyl (1S,3R,4S)-3-amino-4-hydroxycyclohexane-1-carboxylic acid. Domain abbreviations are as follows: Cy, heterocyclisation domain; C/E, dual condensation–epimerisation domain; H, flavin-dependent halogenase.

SYNZIPs

SYNZIPs are a synthetic biological tool developed via computational methods that enable engineering of protein–protein interactions (Grigoryan et al., 2009; Park et al., 2017; Thompson et al., 2012). The original development of this tool was inspired by...
basic-region leucine zippers (bZIPs); transcription factors with a high level of basic amino acids that allow binding of DNA (Grigoryan et al., 2009). There are around 53 bZIP proteins in humans. These are implicated in many biological processes and are therefore attractive targets for inhibition (Grigoryan et al., 2009). Work to develop protein inhibitors of bZIPs led to production of a set of synthetic bZIPs, or SYNZIPs; coiled–coil peptides that bind favourably to bZIPs, inhibiting their interaction with DNA (Grigoryan et al., 2009; Thompson et al., 2012). The coiled–coil SYNZIPs have a repeating seven residue, or heptad pattern (abcdefg), containing hydrophobic amino acids, usually at positions a and d, or e and g (Park et al., 2017; Thompson et al., 2012). On assessing the developed library of SYNZIPs for pairwise interactions amongst themselves, certain pairs were shown to bind with high affinity, forming coiled–coil interaction interfaces (Thompson et al., 2012).

Use of high affinity SYNZIP pairs to mediate interaction between PKS modules has been demonstrated. The DEBS PKS has been used as a platform to explore this (Klaus et al., 2019). SYNZIP pairs were able to promote interaction between non-cognate DEBS module 1 and module 6, enhancing turnover compared to use of DDs at this interface (Fig 10a and b). Furthermore, SYNZIPs enabled artificial splitting of DEBS module 1 between the AT and KR domains achieving turnover similar to that of covalently tethered module 1 (Fig. 10c and d) (Klaus et al., 2019). This shows the potential of SYNZIPs, and other computationally developed protein–protein interaction tools, to advance genetic engineering of multienzymes. With continued optimisation, SYNZIPs may prove to be a valuable tool for biosynthetic engineering.

**DNA- Templating**

A recently developed method of engineering protein–protein interactions in natural product pathways involves use of DNA-templating. It is worth noting that DNA-templating has been applied to chemical synthesis for several years, using the principle that single-stranded DNA (ssDNA) templates are tethered to building blocks which, as complementary ssDNA-tags anneal, allows control of the order of connectivity (Goodnow et al., 2017).

Recently, the DNA-templating methodology has been applied to the gramicidin S NRPS (Huang et al., 2020). Here, the multimodular GrsB was split into standalone modules, allowing controlled production of targeted shunt metabolites. A zinc finger (ZF) was tethered to each module. This recognises a specific 9 base pair DNA motif, providing affinity for a double-stranded DNA template which functions as a reaction surface to which the ZF-bearing NRPS modules are bound. A four-module system with three ZF-bound NRPS subunits was generated. A modified SLiM–βH domain pair from the InxA/B interface from the Xenorhabdus inexxi rhapsodopeptide/xenortide-like pathway enabled interaction between the ZF-bearing modules, while a non-cognate but interacting COM domain pair enabled interaction between the ZF-lacking GrsA and downstream TycB1 (Fig. 10e). This system was able to achieve turnover at one-third of the rate of wild-type GrsA-B (Huang et al., 2020). However, optimisation of this technique, as already demonstrated via modification of spacing between ZFs on the DNA-template, has the potential to further increase product titre. Replacement of GrsB1 with TycB1, a functionally analogous but more highly expressed protein, may have contributed to pathway inhibition. DNA-templating of NP pathways, while in its infancy, is highly
promising, especially as initial data shows it may be possible for it to be used in conjunction with chimeric systems to enhance turnover.

Conclusions and Future Perspectives

Since identification of the first PKS DD in 1996, termed a linker region rather than a DD at the time, there has been a dramatic increase in the number and type of DDs identified (Aparicio et al., 1996). Whilst some classes of DD have currently only been identified in a single type of multienzyme, bioinformatics analyses have begun to show that certain DD classes are prevalent across PKS, NRPS and hybrid PKS–NRPS systems. There are clearly several different types of DD. Some exist as CDD–NDD pairs, such as four α-helix bundles, eight α-helix bundles, double helix pairs and PAX DDs. In others, the CDD interacts directly with the downstream domain, such as helix-hand DDs and DHD domains, or a short peptide CDD interacts with a much larger, structured NDD, as observed in SLiM–βHD domain pairs. Perhaps additional mechanisms of interaction will begin to emerge as protein subunit junctions are studied further.

Experiments using DDs to engineer biosynthetic pathways have increased in success as knowledge of their compatibility and mechanisms of interaction has advanced. How malleable these systems are towards engineering, and the extent to which DDs can help to facilitate this, is not yet fully understood. In most cases, it is still not known what secondary points of contact occur across these junctions, both between pairs of catalytic domains and between DDs and catalytic domains. Once understood, those regions that should be kept consistent to enable conformational changes necessary for protein function will be clarified. This will allow higher product yields to be achieved. For simply bringing two proteins into proximity so as to enable substrate transfer, SYNZIPs are a valuable tool. However, the extent to which synthetic biology tools can be used in parallel to DDs to engineer chimeric systems is only just beginning to be understood.

Another factor to consider is subunit dimerisation, both in the context of NRPSs versus PKSs, but also in cis- versus trans-AT PKSs. The DDs in each type of system interact differently and this must be considered when using them for engineering. Substrate scope is also an important factor to consider in parallel to DD choice. It transpires that subtle changes are needed in the domains themselves to broaden their substrate scope and enable biosynthetic pathway engineering in its fullest sense. Indeed, recent mutagenesis studies on the active sites of KS domains from the DEBS cis-AT PKS to broaden their substrate scope showed that KS substate scope is a crucial factor to consider alongside maintenance of protein–protein interactions to achieve high turnover when undertaking PKS engineering (Klaus et al., 2020)—a phenomenon which has also been observed for KS domains from trans-AT PKSs (Jenner et al., 2013, 2015; Nguyen et al., 2008).

In conclusion, understanding of DDs, their structures, key amino acid interactions and the interfaces at which they occur has greatly increased over recent years. Use of DDs to engineer biosynthetic pathways has been undertaken with varying degrees of success. However, factors that must be maintained to successfully engineer a pathway are just beginning to be clearly defined.
Funding
This work was financially supported by the Biotechnology and Biological Sciences Research Council (BBSRC) (BB/M017982/1 to G.L.C., BB/L022761/1 to J.R.L., BB/RO10218/1 to G.L.C. and J.R.L.), Discovery Fellowship (BB/RO1212/1 to M.J.), a Midlands Integrative Bioscience Doctoral Training Partnership Studentship (BB/M01116X/1 to M.J.), the MRC (IBR Doctoral Training Partnership Grant (MR/N014294/1 Studentship to H.G.S.)), and the ERC (Starting Grant 639907 to J.R.L.).

Conflict of Interest
G.L.C. is a co-director of Erebagen Ltd. The other authors declare no conflict of interest.

References
Aparicio, J. F., Molnár, I., Schwecke, T., König, A., Haydock, S. F., Ee Khaw, L., Staunton, J., & Leadlay, P. F. (1996). Organization of the biosynthetic gene cluster for rapamycin in Streptomyces hygroscopicus: Analysis of the enzymatic domains in the modular polyketide synthase. Gene, 169(1), 9–16. https://doi.org/10.1016/1387-1877(95)00800-4.

Bretschnieder, T., Heim, J. B., Heine, D., Winkler, R., Busch, B., Kusebauch, B., Stehle, T., Zocher, G., & Hertweck, C. (2013). Vinylogous chain branching catalysed by a dedicated polyketide synthase module. Nature, 502(7469), 124–128. https://doi.org/10.1038/nature12588.

Broadhurst, R. W., Nietlispach, D., Wheatcroft, M. P., Leadlay, P. F., & Weissman, K. J. (2003). The structure of docking domains in modular polyketide synthases. Chemistry & Biology, 10(8), 723–731. https://doi.org/10.1016/s1074-5521(03)00156-x.

Buchholz, T. J., Geders, T. W., Bartley, F. E., Reynolds, K. A., Smith, J. L., & Sherman, D. H. (2009). Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. ACS Chemical Biology, 4(1), 41–52. https://doi.org/10.1021/cb8002607.

Cai, X., Zhao, L., & Bode, H. B. (2019). Reprogramming promiscuous nonribosomal peptide synthetases for production of specific peptides. Organic Letters, 21(7), 2116–2120. https://doi.org/10.1021/acs.orglett.9b00395.

Challis, G. L., Wagner, D. T., Meinke, J. L., Zogzas, C. E., Gay, G. R., Gay, D. C., Wagner, D. T., Meinke, J. L., Zogzas, C. E., Gay, G. R., & Weissman, K. J. (2003). The structure of docking domains in a polyketide synthase module. Nature, 426(6965), 468–476. https://doi.org/10.1038/nature02156.

Cai, X., Zhao, L., & Bode, H. B. (2019). Reprogramming promiscuous nonribosomal peptide synthetases for production of specific peptides. Organic Letters, 21(7), 2116–2120. https://doi.org/10.1021/acs.orglett.9b00395.

Challis, G. L., Wagner, D. T., Meinke, J. L., Zogzas, C. E., Gay, G. R., Gay, D. C., Wagner, D. T., Meinke, J. L., Zogzas, C. E., Gay, G. R., & Weissman, K. J. (2003). The structure of docking domains in a polyketide synthase module. Nature, 426(6965), 468–476. https://doi.org/10.1038/nature02156.

Cai, X., Zhao, L., & Bode, H. B. (2019). Reprogramming promiscuous nonribosomal peptide synthetases for production of specific peptides. Organic Letters, 21(7), 2116–2120. https://doi.org/10.1021/acs.orglett.9b00395.
Newman, D. J. & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. Journal of Natural Products, 83(3), 770–803. https://doi.org/10.1021/acs.jnatprod.9b01285.

Nguyen, T., Ishida, K., Jenke-Kodarna, H., Dittmann, E., Gurgiu, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., & Piel, J. (2008). Exploiting the mosaic structure of transacyltransferase polyketide synthases for natural product discovery and pathway dissection. Nature Biotechnology, 26(2), 225–233. https://doi.org/10.1038/nbt1379.

Park, W. M., Bedewy, M., Berggren, K. K., & Keating, A. E. (2017). Modular assembly of a protein nanotriangle using orthogonally interacting coiled coils. Scientific Reports, 7(1), 1–10. https://doi.org/10.1038/s41598-017-10918-6.

Partida-Martinez, L. P. & Hertweck, C. (2007). A gene cluster encoding rhizoxin biosynthesis in “Burkholderia rhizoxina”, the bacterial endosymbiont of the fungus Rhizopus microsporus. ChemBioChem, 8(1), 41–45. https://doi.org/10.1002/cbic.200600393.

Reimer, J. M., Haque, A. S., Tarry, M. J., & Schmeing, T. M. (2018). Piecing together nonribosomal peptide synthesis. Current Opinion in Structural Biology, 49, 104–113. https://doi.org/10.1016/j.sbi.2018.01.011.

Richter, C. D., Nietlispach, D., Broadhurst, R. W., & Weissman, K. J. (2008). Multienzyme docking in hybrid megasynthetases. Nature Chemical Biology, 4(1), 75–81. https://doi.org/10.1038/nchembio.2007.61.

Risser, F., Collin, S., Dos Santos-Moraiz, R., Gruen, A., Chagot, B., & Weissman, K. J. (2020). Towards improved understanding of inter-subunit interactions in modular polyketide biosynthesis: Docking in the enacyloxin IIa polyketide synthase. Journal of Structural Biology, 212(1), 107581. https://doi.org/10.1016/j.jsb.2020.107581.

Siewers, V., San-Bento, R., & Nielsen, J. (2010). Implementation of communication-mediating domains for non-ribosomal peptide production in Saccharomyces cerevisiae. Biotechnology and Bioengineering, 106(5), 841–844. https://doi.org/10.1002/bit.22739.

Smith, S. & Tsai, S.-C. (2007). The type I fatty acid and polyketide synthases: A tale of two megasynthases. Natural Product Reports, 24(5), 1041–1072. https://doi.org/10.1039/b603600g.

Stachelhaus, T., Mootz, H. D., Bergendah, V., & Marahiel, M. A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis: Catalytic role of the condensation domain. Journal of Biological Chemistry, 273(35), 22773–22781. https://doi.org/10.1074/jbc.273.35.22773.

Stachelhaus, T., Mootz, H. D., & Marahiel, M. A. (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. Chemistry & Biology, 6(8), 493–505. https://doi.org/10.1016/S1074-5521(98)80082-9.

Staunton, J. & Weissman, K. J. (2001). Polyketide biosynthesis: A millennium review. Natural Product Reports, 18(4), 380–416. https://doi.org/10.1039/a009079g.

Süssmuth, R. D. & Mainz, A. (2017). Nonribosomal peptide synthesis—Principles and prospects. Angewandte Chemie, International Edition, 56(14), 3770–3821. https://doi.org/10.1002/anie.201609079.

Tanovic, A., Samel, S. A., Essen, L. O., & Marahiel, M. A. (2008). Crystal structure of the termination module of a non-ribosomal peptide synthetase. Science, 321(5889), 659–663. https://doi.org/10.1126/science.1159850.

Thompson, K. E., Bashor, C. J., Lim, W. A., & Keating, A. E. (2012). SYNZIP protein interaction toolbox: In vitro and in vivo specifications of heterospecific coiled-coil interaction domains. ACS Synthetic Biology, 1(4), 118–129. https://doi.org/10.1021/sb200015u.

Tsujii, S. Y., David, E., Cane, D. E., & Khosla, C. (2001). Selective protein–protein interactions direct channeling of intermediates between polyketide synthase modules. Biochemistry, 40(8) 2326–2331. https://doi.org/10.1021/bi0102463N.

Watzel, J., Hacker, C., Duchardt-Ferner, E., Bode, H. B., & Wöhnhert, J. (2020). A new docking domain type in the peptidemicrobial-Xenorhabdus peptide producing nonribosomal peptide synthetase from Xenorhabdus hovienii. ACS Chemical
Smith et al., Biology, 15(4), 982–989. https://doi.org/10.1021/acschembio.9b01022.

Weissman, K. J. & Muller, R. (2008). Protein–protein interactions in multienzyme megasynthetases. ChemBioChem, 9(6), 826–848. https://doi.org/10.1002/cbic.200700751.

Whicher, J. R., Smaga, S. S., Hansen, D. A., Brown, W. C., Gerwick, W. H., Sherman, D. H., & Smith, J. L. (2013). Cyanobacterial Polyketide synthase docking domains: A tool for engineering natural product biosynthesis. Chemistry & Biology, 20(11), 1340–1351. https://doi.org/10.1016/J.CHEMBIOL.2013.09.015.

Yan, J., Gupta, S., Sherman, D. H., & Reynolds, K. A. (2009). Functional dissection of a multimodular polypeptide of the pikromycin polyketide synthase into monomodules by using a matched pair of heterologous docking domains. ChemBioChem, 10(9), 1537–1543. https://doi.org/10.1002/cbic.200900098.

Zeng, J., Wagner, D. T., Zhang, Z., Moretto, L., Addison, J. D., & Keatinge-Clay, A. T. (2016). Portability and structure of the four-helix bundle docking domains of trans-acyltransferase modular polyketide synthases. ACS Chemical Biology, 11(9), 2466–2474. https://doi.org/10.1021/acschembio.6b00345.