Towards improved understanding of intersubunit interactions in modular polyketide biosynthesis: Docking in the enacyloxin IIa polyketide synthase

Fanny Risser, Sabrina Collin, Raphael dos Santos-Morais, Arnaud Gruez, Benjamin Chagot, Kira J. Weissman

Towards improved understanding of intersubunit interactions in modular polyketide biosynthesis: docking in the enacyloxin IIa polyketide synthase

Fanny Risser, Sabrina Collin, Raphael Dos Santos-Morais, Arnaud Gruez*, Benjamin Chagot* and Kira J. Weissman*

Université de Lorraine, CNRS, IMoPA, F-54000 Nancy, France

*Corresponding authors.
E-mail address: arnaud.gruez@univ-lorraine.fr (A. Gruez), benjamin.chagot@univ-lorraine.fr (B. Chagot) or kira.weissman@univ-lorraine.fr (K.J. Weissman)

Present address, Fanny Risser: Department of Chemistry, University of Basel, Mattenstrasse 24a, CH-4058 Basel, Switzerland; Department of Biosystems Science and Engineering, ETH-Zürich, Mattenstrasse 26, CH-4058 Basel, Switzerland

Present address, Raphael Dos Santos-Morais: Université de Lorraine, LIBio, F-54000 Nancy, France

ABSTRACT

Modular polyketide synthases (PKSs) are molecular-scale assembly lines comprising multiple gigantic polypeptide subunits. Faithful ordering of the subunits is mediated by extreme C- and N-terminal regions called docking domains (DDs). Decrypting how specificity is achieved by these elements is important both for understanding PKS function and modifying it to generate useful polyketide analogues for biological evaluation. Here we report the biophysical and structural characterisation of all six PKS/PKS interfaces in the unusual, chimaeric *cis*-AT/*trans*-AT PKS pathway responsible for biosynthesis of the antibiotic enacyloxin IIa in *Burkholderia ambifaria*. Taken together with previous work, our data reveal that specificity is achieved in the enacyloxin PKS by deploying at least three functionally orthogonal classes of DDs. We also demonstrate for the first time that *cis*-AT PKS subunits incorporate DDs with intrinsically disordered character, reinforcing the utility of such regions for achieving both medium affinity and high specificity at PKS intersubunit junctions. Overall, this work substantially increases the number of orthogonal DDs available for creating novel, highly-specific interfaces within *cis*- and *trans*-AT PKSs and their hybrids. It also reveals unexpected sequence/structure relationships in PKS DDs, identifying major current limitations to both accurately predicting and categorising these useful protein-protein interaction motifs.

Abbreviations: PKS, polyketide synthase; AT, acyl transferase; DD, docking domain; 4HB, 4 α-helix bundle; DHD, dehydratase docking; short linear motif (SLiM); βHD, β-hairpin domain; ACP, acyl carrier protein; KS, ketosynthase; NRPS, nonribosomal peptide synthetase; C, condensation; PCP, peptidyl carrier protein; KR, ketoreductase; SEC-SLS, size exclusion chromatography-static light scattering

1. Introduction

Modular polyketide synthases (PKSs) are bacterial mega-enzymes responsible for synthesis of complex natural products, many of which have found clinical application (Davison and Brimble, 2019). A striking feature of these systems is assembly-line (division-of-labour) organisation, in which each cycle of chain building is carried out by a specific module of enzymatic domains (Staunton and Weissman, 2001; Weissman, 2009). To date, two classes of modular PKSs have been identified: the '*cis*-acyl transferase (AT)' and '*trans*-AT' systems, which appear to have distinct evolutionary origins (Jenke-Kodama et al., 2005; Jenke-Kodama and Dittmann, 2009; Nguyen et al., 2008). The defining difference between the two families is that in *trans*-AT PKSs, the eponymous ATs are not present

within the multienzymes, but are discrete enzymes which act iteratively in *trans* (Helfrich and Piel, 2016).

In both *cis*-AT and *trans*-AT PKSs, the multiple modules that constitute the systems are distributed among several gigantic polypeptides called 'subunits' (**Fig. 1**). As a result, the fidelity of the biosynthesis depends on contacts between the various subunits which allow them to adopt the correct sequence. These high-specificity interactions are mediated by short sequences situated at the extremities of the subunits called 'docking domains' (Broadhurst et al., 2003) (Throughout this article we will use the terminology ^{C}DD to refer to a C-terminal docking domain, and ^{N}DD to an N-terminal DD). Conversely, rational modification of DD-mediated communication by genetic engineering to enforce unnatural subunit ordering is a promising strategy towards generating novel, high-value polyketide analogues (Weissman, 2015). This approach, in which subunits/modules are employed as intact engineering units, has gained traction with the discovery that the functional domains within PKS modules are highly interdependent (Dutta et al., 2014; Whicher et al., 2014).

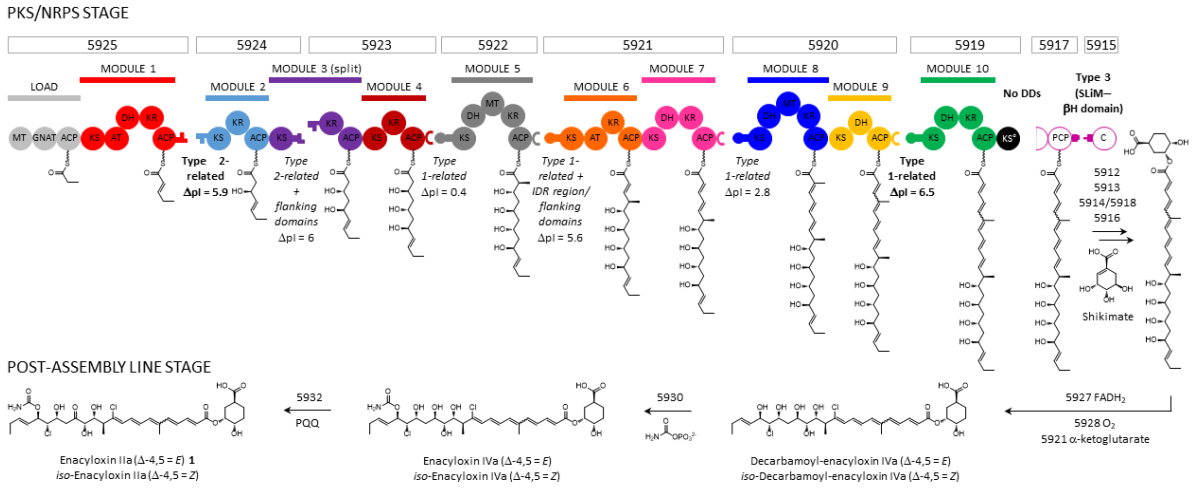


Fig. 1. The enacyloxin biosynthetic pathway. The biosynthesis occurs in two stages: assembly of the core by a hybrid polyketide synthase (PKS) (solid spheres)/nonribosomal peptide synthetase (NRPS) (hollow spheres), followed by elaboration of the released chain by a series of post-assembly line enzymes. The PKS portion of the pathway encompasses seven large subunits (Bamb_5925–5919, boxed), the first six of which belong to the *cis*-AT PKS class, and the last of which is a *trans*-AT PKS (Mahenthiralingam et al., 2011). Each subunit contains one or more functional modules (one loading and 10 chain extension), which are themselves composed of catalytic and acyl carrier protein (ACP) domains. As determined in this work, communication across the PKS intersubunit interfaces is mediated by a collection of docking domains belonging to at least two different types: 1-related and 2-related (DD types were assigned based on structure elucidation at high resolution (in bold), or detailed biophyiscal and bioinformatics analysis), while a type 3 DD (or SLiM–β-hairpin (Kosol et al., 2019)) was recently shown to operate at the PCP (Bamb_5927)/C (Bamb_5915) interface involved in chain off-loading. Key: MT, *C*-methyl transferase; GNAT, Gcn5-related *N*-acetyl transferase; KS, ketosynthase (the ° indicates a condensation-inactive version which serves as an ACP-transacylase); AT, acyl (malonyl) transferase; DH, dehydratase; KR, ketoreductase; PCP, peptidyl carrier protein; C, condensation. ΔpI: the difference in calculated pIs (Bjellqvist et al., 1994, 1993; Walker, 2005) for the identified ^{C}DDs and ^{N}DDs acting at each interface.

To date, efforts in several laboratories have identified multiple types of DDs operating in modular PKS systems, typically by solving the NMR or crystal structures of complexes in which the ^{C}DDs and

^{N}DDs are covalently joined. The *cis*-AT PKSs incorporate types 1a (Broadhurst et al., 2003), 1b (Buchholz et al., 2009), 2 (Whicher et al., 2013), and 3 (Hacker et al., 2018; Richter et al., 2008) DDs (**Fig. 2A**), and the *trans*-AT PKSs, 4 α-helix bundle (4HB) (Dorival et al., 2016; Zeng et al., 2016) and dehydratase docking (DHD) (Jenner et al., 2018) domains (**Fig. 2B**). For types 1a/1b, 2 and 4HB, formation of the docking domain complexes relies on contacts between α-helices (**Fig. 2**). However, there are important structural differences between the classes which can be used to discriminate between them (**Table 1** and **Fig. 2**). Specifically, the type 1 docking interactions involve a single α-helix contributed by both the ^{C}DD and the ^{N}DD, and the overall complex is heterotetrameric (Broadhurst et al., 2003; Buchholz et al., 2009). (The C-termini of structurally-characterized type 1a/1b docking domains include an additional four α-helix bundle dimerization motif upstream of the docking α-helix (**Fig. 2A**), but it is not thought to be involved in docking) (Broadhurst et al., 2003; Weissman, 2006). Type 2 complexes are also heterotetrameric, but based on the structure of the prototypical CurG ^{C}DD–CurH ^{N}DD complex (Whicher et al., 2013), both the ^{C}DD and ^{N}DD comprise two α-helices.

Finally, the 4HB type interactions are the unique heterodimers, and formed by two α-helices contributed by both the ^{C}DD and ^{N}DD (**Fig. 2B**) (Dorival et al., 2016; Zeng et al., 2016). In addition, the 4HB-type ^{N}DDs have been shown to exhibit characteristics of intrinsically-disordered regions (IDRs) (Dorival et al., 2016). As for the type 3 DDs, the mode of interaction is highly distinct (**Fig. 2A**). In this case, the ^{C}DD which is a short linear motif (SLiM) within an overall disordered C-terminus of the subunit (Kosol et al., 2019), inserts as a β-strand into a β-hairpin formed by the N-terminal docking domain, with additional contacts to an ^{N}DD α-helix (Hacker et al., 2018). Pairs of this type are thus also referred to as SLiM/βHD domains (Kosol et al., 2019). Importantly, the distinct architectures and modes of interaction within each type of DD mean that the various types are intrinsically orthogonal to each other.

To both understand how specificity is achieved and manipulate intersubunit interactions in a given modular PKS, it is necessary to identify the full range of DD types operating in the system. At present, however, sequence-based classification and pairwise interaction prediction can only be carried out with some confidence for type 1 domains (Wang et al., 2019). This is due to the fact that despite the growing number of high-resolution DD structures solved to date, there is only one or at most two representatives of each type. Furthermore, even DD members assigned to the same class can exhibit substantial architectural differences (as with the type 2 CurG ^{C}DD/CurH ^{N}DD and CurK ^{C}DD/CurL ^{N}DD pairs (Whicher et al., 2013)). Thus, it is currently impossible to establish the sequence/structure relationships for these elements necessary for reliable sequence-based

classification of all DD types, a situation compounded by the apparent high sequence variability within several proposed DD classes (Kosol et al., 2019; Wang et al., 2019; Whicher et al., 2013).

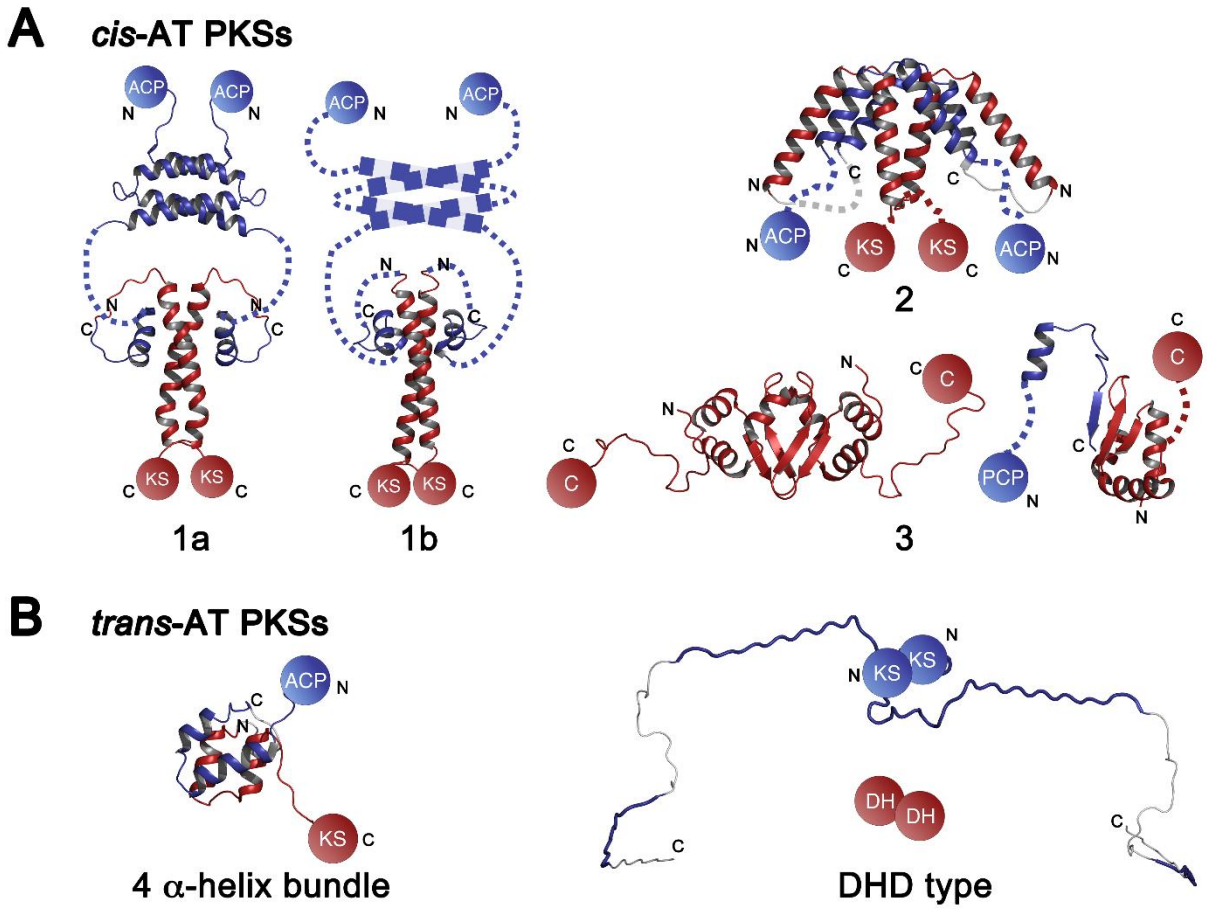


Fig. 2. Docking domains (DDs) in modular PKSs. (A) The four structural classes of DDs characterised to date from *cis*-AT PKSs (Broadhurst et al., 2003; Buchholz et al., 2009; Whicher et al., 2013), are shown in the presence of typical up- and downstream flanking domains. The C-terminal DDs (^{C}DDs) are illustrated in blue, and the ^{N}DDs in red. Class 3 is notably present in hybrid PKS/NRPS systems (where the NDD is homodimeric) (Richter et al., 2008), as well as pure NRPSs (where the ^{N}DD is monomeric) (Hacker et al., 2018). (B) Classes of DDs identified to date in *trans*-AT PKS systems. The 4 α-helix bundle type is formed by two α-helices contributed by each of the ^{C}DD and ^{N}DD (Dorival et al., 2016). The DHD domains (for 'dehydratase (DH) docking') (shown in thick blue lines) which are present at KS/DH junctions, are intrinsically disordered regions (IDRs) which directly contact the downstream DH domain (Jenner et al., 2018). The ^{N}DD of the four α-helix bundle type of docking domain (shown in red) is also an IDR. Key: ACP, acyl carrier protein; KS, ketosynthase; PCP, peptidyl carrier protein; C, condensation.

Here we aimed to boost our understanding of DD sequence/structure/specificity relationships by analysis of the enacyloxin IIa mixed *cis*-AT/*trans*-AT PKS from *Burkholderia ambifaria* (Mahenthiralingam et al., 2011). This system is notable for incorporating 8 interprotein interfaces, a relatively large number compared to other assembly lines (Helfrich and Piel, 2016; Medema et al., 2015). Six of these occur between PKS subunits (Bamb_5925–5919, hereafter referred to only by their numbers), five of which involve flanking acyl carrier protein (ACP) and ketosynthase (KS) domains. Such an organisation which would seem to impose particularly strict specificity demands on

the system. The remaining two interfaces include the nonribosomal peptide synthetase (NRPS) domains that participate in enacyloxin's unusual chain release mechanism: a peptidyl carrier protein (PCP, Bamb_5917) and a condensation domain (C, Bamb_5915)) (**Fig. 1**) (Kosol et al., 2019; Masschelein et al., 2019). In elegant recent work (Kosol et al., 2019), communication between 5917 and 5915 was shown to be mediated by a type 3 (SLiM/βHD domain) pair of DDs. In contrast, no docking domain appears to be required at the 5919/5917 junction (Kosol et al., 2019). A second feature of substantial interest is that the enacyloxin PKS is one of the rare known hybrid *cis*-AT/*trans*-AT PKSs (Kuo et al., 2016; Musiol et al., 2011). Studying the *cis*-AT/*trans*-AT interface (between subunits 5920 and 5919) was thus expected to identify the DDs capable of mediating communication between the two principal classes of modular PKSs.

Using a multidisciplinary approach combining biophysical analysis, NMR, small angle X-ray scattering (SAXS), and site-directed mutagenesis, we have characterised docking within the PKS portion of the enacyloxin system.

2. Results and Discussion

2.1. In silico *analysis of docking in the enacyloxin PKS*

The starting point for this work was to identify putative docking domains in the enacyloxin PKS using bioinformatics by analysing the sequences of the C- and N-termini of all PKS subunits within the system (here it is important to note that the subunits are numbered in descending order, as in the original publication (Mahenthiralingam et al., 2011)). In the case of subunits 5925 and 5923–5920, the most C-terminal domain is an ACP, while for subunits 5924 and 5922–5919, the N-terminal domain is a KS. The single 'split' module interface in the system (between subunits 5924 and 5923) is flanked by KS and ketoreductase (KR) domains. Thus, in order to identify regions at the extremities of each subunit that might contain docking domains, we established the boundaries of the adjacent functional domains. For this, the sequence of the last domain of each subunit was compared to the sequences of appropriate homologous domains whose structures have been solved at high resolution (**Supplementary Fig. S1A–C**). This analysis was straightforward for all subunits ending in ACP domains (5925, 5923–5920). On the other hand, close inspection of the C-terminus of 5924 revealed a large unassigned region downstream of the conserved KS domain (**Supplementary Fig. S1D**). Secondary structure prediction within this region (**Supplementary Fig. S1E**) revealed a series of putative secondary structure elements, followed by a 40+ amino acid gap before the last strongly-predicted α-helix. We therefore focused on this extreme C-terminal region as potentially containing a ^{C}DD.

On the basis of the length of the potential ^{N}DD-containing regions of subunits 5924, 5922–5919 (29–43 aa), the presence of additional SLiM/βHD type of DDs (avg. 50+ residues (Kosol et al., 2019; Richter et al., 2008)) could already be excluded. Indeed, none of these was flagged up in the recent, comprehensive database analysis for such DDs (Kosol et al., 2019). Although the 5923 N-terminal region is long enough to encompass a βHD domain, it shows poor homology to validated domains of this type (**Supplementary Fig. S1F**). In addition, the sequence of the extreme 5924 C-terminal region (LEDLQAE**, Supplementary Fig. S1G**) potentially containing a partner ^{C}DD does not correspond to consensus SLiM motifs (LxGxI (Hacker et al., 2018) and D/ExxEExL (Kosol et al., 2019; Richter et al., 2008)). To determine whether known classes could be predicted within the identified extremities, both the C- and N-terminal enacyloxin sequences were aligned against 12 putative members of each class (1a, 1b, 2 and 4HB) ('putative' refers to the fact that certain members of each class were only attributed previously based on sequence homology to a member whose structure had been solved). However, no confident assignments could be made solely on this basis (see **Supplementary Fig. S1G–J** for ^{C}DD alignments).

We therefore aimed to characterise all of the DDs directly, by obtaining them in recombinant form from *E. coli*. To enable successful heterologous expression, it was considered vital to identify reasonable boundaries for the functional DD regions. For this, a series of predictive software was used to analyse all of the putative DDs for: i) secondary structure (PSIPRED (Buchan and Jones, 2019)); ii) propensity for disorder/acting as an interaction motif (IUPred2A (Mészáros et al., 2018)); and in the case of the N-terminal DDs for which these motifs have been described, iii) coiled-coil regions (coiled-coil prediction, NPS@ web server) (Lupas et al., 1991) (**Supplementary Fig. S2**). Based on this analysis, two versions of all but the 5924 ^{C}DD were expressed: full-length and a shorter version based on the secondary structure analysis (**Tables S1** and **S2**). In the case of the ^{N}DDs, we uniformly expressed all of the sequences upstream of the first conserved catalytic domains. However, for 5920 ^{N}DD, a longer version was also obtained, as a new enacyloxin PKS sequence annotation was released during the course of this work, which suggested an alternative start site (**Tables S1** and **S2**).

2.2. Expression of the putative enacyloxin DDs

At the outset, eighteen constructs were amplified from *B. ambifaria* genomic DNA and cloned into vectors pBG102 or pLM302 for expression in *E. coli* BL21 (DE3) as His_6-SUMO-tagged or His_6-MBP tagged proteins, respectively. These comprised: 5925 ^{C}DD long and short versions, 5924 ^{N}DD, 5924 ^{C}DD, 5923 ^{N}DD, 5923 ^{C}DD long and short versions, 5922 ^{N}DD, 5922 ^{C}DD long and short versions, 5921

^{N}DD, 5921 ^{C}DD long and short versions, 5920 ^{N}DD long and short versions, 5920 ^{C}DD long and short versions, and 5919 ^{N}DD. Based upon initial results (*vide infra*), six additional constructs were then cloned: three covalent fusions of 5925 ^{C}DD and 5924 ^{N}DD containing a $(G_3S)_2$, $(G_3S)_3$ or a zero-length linker; a fusion of 5920 ^{C}DD and 5919 ^{N}DD via a $(G_3S)_2$ linker; 5923 ^{N}DD-KR; and 5922 ACP-^{C}DD. (The sequences of all constructs are given in **Table S2**.)

To allow both efficient monitoring of the DDs by UV during purification and dosing for binding studies by ITC, a tyrosine residue was added at the N-terminal ends of certain ^{C}DDs and ^{N}DDs (**Table S2**). Following purification by nickel chromatography, and cleavage of the affinity/solubility tags using human rhinovirus 3C (H3C) protease (**Supplementary Fig. S3**), the N-termini of all of the proteins also harboured short, non-native sequences (GPGS for all but two constructs, **Table S2**). The GPGS sequence is highly similar and occupies the same position relative to the ^{N}DDs as the $(G_3S)_n$ sequence used to fuse certain ^{N}DDs covalently to their upstream ^{C}DDs. We also show by NMR (*vide infra*), that the linkers have no effect on the docking interaction. In the case of the ^{C}DDs, the added residues lie upstream of the additional Y (**Table S2**). Unfortunately, despite extensive effort, it was not possible to obtain the following constructs as recombinant proteins from *E. coli*: 5924 ^{C}DD, 5923 ^{N}DD and 5920 ^{C}DD. These constructs were therefore sourced as synthetic peptides, along with a version of 5923 ^{N}DD missing the first 5 residues, to correspond with an alternative start site published during this work (**Table S2**).

2.3. Identification of the physiologically relevant version of each DD by ITC

As a first step in characterising the DDs, we aimed to determine which of the alternative ^{C}DD/^{N}DD versions was physiologically relevant. For this, binding studies were carried out by isothermal titration calorimetry (ITC). In several experiments, the measured binding stoichiometry was different depending on which partner was in the cell (e.g. for the 5925 ^{C}DD short/5924 ^{N}DD pair, N= 0.5 with 5924 ^{N}DD in the cell and N = 1.2 with 5925 ^{C}DD in the cell (**Supplementary Fig. S4A**, average of three measurements)). In the case of the 5925 ^{C}DD/5924 ^{N}DD interaction, the ca. 1:1 stoichiometry was later determined to be correct. The analysis that follows is therefore based, to the extent possible, on experiments performed with the ^{C}DDs in the cell (**Supplementary Fig. S4A**)). (We were ultimately able to explain the anomalous result of titrating ^{C}DDs as arising from their intrinsically disordered character (Sahu et al., 2016) (*vide infra*)).

Overall, a measurable interaction was detected for all pairs, except between 5924 ^{C}DD and 5923 ^{N}DD (the pair present at the exceptional split-module interface). In addition, the variants of the ^{C}DDs and ^{N}DDs gave clearly interpretable differences in binding affinity, allowing us to identify the physiologically relevant pairs (**Table 1**) and thus their functional extremities (**Table S2**). The measured

binding affinities for all of the interactions were in the range of 2–90 μM (**Table 1**, **Supplementary Fig. S4A**)), consistent with previous measurements between multiple types of DDs (Buchholz et al., 2009; Dorival et al., 2016; Gokhale et al., 1999; Whicher et al., 2013) and transient protein-protein interactions (Ozbabacan et al., 2011; Nooren and Thornton, 2003; Straight et al., 2007; Taniguchi et al., 2010). The binding stoichiometry for each pair was approximately 1 (**Table 1**).

In the case of the weakest affinity interaction (5922 ^{C}DD short (21 residues) with 5921 ^{N}DD, K_d = 90 μM), use of 5922 ^{C}DD long (70 residues) gave improved binding (K_d = 57 μM). Analysis of the additional ca. 50 residues in 5922 ^{C}DD long (**Supplementary Fig. S2**) suggests that they constitute an intrinsically-disordered region (IDR). Given that this affinity remained weak relative to previously measured DD interactions, we additionally evaluated the contribution of the upstream ACP to the junction (binding of 5922 ACP-^{C}DD to 5921 ^{N}DD). Although binding could not be saturated due to the tendency of 5921 ^{N}DD to aggregate and the poor availability of 5922 ACP-^{C}DD, the measured K_d of 19 μM (**Supplementary Fig. S4A**) showed that the presence of the ACP increased the interaction strength by at least three-fold. Thus, the flanking ACP appears to play a role in mediating the intersubunit interaction, as demonstrated previously for the erythromycin *cis*-AT PKS (Wu et al., 2002), but the mechanism by which this occurs remains to be determined.

To increase our confidence that the correct DD regions had been identified, the interactions were next evaluated for their specificity. For this, cross-talk was assessed between 5925 ^{C}DD and 5921 ^{C}DD and all ^{N}DDs of their non-cognate subunits by ITC (**Supplementary Fig. S4B**)). These experiments were carried out under the same conditions for which binding was detected between the native ^{C}DD/^{N}DD pairs. With one exception – that between the 5921 ^{C}DD and 5924 ^{N}DD which yielded a very weak affinity (K_d >100 μM; **Supplementary Fig. S4B**)) – no evidence for any non-native interaction was observed.

2.4. Analysis of DD oligomeric states

As the oligomeric state of the ^{N}DDs can be used to distinguish between 4HB-type DDs and the other types, we next determined whether the ^{N}DDs were monomeric or homodimeric, using a combination of circular dichroism (CD) and ITC (**Supplementary Figs. S5** and **S6A**). Here, evidence was sought for destructuration as the DDs were diluted, consistent with homodimer dissociation. For the dilution CD experiments, spectra were typically acquired of each ^{N}DD at 100 and 10 μM concentration, while dilution by ITC was evaluated by injecting the ^{N}DD alone into buffer. Certain ^{N}DDs were additionally analysed by size exclusion chromatography-static light scattering (SEC-SLS)

(**Supplementary Fig. S7**). The ^{C}DDs were also uniformly investigated by dilution ITC, in order to confirm their expected monomeric character (**Table 1**, **Supplementary Fig. S6B**).

This set of experiments revealed that all of the ^{N}DDs except 5921 ^{N}DD were homodimeric (**Table 1**), thus ruling out 4HB-type DDs at the majority of interfaces. With the notable exception of the highly-stable 5924 ^{N}DD homodimer, the ^{N}DDs showed evidence for dilution-induced dissociation by both CD and ITC (**Supplementary Figs. S5** and **S6A**). The physiological relevance of this finding is unclear, however, as its likely that ^{N}DD homodimerisation would be promoted by the presence of the downstream homodimeric KS domains in all subunits but 5923. In the case of 5923, by contrast, homodimerisation of 5923 ^{N}DD may be particularly important as its downstream domain is a presumably monomeric KR and not a KS, and the module lacks a dimerisation motif which is often present upstream of KR domains in *cis*-AT PKSs (Zheng et al., 2013). Indeed, by SEC-SLS, a construct incorporating both the 5923 ^{N}DD and the downstream KR (5923 ^{N}DD-KR) was shown to be homodimeric, as anticipated (**Supplementary Fig. S7**).

In contrast to the ^{N}DDs, the ^{C}DDs showed no evidence for dilution by ITC, consistent with their presumed monomeric character (**Table 1**, **Supplementary Fig. S6B**), a result confirmed by small-angle X-ray scattering (SAXS) (**Table S3**).

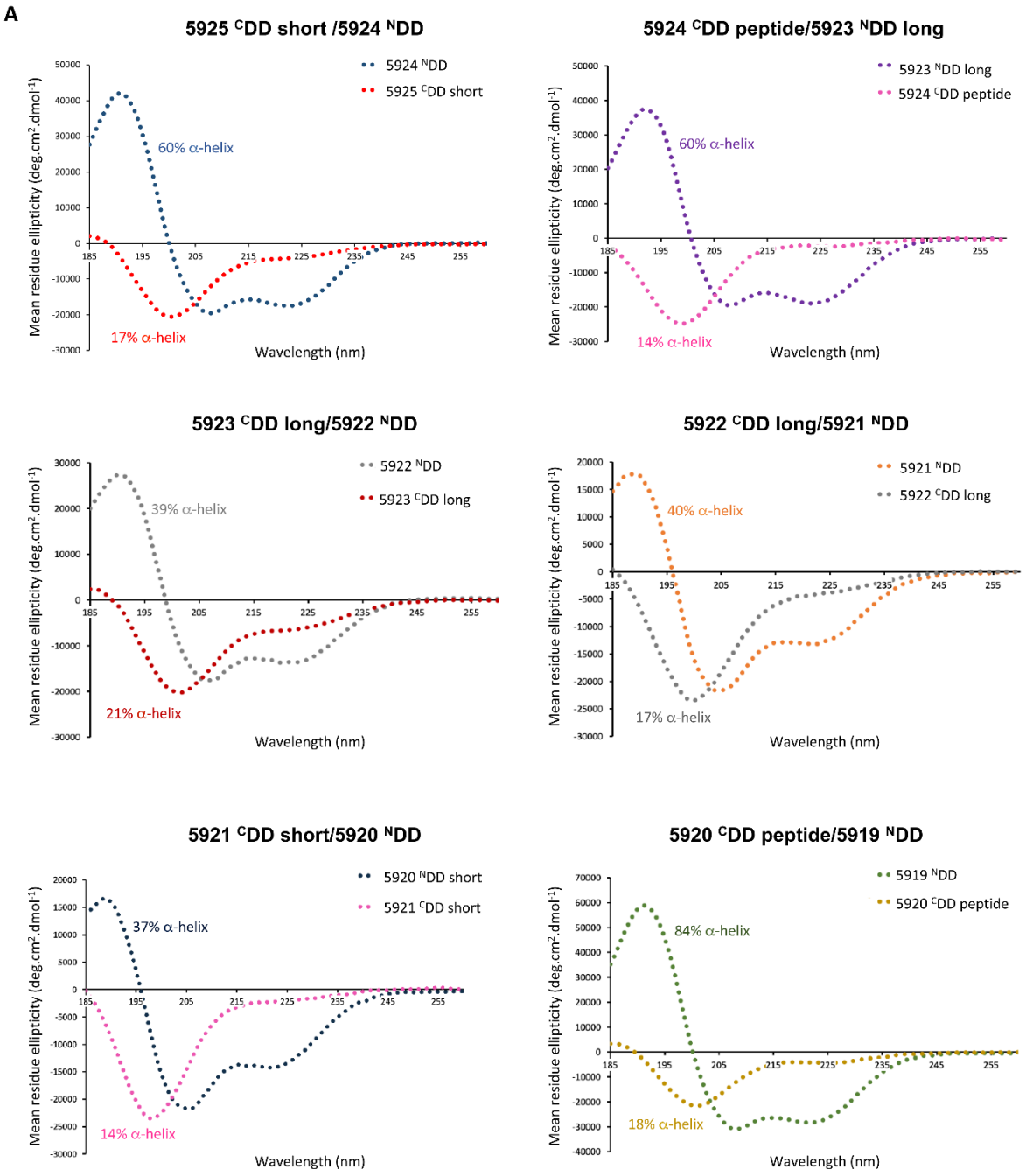
2.5. Analysis of independent and induced folding by CD

Within the solved structures of the covalent ^{C}DD–^{N}DD complexes, the interaction regions of types 1 and 2 ^{C}DDs and ^{N}DDs are well-folded and uniformly α-helical (Broadhurst et al., 2003; Buchholz et al., 2009; Whicher et al., 2013). We have recently shown, however, that isolated ^{N}DD of a representative 4HB-type domain from a *trans*-AT PKS is an intrinsically disordered region (IDR), which undergoes induced folding in the presence of its ^{C}DD partner (Dorival et al., 2016). This observation raises the question of whether other types of DDs incorporate IDRs, but that this behaviour might have been missed because the DDs were studied structurally only in the context of fused complexes with their partners, and thus were stably folded.

The acquired CD data on both the isolated ^{C}DDs and ^{N}DDs were thus analysed for evidence of independent folding (**Fig. 3A**, **Supplementary Fig. S5**). This analysis revealed that while the ^{N}DDs indeed exhibited substantial α-helical content, all of the ^{C}DDs behaved largely as random-coil regions. These data suggested that the ^{C}DDs of the enacyloxin system might be IDRs.

If the ^{C}DDs are indeed IDRs, they should undergo induced folding in the presence of their ^{N}DD partners, leading to an increase in the overall α-helical content of the docked complex relative to that of the individual domains (Dorival et al., 2016). To evaluate this hypothesis, we carried out

differential CD analysis, acquiring combined spectra of the isolated ^{C}DDs and ^{N}DDs, and comparing them to spectra of the ^{C}DD/^{N}DD mixtures (**Fig. 3B**), as described previously (Dorival et al., 2016). This analysis showed that structuration increased for all of the pairs, with the exception of 5924 ^{C}DD/5923 ^{N}DD. This one negative result is in accord with the fact that no interaction was detected for these domains by ITC (**Supplementary Fig. S4A**). As the ^{N}DDs of each pair are already partially-folded at the concentrations used in these experiments, we hypothesise that the majority of structuration observed when the domains are mixed is due to induced folding of the ^{C}DD IDRs.



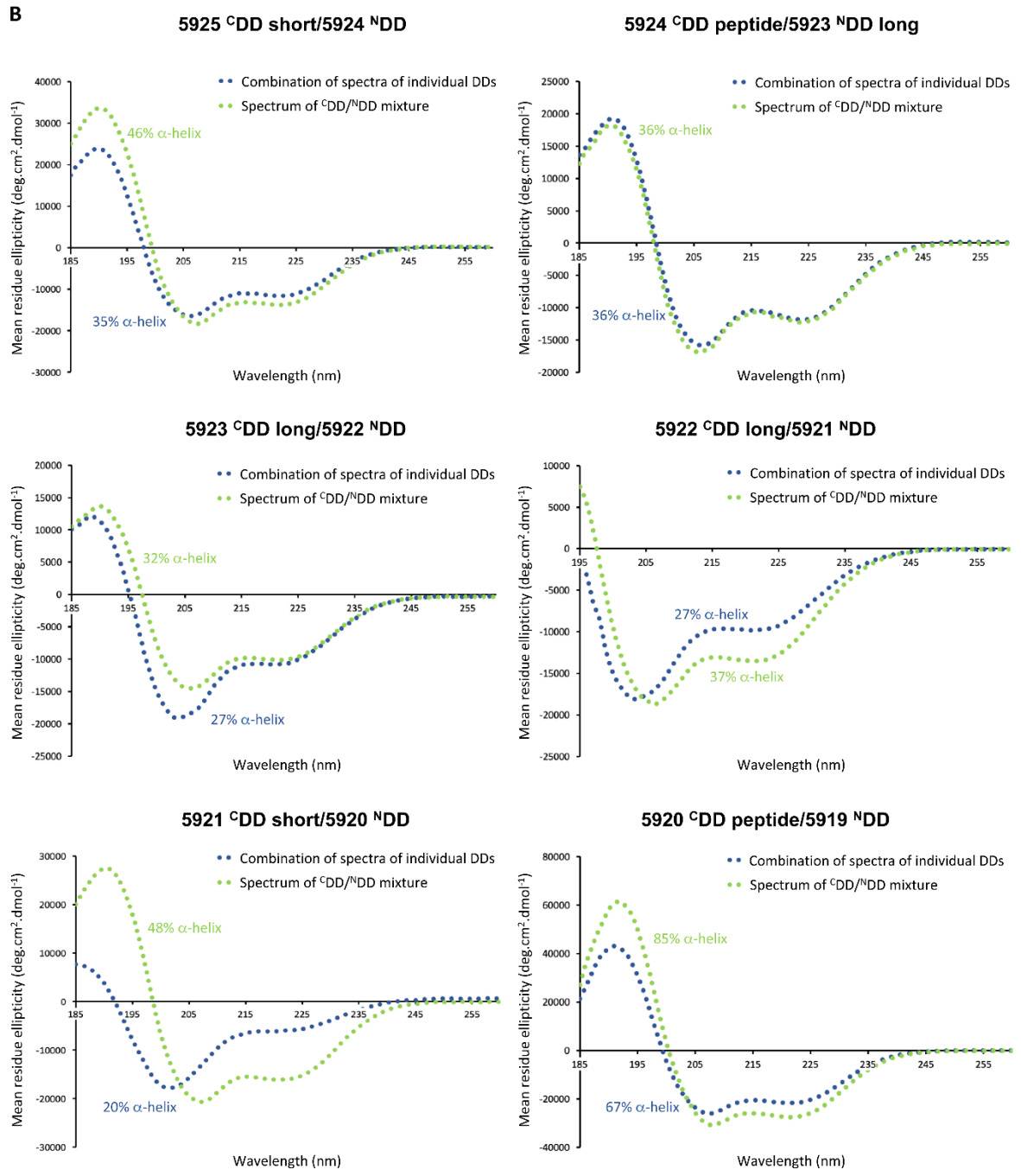


Fig. 3. Analysis of independent and induced folding of DDs by CD. (A) CD analysis of the discrete ^{C}DDs and ^{N}DDs from each interface (the investigated construct is indicated). The % α-helix calculated with the CDNN software (Böhm et al., 1992) is indicated. (B) Differential CD analysis of the DD pairs. In each case, a combined spectrum was obtained by simultaneously analysing the two isolated DDs (both at 100 μM). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μM, was used to identify increased structuration. The % α-helix calculated with the CDNN software (Böhm et al., 1992) under the two conditions is indicated.

Additional evidence for the IDR character of both the 5925 ^{C}DD and the 5920 ^{C}DD was then obtained by SAXS (**Supplementary Fig. S8**, **Table S3**). Specifically, the experimentally determined R_g values are in strong agreement with those predicted for denatured regions (Receveur-Bréchot and

Durand, 2012) (5925 ^{C}DD: R_g (calc'd) = 15.90 Å; R_g (expt.) = 15.63 Å; 5920 ^{C}DD: R_g (calc'd) = 11.58 Å; R_g (expt.) = 12.43 Å), while the Kratky plots and *ab initio* form calculations confirm the highly-extended nature of the ^{C}DDs (**Supplementary Fig. S8**).

While the 5920 ^{C}DD was correctly predicted to be disordered by IUPred2A (Mészáros et al., 2018), the disordered character of 5925 ^{C}DD was not anticipated, and indeed disorder was only predicted in three of the six of the ^{C}DDs (**Supplementary Fig. S2**) which all showed this behavior by CD. It may thus be necessary in future to employ multiple disorder prediction programs in order to obtain a more accurate overall prediction for DD regions.

2.6. Tentative assignment of DD types

We next combined the accumulated biophysical data on all of the interfaces (**Table 1**) with supervised realignment of the enacyloxin DD sequences, as this was now possible having identified the physiologically relevant regions (*vide infra*, as the alignments were subsequently refined with high-resolution structural information on two of the junctions). Realignment in this case was carried out using only the sequences of type 1 and type 2 DDs whose structures have been solved, and thus whose classification was certain. In this way, each DD pair was tentatively attributed to a class: type 2-related DDs at the 5925/5924 and 5924/5923 junctions, and type 1-related DDs at the remaining interfaces (**Table 1**).

Direct evidence was next sought for the DD attributions by characterising in detail the structures of a representative complex of each type of domain. For the putatively type 2-related DDs, the 5925/5924 interface was selected, and for the type 1-related DDs, the particularly interesting 5920/5919 *cis*-AT/*trans*-AT junction.

2.7. Characterisation of the 5925 ^{C}DD/5924 ^{N}DD interaction by NMR, SAXS and site-directed mutagenesis

To judge the feasibility of solving the structure of the 5925/5924 complex, we initially acquired ^{1}H-^{15}N HSQC NMR spectra of ^{15}N-labeled 5925 ^{C}DD in the absence/presence of ^{14}N 5924 ^{N}DD, and ^{15}N 5924 ^{N}DD in the absence/presence of ^{14}N 5925 ^{C}DD. Thus, in each case, all signals arising could be attributed to the ^{15}N-labeled partner (**Fig. 4AB**). These experiments notably confirmed that 5925 ^{C}DD is well-folded in the presence of its 5924 ^{N}DD partner and that the two form a stable complex. We thus went on to solve the structure of the complex by NMR, as well as that of the independently-folding 5924 ^{N}DD dimer. For this, structure calculation was performed using a two-step procedure consisting of in-

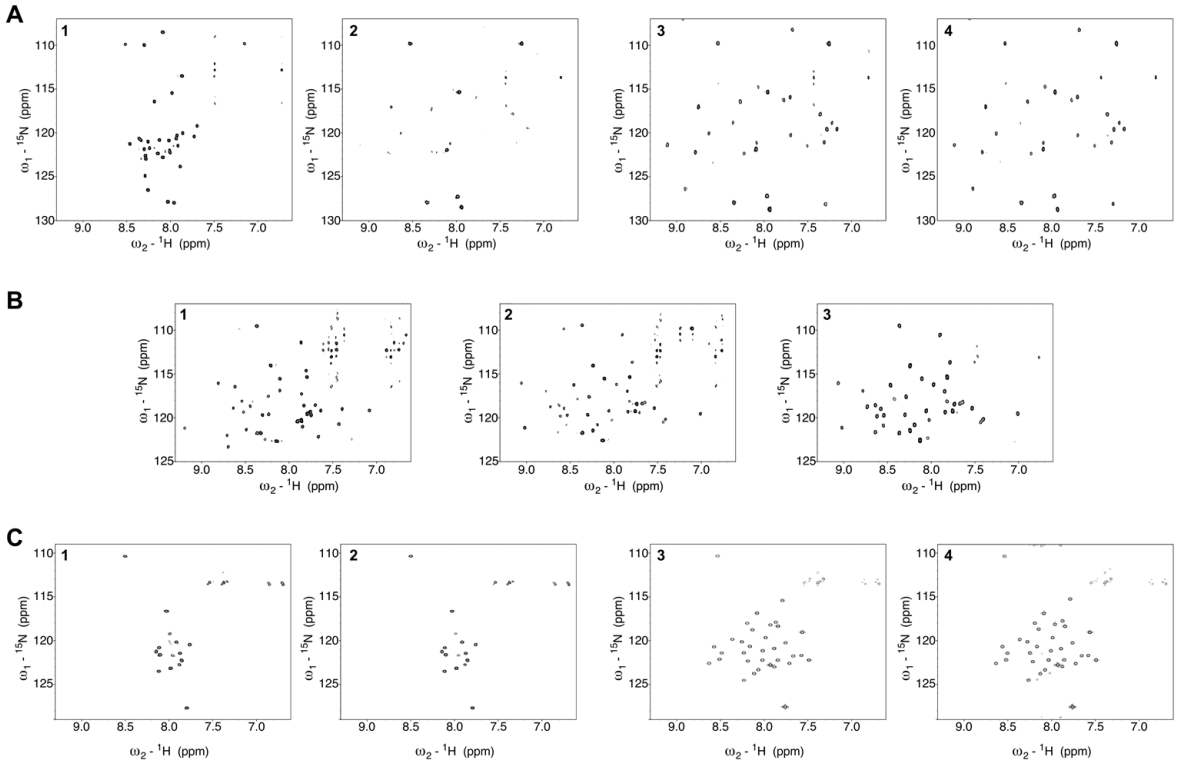


Fig. 4. Characterisation of the 5925 ^{C}DD/5924 ^{N}DD and 5920 ^{C}DD/5919 ^{N}DD interactions by NMR. (A) Saturation experiment with ^{15}N-labeled 5925 ^{C}DD in the presence of unlabelled 5924 ^{N}DD (concentration ratios (^{C}DD:^{N}DD): (1) 400 μM:0 μM; (2) 200 μM:240 μM; (3) 100 μM:400 μM; and, (4) 60 μM:600 μM). (B) Saturation experiment with ^{15}N-labeled 5924 ^{N}DD in the presence of unlabelled 5925 ^{C}DD (concentration ratios (^{N}DD:^{C}DD): (1) 400 μM:0 μM; (2) 300 μM:350 μM; and, (3) 76 μM:600 μM). (C) Saturation experiment with ^{15}N-labelled 5919 ^{N}DD in the presence of unlabelled 5920 ^{C}DD (concentration ratios (^{N}DD:^{C}DD): (1) 150 μM:0 μM; (2) 150 μM:300 μM; (3) 200 μM:1000 μM; (4) 200 μM:4000 μM).

itial structure generation using CYANA (Güntert, 2004), followed by restrained molecular dynamics refinement within AMBER (Case et al. 2014; Case et al., 2005). The quality of the structures was assessed using PROCHECK-NMR (Laskowski et al., 1996) and the MolProbity (Chen et al., 2010) server (**Table S4**), with additional confirmatory evidence obtained by SAXS analysis.

The solved structure of 5924 ^{N}DD (**Fig. 5A**, and for supporting SAXS analysis see **Supplementary Fig. S9AB**) shows that the ^{N}DD encompasses two consecutive α-helices, αA (W10–L18) and αB (K23–E39). αB from the two monomers forms a parallel coiled-coil, while each copy of αA contacts the αB of the opposite monomer, to form an overall 4 α-helix bundle. The first 13 residues of the construct which include four non-native residues (**Table S2**), are unstructured. Dimer formation is ensured by predominantly hydrophobic interactions between the second α-helix of one monomer and both α-helices of the other monomer. Hydrophobic residues involved in the packing are W10, Y14, L15 and L18 located in αA, L26, L29 and L33 from αB of the same monomer, and L26, M27, A30, L31, K34 and L38 within αB of the other monomer (**Fig. 5B**). Residue K23 located at the beginning of

αB of each monomer interacts with the carbonyl oxygen of S19 and L21 located within the loop between the α-helices of the other monomer. K11 of one monomer lies in the vicinity of Q35 of the other monomer, and may participate in a polar interaction.

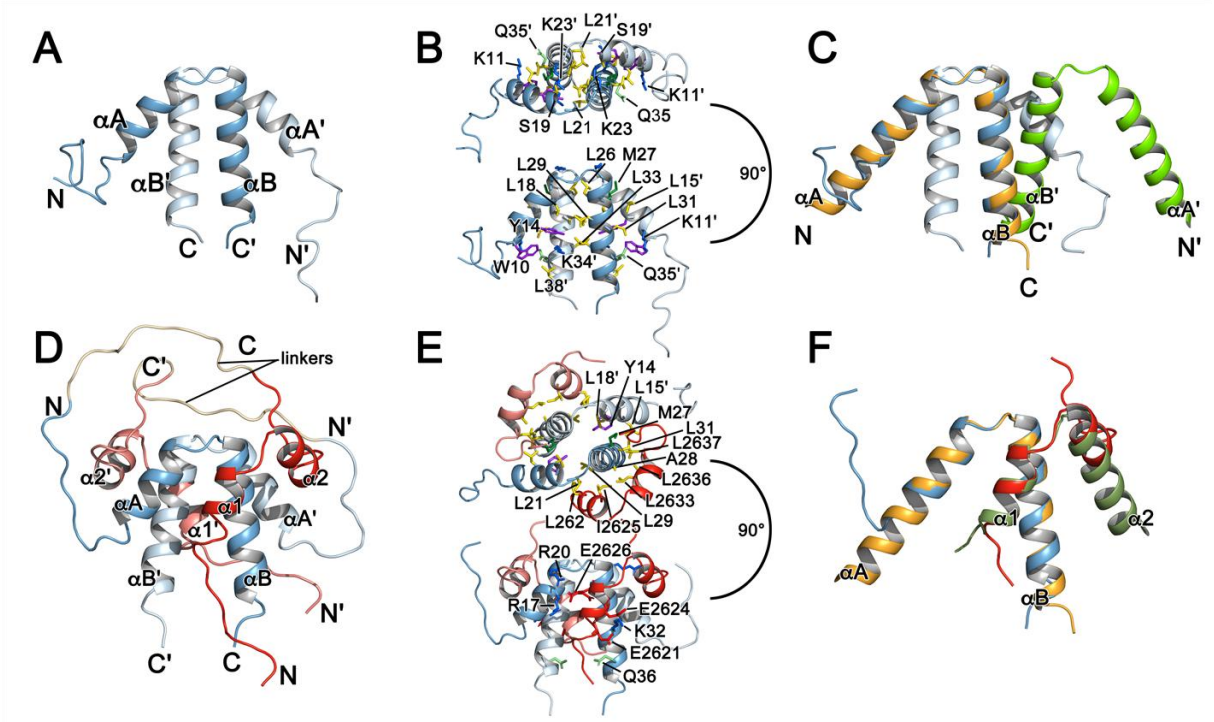


Fig. 5. NMR structures of 5924 ^{N}DD and the 5925 ^{C}DD–5924 ^{N}DD complex. (A) Cartoon representation of the 5924 ^{N}DD dimer, with one monomer coloured in blue and the other in light blue (residues indicated with primes). (B) Interactions between the protein chains. Top and side views of the residues (represented as sticks) involved in dimer formation. Basic residues are shown in blue, polar in light green, aromatic in purple and aliphatic in yellow. (C) Comparison of dimer topology between CurH ^{N}DD and 5924 ^{N}DD. The 5924 ^{N}DD dimer is represented as in (A). The CurH ^{N}DD chains are coloured in orange and green. To emphasise the overall difference in architecture, one monomer of CurH ^{N}DD is superimposed on one monomer of 5924 ^{N}DD. (D) Cartoon representation of the 5925 ^{C}DD–5924 ^{N}DD dimer. 5925 ^{C}DD is shown in red (dark and light for the two monomers), 5924 ^{N}DD in blue, and the $(G_3S)_3$ linker in beige. The α-helices present in 5925 ^{C}DD are indicated as α1 and α2, and those in 5924 ^{N}DD as αA and αB. The primes indicate the second monomer of each DD. (E) Interactions between protein chains in the 5925 ^{C}DD–5924 ^{N}DD dimer. Residues involved in hydrophobic interactions (top view) and in polar interactions (bottom view) are shown as sticks. Basic residues are coloured in blue, polar in light green, methionine in dark green, aromatic in purple and aliphatic in yellow. (F) Comparison of the mode of ^{C}DD/^{N}DD interaction between CurG ^{C}DD–CurH ^{N}DD and 5925 ^{C}DD–5924 ^{N}DD. The figure shows a monomer of CurH ^{N}DD (orange) interacting with a monomer of CurG ^{C}DD (green), superimposed on monomers of 5925 ^{C}DD and 5924 ^{N}DD, as shown in (D). For clarity of presentation, certain linkers in the structures have been removed.

Overall, the monomer structure of 5924 ^{N}DD is similar to that of the prototypical class 2 ^{N}DDs from the curacin PKS (Whicher et al., 2013) (RMSD of 0.55 Å between W10–Q36 of 5924 ^{N}DD and Q48–T74 of CurH (PDB code 4MYY)). However, the mode of dimerisation differs substantially: the CurH ^{N}DD monomers self-associate purely via formation of a coiled-coil, whereas in the 5924 ^{N}DD structure, helices αA also contact the coiled-coil formed by αB to yield an interleaved 4-α-helix bundle (**Fig. 5C**). Considering the coiled-coil regions, a possible explanation for this difference is a

change in register between 5924 ^{N}DD and CurH ^{N}DD in terms of the start of αB and the first residue to contribute to coiled-coil formation. In 5924 ^{N}DD, αB begins at K27 and the coiled-coil starts at L30, while in CurH ^{N}DD, αB starts at P61 and the coiled-coil at L62. The overall result is that distinct faces of the α-helices in 5924 ^{N}DD and CurH ^{N}DD are used to form the coiled-coil motifs, which in the case of 5924 ^{N}DD, also permit interaction with αA. In addition, in 5924 ^{N}DD αA, the presence of three hydrophobic residues W10, Y14 and L15, which align structurally with the hydrophilic residues Q48, E52 and K53 within CurH αA, forces the formation of a more compact dimer to allow burial of these residues in the hydrophobic core of the complex (**Fig. 5B**). Moreover, the polar interaction occurring between K23 of one 5924 ^{N}DD and S19/L21 of the other is not possible within CurH, as the equivalent residue to K23 is non polar (P61).

In order to stably capture the docked complex, three constructs were investigated in which the 5925 ^{C}DD was covalently fused to the 5924 ^{N}DD: i) directly (via a zero-length linker), ii) using a $(G_3S)_2$ linker, as previously described (Dorival et al., 2016; Whicher et al., 2013), or, iii) with a $(G_3S)_3$ linker (**Table S2**). The last construct yielded a stably-docked DD complex. Furthermore, comparative analysis of a mixture of the two independent DDs and the $(G_3S)_3$ covalent complex by CD showed essentially no difference, and thus the artificial fusion had not fundamentally altered the interaction (**Supplementary Fig. S9C**). Full structure elucidation by NMR as described above revealed an overall topology again close to type 2 docking interactions, as predicted (**Fig. 5D** and **Table S4**). Within the complex, the structure of the 5924 ^{N}DD is equivalent to that of the isolated domain, while 5925 ^{C}DD also adopts two α-helices (α1 E2621–G2627 and α2 D2630–L2637). The second α-helix binds in a perpendicular orientation to an αB from the coiled-coil dimer, while helix α1 contacts both αA and αB, to form an overall 8 α-helix bundle (Whicher et al., 2013). We further confirmed the overall shape of the complex and its molecular weight by SAXS (**Table S3**, and see **Supplementary Fig. S9D–G** for a detailed analysis of these results).

The docking interface is hydrophobic and comprises residues L2622 and I2625 from α1 of 5925 ^{C}DD interacting with Y14, L18 from 5924 ^{N}DD αA, and L21, A28 and L29 from 5924 ^{N}DD αB, as well as L2633, L2636 and L2637 from α2 of 5925 ^{C}DD interacting with M27, A28 and L31 of 5924 ^{N}DD αB and L15 of αA of the other monomer (**Fig. 5E**). The structure additionally identified multiple polar interactions potentially contributing to docking specificity (5925 ^{C}DD/5924 ^{N}DD): E2626/R17 and/or R20, and E2621 and/or E2624/K32. More specifically, for these interactions, the NMR structure suggested that a charged residue of one partner was sandwiched between two oppositely-charged residues of the other partner (**Fig. 5E**). As mentioned earlier, consistent with the important role of such contacts in docking, the overall calculated pI (Bjellqvist et al., 1994, 1993; Walker, 2005) for the ^{C}DD shows it to be acidic (pI = 3.8) and the ^{N}DD to be basic (pI = 9.7) (**Table 1**).

In order to evaluate the roles of the contribution of these interactions to the DD complex, we targeted the ^{N}DD residues. The ^{C}DD IDR was not mutated, as we anticipated that we would not be able to distinguish mutations which impacted the docking interaction from those which influenced the induced structuration of the ^{C}DD, or both (Shammas et al., 2016). Based on analysis of the NMR structure, the ^{N}DD residues R17 and K32 were mutated to alanine and glutamate. However, of the two K32 mutants, only the K32A construct proved soluble. Far-UV CD analysis of the obtained mutants confirmed that their structures were not perturbed significantly relative to wild type 5924 ^{N}DD (**Supplementary Fig. S9H**), while SAXS (**Supplementary Fig. S9I–L**) or SEC-SLS analysis (**Supplementary Fig. S7**) (of all constructs except R17E) demonstrated that they remained stably homodimeric.

Following these controls, the binding of each 5924 ^{N}DD mutant to wild type 5925 ^{C}DD was investigated by ITC (experiments in triplicate), as well as the ability of the mutant to induce structuration of the 5925 ^{C}DD by CD. None of the mutants produced a substantial decrease in the measured K_ds for the interaction or a change in binding stoichiometry (**Supplementary Fig. S9M–O**) (R17A: 2.8 $\pm$ 0.6 μM, N = 1.3; R17E: 5 $\pm$ 1 μM, N = 1.2; K32A: 6.3 $\pm$ 0.5 μM, N = 1.2). In addition, all of the ^{N}DD mutants still provoked proper structuration of 5925 ^{C}DD, as measured by CD (**Supplementary Fig. S9P–R**). These data support the idea that the docking interaction cannot be dissected into individual pairs of contacting residues, but rather arises from an ensemble of contacts between the surfaces (*e.g.* tripartite rather than bipartite contacts) (Longhi et al., 2003). Such a feature would appear to make it intrinsically robust to perturbation by single-site mutagenesis.

Having shown that the coiled-coil and interleaved 4-helix bundle mode-of-dimerisation of 5924 ^{N}DD lead to divergent α-helical orientations relative to the canonical type 2 CurH ^{N}DD, we lastly aimed to evaluate the effects of these modifications on the docking interactions in the 5925 ^{C}DD–5924 ^{N}DD structure relative to those in the CurG ^{C}DD–CurH ^{N}DD complex (Whicher et al., 2013). For this, the RMSD between the backbones of various portions of the structures was determined, by specifically including or excluding helix α2 of 5925 ^{C}DD. Comparing residues T2620–L2637 of 5925 ^{C}DD encompassing both helices α1 and α2 and K11–L38 of 5924 ^{N}DD, to those of K4–I21 of CurG ^{C}DD and I49–L76 of CurH ^{N}DD, gave a backbone RMSD of 0.99 Å. The same calculation with only the residues encompassing α1 of 5925 ^{C}DD (i.e. 5925 ^{C}DD T2620–L2628 and 5924 ^{N}DD K11–38 vs. K4–S13 of CurG ^{C}DD and I49–L76 of CurH ^{N}DD) yielded an improved RMSD of 0.60 Å. Thus, overall, the topology of the ^{C}DD α1 relative to the ^{N}DD is identical for the two complexes, while that of ^{C}DD α2 differs substantially (**Fig. 5F**). Thus overall, our data reveal that plasticity in the interactions leading to coiled-coil mediated homodimerisation in 5924 ^{N}DD translate into an overall, novel architecture of the 5925 ^{C}DD–5924 ^{N}DD complex relative to that of CurG ^{C}DD–CurH ^{N}DD. An additional consequence

is to alter the surface properties of the ^{N}DD necessary to induce proper folding of the ^{C}DD partner – potentially explaining the high interaction specificity observed for DD pairs within the 'type-2' related family.

2.8. Characterisation of the 5920 ^{C}DD/5919 ^{N}DD interaction by NMR and SAXS

For the 5920 ^{C}DD/5919 ^{N}DD interaction, ^{15}N-labeled 5919 ^{N}DD was initially investigated in the absence and presence of increasing amounts of unlabelled 5920 ^{C}DD by NMR. Analysis of 5919 ^{N}DD on its own revealed a probable monomer/homodimer equilibrium on the intermediate time-scale (**Fig. 4C**), consistent with relative instability of the association determined by CD and ITC (**Supplementary Figs. S5 and 6A**). This experiment also demonstrated that the complex could be saturated with a large excess of 5920 ^{C}DD (**Fig. 4C**), motivating us to solve the structure of the docking complex. For this, a construct was created in which the two DDs were fused using a $(G_3S)_2$ linker (Whicher et al., 2013), which yielded a stable protein. Analysis by SAXS confirmed that, as expected based on the behaviour of 5919 ^{N}DD, the fusion construct was dimeric (**Supplementary Fig. S10AB**).

Full structure elucidation of the complex was then carried out by heteronuclear multi-dimensional NMR (**Fig. 6A** and **Table S4**), as described earlier. Within the docking complex, residues Q2–A28 of 5919 ^{N}DD form a coiled-coil, via hydrophobic contacts involving I4, L8, I15, L18 and L25 of each monomer. The bound form of 5920 ^{C}DD is structured, with residues I2961–E2971 forming a single α-helix.

The docking interface comprises a series of hydrophobic interactions: I2961, L2964, L2968 and M2972 of 5920 ^{C}DD contact I4, L8, A9 and L12 of one 5919 ^{N}DD monomer and L7 of the other (**Fig. 6B**). In addition, polar interactions are observed: E2971 of one 5920 ^{C}DD monomer forms salt bridges with K16 and K19 of one 5919 ^{N}DD monomer, while D2974 of the same 5920 ^{C}DD monomer potentially contacts K10 of the second 5919 ^{N}DD monomer (**Fig. 6C**). Based on this analysis, while the 5920 ^{C}DD–5919 ^{N}DD complex is clearly type 1, as anticipated, it is not possible to conclusively designate it as being 1a or 1b. For example, despite the presence of a single docking α-helix (type 1a-like), it is of intermediate length between those described for 1a and 1b structures (**Fig. 1**, **Table 1**) (Broadhurst et al., 2003; Buchholz et al., 2009). Furthermore, certain polar contacts at the interface are consistent with both type 1a and type 1b interactions (**Supplementary Fig. S10C**), and therefore cannot clearly be sub-typed. The interface is thus 1-related. Furthermore, in contrast to the DEBS and Pik ^{C}DDs which incorporate three successive α-helices (**Fig. 1**) (Broadhurst et al., 2003; Buchholz et

al., 2009), the 5920 ^{C}DD only contains a single α-helix. Thus, the 5920 ^{C}DD lacks the classic subunit homodimerisation motif of type 1 DDs.

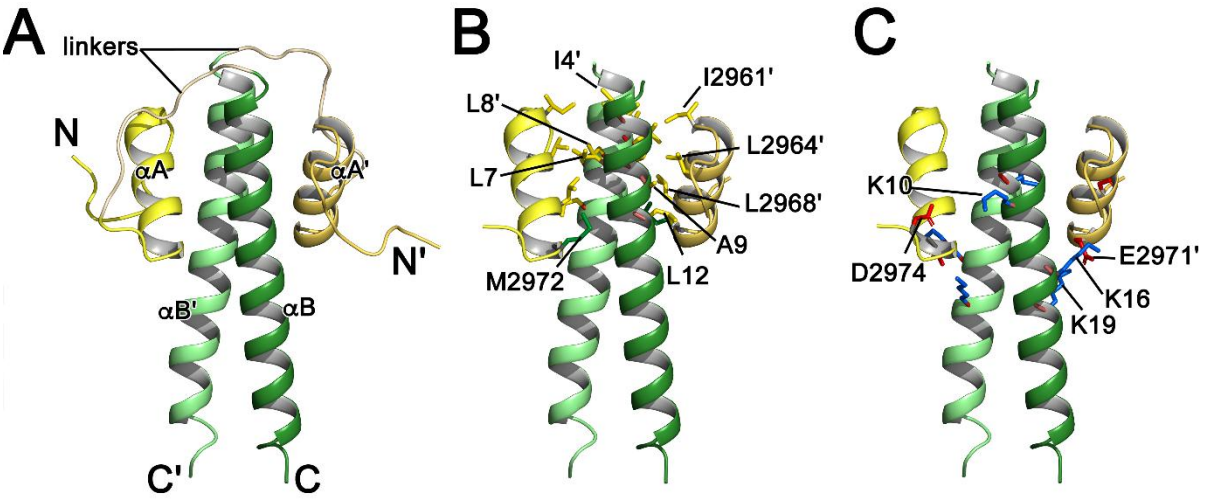


Fig. 4. NMR structure of the 5920 ^{C}DD–5919 ^{N}DD complex. (A) Cartoon representation of the 5920 ^{C}DD–5919 ^{N}DD fusion construct, with 5920 ^{C}DD shown in yellow and 5919 ^{N}DD in green. (B) Hydrophobic interactions between 5920 ^{C}DD and 5919 ^{N}DD. Residues are shown as sticks. (C) Polar interactions between 5920 ^{C}DD and 5919 ^{N}DD. Residues are shown as sticks. In (B) and (C), aliphatic residues are coloured in yellow, methionine in green, basic in blue and acidic in red. For clarity of presentation, linkers in panels (B) and (C) have been removed.

These data demonstrate that type 1-related DDs, which to date were known only from *cis*-AT PKSs, can mediate interaction between an upstream *cis*-AT PKS and a downstream *trans*-AT PKS subunit. Furthermore, sequence analysis of the DDs at a second hybrid interface (the *trans*-AT/*cis*-AT junction between KirAV and KirAVI (Weber et al., 2008)), suggests they are also type 1 (**Supplementary Fig. S11**). It is also notable that several more of the putatively 1-related domains from the enacyloxin system also comprise only the third α-helix of classical type 1a (Broadhurst et al., 2003) and 1b (Buchholz et al., 2009) domains (**Table 1**), arguing that this is a viable PKS interaction element. Thus, this type of DD should be a generally suitable element for creating additional such hybrid systems as a means of boosting polyketide diversity.

2.9. The basis for docking specificity in the enacyloxin PKS

As a final step in our analysis, we refined our DD sequence alignments to include the structural information obtained on the 5920 ^{C}DD/5919 ^{N}DD and 5925 ^{C}DD/5924 ^{N}DD interfaces (**Supplementary Fig. S12**). Incorporation of these data fully supported our preliminary attributions for the remaining enacyloxin junctions (**Table 1**). With the structural data in hand, we were also able to evaluate the quality of the secondary structure predictions for the two interfaces (**Supplementary Fig. S13**). This analysis showed that although the prediction was highly accurate for the type-1 related 5920

^{C}DD/5919 ^{N}DD junction, it was weaker for the type 2-related 5925 ^{C}DD/5924 ^{N}DD interface, both in terms of the number and the boundaries of the α-helical elements. These observations further support the necessity of direct experimental characterisation of PKS DDs.

Taken together with previous work (Kosol et al., 2019), our data show that the enacyloxin PKS incorporates at minimum 3 different types of DDs (1, 2 and 3 (Kosol et al., 2019)). Although multiple interfaces apparently employ the same types of DDs, features can be distinguished that might underlie docking specificity. Concerning the four type 1-related junctions, the most notable difference is the ΔpIs among the DD pairs (where ΔpI = theoretical pI ^{N}DD−theoretical pI ^{C}DD (Bjellqvist et al., 1994, 1993; Walker, 2005)): 5923/5922 = 0.4, 5922/5921 = 5.6, 5921/5920 = 2.8 and 5920/5919 = 6.5. This observation suggests that divergent electrostatics may be a major contributor to partner selection. In the case of the 5922/5921 and 5920/5919 interfaces for which the ΔpIs are similar, a further distinguishing aspect is that the 5922/5921 junction also depends on a putative, extended IDR domain within the C-terminal region of 5922 (directly upstream of the identified ^{C}DD), as well as potential contacts between the flanking ACP and KS domains.

For the two interfaces mediated by type 2-related domains, only those at the 5925/5924 interface interact as discrete domains (K_d = 1.1 ± 0.1 μM), while no binding is observed between 5924 ^{C}DD and 5923 ^{N}DD in the absence of their neighbours. Several explanations for this latter result are possible, including that the DDs only function when linked to their respective up- and downstream domains, that they do not contact each other directly but instead interact with the flanking domains (as with the dehydratase docking domains (DHDs) described for *trans*-AT PKSs (Jenner et al., 2018)), or indeed that DDs do not contribute to reconstituting a functional module at this point in the assembly line.

2.10. Implications of our data for structure/sequence relationships, and the classification and prediction of docking domains

The data obtained in this work show that sequence variation within DD types (Wang et al., 2019) can manifest itself in different ways. In the case of the type 2-related 5925/5924 DDs, the general character of the interaction is preserved (i.e. formation of an 8-α-helix bundle), but the overall architecture of the complex differs. Discrepancies include the manner of ^{N}DD homodimerisation and the length and orientation of α-helical secondary structure motifs, and thus the specific residues constituting the contact interfaces. In contrast, the type 1-related DDs at the 5920/5919 junction, whose sequences also diverge substantially from canonical type 1 DDs, display a fold highly-

reminiscent of those described previously but the pattern of charge:charge interactions is distinct (Broadhurst et al., 2003; Buchholz et al., 2009).

An important question raised by these results is to what extent such sequence/architectural divergences give rise to sub-types of DDs (such as 1a (Broadhurst et al., 2003) and 1b (Buchholz et al., 2009)), where sub-types are defined as being mutually non-interacting. Specifically, can a few residues within a conserved docking framework suffice to ensure specificity, or are larger-scale architectural variations required to control partner choice within a PKS assembly line? Despite the advances made here, we still lack a full description of the types and sub-types of DDs that exist, and the differences that serve to distinguish them. Filling this information gap, which has important implications for engineering, will depend on solving further high-resolution structures of PKS DDs and their complexes, coupled with detailed analysis of their interaction specificities. Recognition of additional sub-types would both help to explain why naïve sequence analysis is insufficient for attributing DD classes (**Supplementary Fig. S1**), and improve future DD type prediction in newly-discovered systems. Better DD classification may also ultimately make it possible to discern predictive amino acid 'codes' which govern partner choice within DD sub-classes, the existence of which may have been masked to date by the combined analysis of several sub-types in typical multiple sequence alignments. Nonetheless, our mutagenesis results with 5924 ^{N}DD for which no effect on affinity was observed, argue that single amino acid changes may not be sufficient to reliably redirect docking interactions within sub-types.

Another notable finding of our work is that all of the ^{C}DDs show strong characteristics of IDRs. Although IDRs were previously identified in *trans*-AT PKSs (Dorival et al., 2016), this is a novel finding for *cis*-AT PKSs. It is now clear that nature favours the use of IDRs as docking elements. These, in combination with subtle or more pronounced architectural variation in the partner domains, serve to establish multiple low affinity, high specificity (Mollica et al., 2016) junctions within single PKS systems. The abundance of IDRs in modular PKSs is, on the other hand, a further-complicating factor for sequence-based prediction of DD types, because motifs in natively disordered regions often misalign when classical methods are applied (Thompson et al., 2011). The presence of IDRs will in addition further hinder efforts to adapt DD interaction specificity within sub-types by site-directed mutagenesis, as the effects of point mutations on induced structuration/docking are difficult to predict. These observations argue for the use of matched native pairs in genetic engineering, including the four pairs of intrinsically non-interacting and functionally-independent DDs identified in this work. It is also encouraging that synthetic coiled-coil (SYNZIP) (Klaus et al., 2019) domains have recently been shown to function in engineered PKSs, as they are intrinsically orthogonal to natural DD types.

3. Materials and Methods

3.1. Sequence analysis

The sequences of the type 1a ^{C}DDs and ^{N}DDs were obtained from (Broadhurst et al., 2003), the type 1b domains from (Buchholz et al., 2009), the type 2 DDs from (Whicher et al., 2013), and additional 4 α-helix type domains from (Zeng et al., 2016). For analysis of the putative enacyloxin ^{C}DDs, the C-terminal boundaries of the upstream domains (ACP or KS) were determined by multiple sequence alignment of the subunit ends using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (Madeira et al., 2019), and comparison with domain boundaries identified in the solved structures of homologous domains (Alekseyev et al., 2007; Davison et al., 2014; Gay et al., 2014; Tang et al., 2006). Similarly, the start sites of the domains downstream of the N-terminal docking domains (KS or KR (Keatinge-Clay, 2007; Keatinge-Clay and Stroud, 2006)) were assigned by alignment of the beginning of the subunits and comparison with the sequences of the solved structures of homologous domains. Sequence alignments shown in figures were generated using Clustal Omega and displayed using BoxShade (http://www.ch.embnet.org/software/BOX_form.html). Secondary structure prediction was carried out with PSIPRED (Buchan and Jones, 2019), analysis for disorder and interaction propensity with IUPred2A (https://iupred2a.elte.hu/) (Mészáros et al., 2018), the presence of coiled-coils predicted with the NPS@ web server (https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_lupas.html) (Lupas et al., 1991), and pIs calculated using ExPASY (https://web.expasy.org/compute_pi/) (Bjellqvist et al., 1994, 1993; Walker, 2005).

3.2. Materials and DNA manipulation and sequencing

Burkholderia ambifaria AMMD was purchased from BCCM/LMG (Ghent University, Belgium) and genomic DNA was extracted using Wizard Genomic DNA Purification Kit (Promega). Biochemicals and media were obtained from ThermoFisher Scientific (Tris, EDTA), Merck (NaPi), Carlo ERBA (NaCl), BD (peptone, yeast extract), VWR (glycerol), and Sigma-Aldrich (imidazole, IPTG). Isotopically-labelled chemicals were sourced from Eurisotop. The enzymes for genetic manipulation were purchased from ThermoFisher Scientific. DNA isolation and manipulation in *Escherichia coli* were carried out according to standard methods (Sambrook and Russell, 2001). PCR amplifications were performed on a Mastercycler Pro (Eppendorf) using Phusion high-fidelity DNA polymerase (Thermo Scientific). Primers were obtained from Sigma-Aldrich (France), and DNA sequencing was carried out by GATC Biotech (Mulhouse, France). Synthetic peptides were purchased from GeneCust (Boynes, France).

3.3. Cloning, expression and purification

Nineteen constructs (5925 ^{C}DD, 5924 ^{N}DD, 5924 ^{C}DD, 5924 KS-^{C}DD, 5923 ^{N}DD, 5923 ^{N}DD-KR, 5923 ^{N}DD-KR-ACP, 5923 ^{C}DD long and short versions, 5922 ^{N}DD, 5922 ^{C}DD long and short versions, 5922 ACP-^{C}DD, 5921 ^{N}DD, 5921 ^{N}DD-KS, 5921 ^{C}DD long and short versions, 5920 ^{N}DD long and short versions, 5920 ^{C}DD long and short versions, 5920 ACP-^{C}DD and 5919 ^{N}DD) were amplified directly from *B. ambifaria* genomic DNA using forward primers incorporating a *Bam*HI restriction site (or in select cases, *a Sal*I restriction site) and a reverse primer including a *Hind*III restriction site (**Table S1**), and were ligated into the equivalent sites of vector pBG102 (or pLM302 for 5922 ACP-^{C}DD only). Vector pBG102 codes for a His_6-SUMO tag and pLM302 codes for a His_6-maltose binding protein (MBP) tag (Centre for Structural Biology, Vanderbilt University). Following cleavage of the tags, the proteins incorporated a non-native N-terminal GPGS sequence when *Bam*HI was used or a GPGSPNSSSVD sequence for *Sal*I. Nonetheless, these additional residues were not expected to impact on docking, because it has been shown repeatedly that it is possible to covalently fuse ^{C}DD and ^{N}DD pairs using variable length linkers, without altering the native interactions (Broadhurst et al., 2003; Buchholz et al., 2009; Dorival et al., 2016; Whicher et al., 2013). From selected vectors and *B. ambifaria* genomic DNA, covalent fusions were constructed (5925 ^{C}DD−5924 ^{N}DD and 5920 ^{C}DD−5919 ^{N}DD) using overlap PCR. Specifically, fragments encoding 5925 ^{C}DD, 5924 ^{N}DD, 5920 ^{C}DD and 5919 ^{N}DD were amplified, and then used as templates in a second PCR reaction to obtain the fusion constructs. The resulting fragments were ligated into vector pBG102. Site-directed mutagenesis of 5924 ^{N}DD was carried out using Phusion high-fidelity DNA polymerase and vector pBG102 containing the wild-type sequence.

The vectors were used to transform *E. coli* BL21 or Rosetta 2 (DE3) (Novagen), and constructs were expressed as His_6-SUMO-tagged or His_6-MBP-tagged fusions by growth in LB medium at 37 °C to an A_{600} of 0.7, followed by induction with 300 μM IPTG and incubation at 20 °C for a further 12−18 h. The *E. coli* cells were collected by centrifugation (3500 *g*) and resuspended in buffer 1 (50 mM Tris (pH 8.0), 300 mM NaCl). Cells were lysed by sonication, and cell debris was removed by centrifugation and filtration (0.45 μm). When soluble protein was obtained, the cell lysates were loaded onto a HisTrap 5 mL column (GE), equilibrated in buffer 1. The column was washed extensively with buffer 1 containing 50 mM or 20 mM imidazole, and His-tagged proteins were eluted using 500 mM imidazole.

All His_6-tagged constructs were then incubated with His-tagged human rhinovirus 3C protease (1 μM) for 12−16 h at 4 °C to cleave the affinity-solubility tags. The constructs were then separated from the remaining His-tagged proteins by loading onto a HisTrap 5 mL column (GE) followed by

elution in buffer 1 containing 10 mM or 20 mM imidazole. Following this step, size-exclusion chromatography on a Superdex 75 26/60 column (GE) in 100 mM sodium phosphate (pH 6.5), or 20 mM Tris and 100 mM NaCl (pH 8.0) for 5923 ^{C}DD long version and 5922 ^{N}DD, or on a Superdex 200 26/60 column (GE) in the case of 5923 ^{N}DD-KR, resulted in a homogeneous preparation of each protein (**Supplementary Fig. S3**).

Production of ^{15}N- and ^{13}C,^{15}N-enriched 5925 ^{C}DD, 5924 ^{N}DD, 5919 ^{N}DD, 5925 ^{C}DD–5924 ^{N}DD (covalent fusion) and 5920 ^{C}DD–5919 ^{N}DD (covalent fusion) for NMR structure elucidation was carried out by growth in M9 minimal medium. The minimal medium was supplemented with $^{15}NH_4Cl$ (0.5 g/L) and $^{13}C_6$-glucose (2.0 g/L) as the sole sources of nitrogen and carbon. The isotopically-labelled constructs were purified to homogeneity as described above (**Supplementary Fig. S3**).

3.4. Validation of construct identities by HPLC-MS

Samples were equilibrated to 0.5% TFA, 2% CAN and diluted, if required, to 1 pM, and 5 µl were fractionated through nano-HPLC on an Ultimate 3000 system equipped with a pepMap 100 C18 desalting column and a 15 cm, pepMap RSLC C18 fractionation column (all from ThermoFischer) and eluted using a 2–45 % CAN gradient over 30 min at 300 nL/min. Eluting material was detected based on its absorbance at 214 nm to verify the relative purity of compounds. Fractions (170, 9 s each) were collected on a Proteineer FcII and elutions were mixed on MTP-1536 TF target spots with α-cyano-4-hydroxycinnamic acid (all from Bruker). Spots corresponding to eluting peaks based on absorbance were investigated by MALDI-TOF and MALDI-TOF/TOF (in LIFT mode) mass spectrometry on an Autoflex IV (Bruker), using near neighbour external calibration.

3.5. Size-exclusion chromatography coupled to static light scattering

SEC-SLS was performed using two different setups (SEC-RALS/LALS or SEC-MALS): (i) an AKTA Explorer HPLC system (GE Healthcare) coupled to both RALS/LALS light scattering and refractive index detectors (Viscotek TDA302, Malvern Panalytical). The SEC columns (Shodex KW-802.5 or KW804, 7 µm, 8 mm ID × 300 mm, void volume ~6 mL, total volume ~12.5 mL) were equilibrated with NaPi buffer (100 mM, pH 6.5). The flow rate ranged from 0.2–0.5 mL/min, and the temperature was 30 °C. Data were processed with the Omnisec software (v5.12, Malvern Panalytical) using RALS/LALS detectors for 5923 ^{N}DD-KR, and only the RALS detector for the other proteins; and (ii) (for the 5923 ^{N}DD) AKTA pure (GE Healthcare) coupled with a MALS detector (miniDAWN TREOS II, Wyatt Technology) and a refractometer (Optilab T-rEx, Wyatt Technology). The SEC column (Superdex 75,

10 mm ID × 300 mm, void volume ~8 mL, total volume ~24 mL) was equilibrated with 20 mM Tris.HCl (pH 7.5), 500 mM NaCl. The flow rate was 0.5 mL/min and the temperature was 25 °C. Data were processed with the Astra software (v. 7.14, Wyatt Technology). The refractometer was used as the concentration detector and the refractive index increment value (dn/dc) used to determine the molecular weight was 0.185 mL/g.

3.6. Analysis by circular dichroism

Circular dichroism (CD) was carried out on a Chirascan CD from Applied Photophysics in 100 mM NaPi (pH 6.5 (or pH 8.0 for 5923 ^{C}DD/5922 ^{N}DD)). Data were collected at 0.5 nm intervals in the wavelength ranges of 185/190/195−260 nm at 20 °C, using a temperature-controlled chamber. For most measurements, a 0.01 cm cuvette containing 30 μL of protein sample at 100 μM was used. In the case of 5922 ^{C}DD, 5921 ^{N}DD, 5920 ^{C}DD and 5919 ^{N}DD, measurements were made on 300 μM samples. For analysis of 10 μM protein, 200 μL of sample were used in a 0.1 cm cuvette, and for 1 μM, 1 mL of sample in a 1 cm cuvette. Measurements were performed at least in triplicate, and sample spectra were corrected for buffer background by subtracting the average spectrum of buffer alone. To evaluate the extent of induced folding when two docking domains were combined (Dorival et al., 2016), we first acquired two sets of spectra corresponding to the docking domains alone, either by placing the ^{C}DD into one cuvette and the partner ^{N}DD into another (30 μL each at 100 μM or 300 μM for 5922 ^{C}DD, 5921 ^{N}DD, 5920 ^{C}DD and 5919 ^{N}DD) and analysing the two cuvettes simultaneously, or by averaging the spectra of the ^{C}DD and ^{N}DD acquired separately (these protocols consistently yielded the same result). We then combined the two docking domains, and placed the mixture into both cuvettes (30 μL at 100 μM or 300 μM for 5922 ^{C}DD, 5921 ^{N}DD, 5920 ^{C}DD and 5919 ^{N}DD), again analysing the two simultaneously. Spectrum deconvolution was carried out using the CDNN software (Böhm et al., 1992).

3.7. Analysis by isothermal titration calorimetry

ITC was performed using a MicroCal ITC 200 (GE) at 20 °C in 100 mM NaPi (pH 6.5 (or pH 8.0 for 5923 ^{C}DD/5922 ^{N}DD)). A 300 μL aliquot of ^{C}DD (wild type) or ^{N}DD (wild type or mutant) at 60 μM, 120 μM or 200 μM was placed in the calorimeter cell, and ^{N}DD (wild type or mutant) or wild type ^{C}DD, at 600 μM, 1.8 mM or 2 mM, was added as follows: 0.4 μL over 0.8 s for the first injection, followed by 19 injections of 2 μL over 4 s at 120 s intervals. The heat of reaction per injection (microcalories/second) was determined by integration of the peak areas using either the Origin

(MicroCal) or Nitpic software (Keller et al., 2012). These software provide the best-fit values for the heat of binding (ΔH), the stoichiometry of binding (N), and the dissociation constant (K_d). The heats of dilution were determined by injecting the DDs alone (initial concentrations of 600 μM, 1.8 mM or 2 mM), and were subtracted from the corresponding binding experiments before curve fitting.

3.8. NMR data acquisition

To minimise the amount of protein needed, 300 μL of protein solution (1.0–1.3 mM DDs in 100 mM NaPi (pH 6.5)) were loaded into a 3 mm NMR tube. All NMR data were recorded at 25 °C on a Bruker DRX600 spectrometer equipped with a cryogenic probe (Unité Mixte de Service (UMS) 2008 Ingénierie-Biologie-Santé en Lorraine (IBSLor)). Backbone and sequential resonance assignments were obtained by the combined use of 2D ^{15}N–^{1}H and ^{13}C–^{1}H HSQC spectra and 3D HNCA, HNCACB, CBCA(CO)NH, HNHA, HBHA(CO)NH, HN(CA)CO, and HNCO experiments. Assignments of aliphatic side chain resonance were based on 2D aromatic ^{13}C–^{1}H HSQC, 3D (H)CC(CO)NH, H(CC)(CO)NH, HNCO-ECOSY, CCH−TOCSY, and HCCH-TOCSY experiments (reviewed in (Cavanagh et al., 2007)). To collect NOE-based distance restraints for the 5924 ^{N}DD, 5925 ^{C}DD–5924 ^{N}DD and 5920 ^{C}DD–5919 ^{N}DD structure calculations, 3D ^{15}N NOESY-HSQC and ^{13}C NOESY-HSQC were recorded on uniformly ^{13}C,^{15}N enriched samples. The mixing time used in all NOESY experiments was set to 120 ms. NMR data were processed using Topspin 3.1 (Bruker) and were analysed using Sparky (Goddard TD and Kneller DG, SPARKY3, University of California, San Francisco, 2003).

3.9. NMR structure calculation

Initial structures were generated using CYANA 3.0 software (Güntert, 2004). Starting from a set of manually-assigned NOEs, the standard CYANA protocol of seven iterative cycles of calculations was performed with NOE assignment by the embedded CANDID routine combined with torsion angle dynamics structure calculation (Herrmann et al., 2002). In each cycle, 100 structures starting from random torsion angle values were calculated with 15 000 steps of torsion angle dynamics-driven simulated annealing. A total of 1824 NOE-based distances and 116 backbone angle restraints were used for structure calculation of the 5924 ^{N}DD domain, whereas 2875 and 2504 NOE-based distances associated with 136 and 148 backbone angle restraints were used for the 5925 ^{C}DD–5924 ^{N}DD and 5920 ^{C}DD–5919 ^{N}DD complexes, respectively. To ensure that the overall structures of 5925 ^{C}DD–5924 ^{N}DD and 5920 ^{C}DD–5919 ^{N}DD were homodimeric, weak restraints (9 Å) were applied to the two monomers of the ^{N}DD regions during the first two cycles of CYANA calculation. The angle restraints

were obtained from $^{13}C\alpha$, $^{13}C\beta$, $^{13}C'$, ^{15}N, ^{1}HN, and $^{1}H\alpha$ chemical shifts using TALOS-N (Shen and Bax, 2013) with an assigned minimum range of ±20°.

The second stage consisted of the refinement of the 50 lowest CYANA target function conformers by restrained molecular dynamic (rMD) simulations in Amber 14 (Case et al., 2014; Case et al., 2005) using the generalised Born solvent model. Possible atoms clashes within CYANA initial structures were regularized by a 1 ps energy minimisation in the Amber force field, and then 20 ps of rMD was performed with the following protocol: the system was heated for 5 ps at 600 K with tight coupling for heating and equilibration (TAUTP = 0.4), 15 ps of cooling to 0 K with 13 ps of slow cooling (loose coupling, TAUTP = 4.0−1.0) followed by 1 ps of faster cooling (TAUTP = 1.0), and a final 1 ps of very fast cooling (TAUTP = 0.1−0.05). During rMD, the restraints were slowly ramped from 10−100% of their final values over the first 3 ps. Force constants for distance and angle restraints were set to 32 kcal mol^{-1} $Å^{-2}$ and 50 kcal mol^{-1} rad^{-2}, respectively. The final representative ensembles correspond to the 25 conformers from each calculation with the lowest restraint energy terms.

3.10. Small angle X-ray data collection

SAXS data were collected on the SWING beamline at the Synchrotron SOLEIL (France). Frames were recorded using an Eiger 4M detector at an energy of 12 keV. The distance between the sample and the detector was set to 2000 mm, leading to scattering vectors q ranging from 0.0005 to 0.5 $Å^{-1}$. The scattering vector is defined as $4\pi/\lambda \sin\theta$, where 2θ is the scattering angle. The protein samples in buffer (20 mM Tris.HCl, 200 or 400 mM NaCl, 5% glycerol (as radioprotectant) (pH 8.0)) were injected using the online automatic sample changer into a pre-equilibrated HPLC-coupled size-exclusion chromatography column (Bio-SEC 100 Å (4.6 mm ID × 300 mm), Agilent) at a temperature of 15 °C.

The online purification system that delivers the eluted fractions into the measurement cell was developed on the SWING beamline (David and Pérez, 2009). After equilibrating the column in the protein buffer, 100 successive frames of 0.75 s were recorded of the buffer background. A 50 µL aliquot of the protein sample (at 10 or 30 mg/mL) was then injected, and complete data sets were acquired. The protein concentration downstream of the elution column was followed via the absorbance at 280 nm with an *in situ* spectrophotometer. In contrast to classical SAXS experiments that are conducted in batch using several protein concentrations within a standard range (*e.g.*, 0.1−10 mg/mL^{-1}), the fact that data collection is coupled to a gel-filtration column means that analysis of the required multiple concentrations of the protein occurs within a single experiment because many different positions within the elution peak are sampled during the course of the measurement (typically 50−100 frames are acquired) (David and Pérez, 2009).

Data reduction to absolute units, frame averaging, and solvent subtraction were performed using FOXTROT, a dedicated in-house application. Each acquisition frame of the experiment yielded a scattering spectrum, which was then analysed using FOXTROT, giving a R_g (radius of gyration) as well as an $I(0)$ value, which depends on the protein concentration at that location in the elution peak from the Guinier law (approximation $I(q) = I(0) \exp(-q^2 R_g^2/3)$ for $qR_g < 1.3$). It is thus the fact that the concentration naturally varies across a gel-filtration peak that provides a full range of dilutions as used in classical SAXS experiments. Under our experimental conditions (buffer, dilution over the gel filtration step, etc.), observation of a constant R_g for a significant proportion of the concentrations present in our gel filtration peaks means that these measurements were independent of concentration and thus that we were effectively under conditions of infinite dilution.

In the following step of data analysis, all the frames that exhibited the same R_g as a function of $I(0)$ were corrected for buffer signal and averaged, thus ensuring that the data reflected only the signal arising from the protein structure and not from intermolecular interactions. The distance distribution function $P(r)$ and the maximum particle diameter D_{max} were calculated by Fourier inversion of the scattering intensity $I(q)$ using GNOM (Svergun, 1992).

3.11. Determination of construct molecular weights by SAXS

Classically, molecular weights can be derived from the SAXS data using the $I(0)$ and the known protein concentration. Although in our experimental setup the protein concentration was known when the peak eluted from the gel-filtration column, the fact that there is a delay between exiting the column and the acquisition of the SAXS data meant that the concentration could have changed prior to the measurement. Therefore, to determine the solution molecular weights of our constructs, we relied on SAXS MoW (Fischer et al., 2010), which depends on integration of the distance distribution function (because this is known to be proportional to molecular weight) or the volume correlation (V_c) method (Rambo and Tainer, 2013). Size exclusion chromatography-static light scattering (SEC-SLS), for which the intensity of light scattered off protein molecules in solution is proportional to their molecular weight (Knobloch and Shaklee, 1997), was also employed for certain constructs.

3.12. Ab initio *shape determination*

Ab initio protein shapes were calculated from the experimental SAXS data using the bead-modelling program DAMMIN (Svergun, 1999). At least 20 to 30 independent fits were carried out without restraint, and with a two-fold symmetry restraint when the dimeric state had been

established by SAXS MoW and/or SEC-SLS. The results of several DAMMIN models were then compared using the program SUPCOMB (Kozin and Svergun, 2001) in order to determine a consensus model. Low-resolution models were then averaged using the DAMAVER and DAMFILT packages (Konarev et al., 2003; Volkov et al., 2003). The quality of the 3D modelling was determined using the discrepancy χ^2, defined according to (Konarev et al., 2003). Values lying in the range of 0.9−1.1 are accepted to indicate a good fit between the models and the data. However, the calculation of χ^2 is inversely proportional to the measurement error, that is, the lower the error in the measurement, the higher the χ^2. Because the detector at SOLEIL yields inherently low-error, high-quality data, the effect is to raise the determined χ^2 values above this standard range (Svergun et al., 2013). Nonetheless, visual inspection of the agreement between the theoretical curves calculated from the average molecular forms and the acquired SAXS data confirms the goodness of fit in all cases. Therefore, we decided not to rescale artificially the error on the measurement to give χ^2 of approximately 1, as has been suggested (Svergun et al., 2013). High-resolution structures were then superimposed into the *ab initio* protein shapes using SUPCOMB (Kozin and Svergun, 2001). Figures displaying the protein structures were generated using PyMOL (Schrödinger, LLC), and the low-resolution bead models from DAMMIN were converted into mesh.

3.13. Accession Numbers

Chemical shifts for 5924 ^{N}DD, 5925 ^{C}DD−5924 ^{N}DD and for 5920 ^{C}DD−5919 ^{N}DD have been deposited in the Biological Magnetic Resonance Bank with accession codes **34446**, **34448** and **34447**, respectively. PDB coordinates for 5924 ^{N}DD, 5925 ^{C}DD−5924 ^{N}DD and for 5920 ^{C}DD−5919 ^{N}DD have been deposited under accession codes **6TDD**, **6TDN** and **6TDM**, respectively. The SAXS data have been deposited under accession codes **SASDJE2** (5925 ^{C}DD), **SASDJF2** (5924 ^{N}DD), **SASDJG2** (5920 ^{C}DD), **SASDH2** (5924 ^{N}DD K32A), **SASDJJ2** (5924 ^{N}DD R17A), **SASDJK2** (5920 ^{C}DD−5919 ^{N}DD), **SASDJL2** (5925 ^{C}DD−5924 ^{N}DD).

CRediT authorship contribution statement

Fanny Risser: Investigation, Formal Analysis, Visualization, Writing – review & editing. **Sabrina Collin:** Investigation, Formal Analysis. **Raphael Dos Santos-Morais:** Investigation, Formal Analysis. **Arnaud Gruez:** Conceptualization, Methodology, Investigation, Formal Analysis, Supervision, Project administration, Writing – Review & Editing. **Benjamin Chagot:** Conceptualization, Methodology, Investigation, Formal Analysis, Supervision, Project administration, Visualization, Writing – Review &

Editing. **Kira J. Weissman:** Conceptualization, Methodology, Formal Analysis, Supervision, Project administration, Visualization, Writing – Original Draft, Review & Editing, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by the Agence Nationale de la Recherche [grant number ANR-16-CE92-0006-01 PKS STRUCTURE to K.J.W.], the Centre National de la Recherche Scientifique (CNRS), and the 'IMPACT Biomolecules' project of the Lorraine Université d'Excellence [Investissements d'avenir – grant number ANR 15-004]. We acknowledge Javier Perez and Aurelien Thureau (Soleil Synchrotron, Swing) for help in data acquisition. The NMR, and portions of the ITC and SEC-SLS analysis were carried out on the Plateforme de Biophysique et Biologie Structurale (B2S) of the UMS2008 IBSLor (Université de Lorraine-CNRS-INSERM – http://umsibslor.univ-lorraine.fr), and we would also like to thank Jean-Baptiste Vincourt from the Proteomics Core Facility. Complementary ITC data and SEC-SLS data were acquired on the Plateforme Approches fonctionnelles et Structurales des InterActions cellulaires (ASIA), of the Faculté des Sciences et Technologies (Université de Lorraine). Appreciation also goes to Alexandre Kriznik for assistance in analysing the ITC and CD data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at….

References

AMBER 14, D. A. Case, V. Babin, J. T. Berryman, R. M. Betz, Q. Cai, D. S. Cerutti, T. E. Cheatham III, T. A. Darden, R. E. Duke, H. Gohlke, A. W. Goetz, S. Gusarov, N. Homeyer, P. Janowski, J. Kaus, I. Kolossváry, A. Kovalenko, T. S. Lee, S. LeGrand, T. Luchko, R. Luo, B. Madej, K. M. Merz, F. Paesani, D. R. Roe, A. Roitberg, C. Sagui, R. Salomon-Ferrer, G. Seabra, C. L. Simmerling, W. Smith, J. Swails, R. C. Walker, J. Wang, R. M. Wolf, X. Wu, P. A. Kollman (2014) University of California, San Francisco.

Alekseyev, V.Y., Liu, C.W., Cane, D.E., Puglisi, J.D., Khosla, C., 2007. Solution structure and proposed domain domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. Protein Sci. 6, 2093–2107. https://doi.org/10.1110/ps.073011407.

Bjellqvist, B., Basse, B., Olsen, E., Celis, J.E., 1994. Reference points for comparisons of two-dimensional maps of proteins from different human cell types defined in a pH scale where isoelectric points correlate with polypeptide compositions. Electrophoresis 15, 529–539. https://doi.org/10.1002/elps.1150150171.

Bjellqvist, B., Hughes, G.J., Pasquali, C., Paquet, N., Ravier, F., Sanchez, J.C., Frutiger, S., Hochstrasser, D., 1993. The focusing positions of polypeptides in immobilized pH gradients can be predicted from their amino acid sequences. Electrophoresis 14, 1023–1031. https://doi.org/10.1002/elps.11501401163.

Böhm, G., Muhr, R., Jaenicke, R., 1992. Quantitative analysis of protein far UV circular dichroism spectra by neural networks. Protein Eng. 5, 191–195. https://doi.org/10.1093/protein/5.3.191.

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

Buchan, D.W.A., Jones, D.T., 2019. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Res. 47, W402–W407. https://doi.org/10.1093/nar/gkz297.

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 Chem. Biol. 4, 41–52. https://doi.org/10.1021/cb8002607.

Case, D.A., Cheatham, T.E., Darden, T., Gohlke, H., Luo, R., Merz, K.M., Onufriev, A., Simmerling, C., Wang, B., Woods, R.J., 2005. The Amber biomolecular simulation programs. J. Comput. Chem. 26, 1668–1688. https://doi.org/10.1002/jcc.20290.

Cavanagh, J., Fairbrother, W. J., Palmer, A. G., Rance, M., Skelton, N.J. (Eds), 2007., in: Protein NMR Spectroscopy (Second Edition). Academic Press, Burlington. https://doi.org/10.1016/B978-012164491-8/50000-2

Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., Richardson, D.C., 2010. MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 66, 12–21. https://doi.org/10.1107/S0907444909042073.

David, G., Pérez, J., 2009. Combined sampler robot and high-performance liquid chromatography: a fully automated system for biological small-angle X-ray scattering experiments at the Synchrotron SOLEIL SWING beamline. J. Appl. Crystallogr. 42, 892–900. https://doi.org/10.1107/S0021889809029288.

Davison, E.K., Brimble, M.A., 2019. Natural product derived privileged scaffolds in drug discovery. Curr. Opin. Chem. Biol. 52, 1–8. https://doi.org/10.1016/j.cbpa.2018.12.007.

Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A., Weissman, K.J., 2014. Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. Chem. Sci. 5, 3081–3095. https://doi.org/10.1039/C3SC53511H.

Dorival, J., Annaval, T., Risser, F., Collin, S., Roblin, P., Jacob, C., Gruez, A., Chagot, B., Weissman, K.J., 2016. Characterization of intersubunit communication in the virginiamycin trans-acyl transferase polyketide synthase. J. Am. Chem. Soc. 138, 4155–4167. https://doi.org/10.1021/jacs.5b13372.

Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., Skiniotis, G., 2014. Structure of a modular polyketide synthase. Nature 510, 512–517. https://doi.org/10.1038/nature13423.

Fischer, H., de Oliveira Neto, M., Napolitano, H.B., Polikarpov, I., Craievich, A.F., 2010. Determination of the molecular weight of proteins in solution from a single small-angle X-ray scattering measurement on a relative scale. J. Appl. Cryst. 43, 101–109. https://doi.org/10.1107/S0021889809043076.

Gay, D.C., Gay, G., Axelrod, A.J., Jenner, M., Kohlhaas, C., Kampa, A., Oldham, N.J., Piel, J., Keatinge-Clay, A.T., 2014. A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. Structure 22, 444–451. https://doi.org/10.1016/j.str.2013.12.016.

Gokhale, R.S., Hunziker, D., Cane, D.E., Khosla, C., 1999. Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. Chem. Biol. 6, 117–125. https://doi.org/10.1016/S1074-5521(99)80008-8.

Güntert, P., 2004. Automated NMR structure calculation with CYANA. Methods Mol. Biol. 278, 353–378. https://doi.org/10.1385/1-59259-809-9:353.

Hacker, C., Cai, X., Kegler, C., Zhao, L., Weickhmann, A.K., Wurm, J.P., Bode, H.B., Wöhnert, J., 2018. Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. Nat. Commun. 9, 4366. https://doi.org/10.1038/s41467-018-06712-1.

Helfrich, E.J.N., Piel, J., 2016. Biosynthesis of polyketides by trans-AT polyketide synthases. Nat. Prod. Rep. 33, 231–316. https://doi.org/10.1039/c5np00125k.

Herrmann, T., Güntert, P., Wüthrich, K., 2002. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J. Mol. Biol. 319, 209–227. https://doi.org/10.1016/S0022-2836(02)00241-3.

Jenke-Kodama, H., Dittmann, E., 2009. Evolution of metabolic diversity: insights from microbial polyketide synthases. Phytochemistry 70, 1858–1866. https://doi.org/10.1016/j.phytochem.2009.05.021.

Jenke-Kodama, H., Sandmann, A., Müller, R., Dittmann, E., 2005. Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. 22, 2027–2039. https://doi.org/10.1093/molbev/msi193.

Jenner, M., Kosol, S., Griffiths, D., Prasongpholchai, P., Manzi, L., Barrow, A.S., Moses, J.E., Oldham, N.J., Lewandowski, J.R., Challis, G.L., 2018. Mechanism of intersubunit ketosynthase-dehydratase interaction in polyketide synthases. Nat. Chem. Biol. 14, 270–275. https://doi.org/10.1038/nchembio.2549.

Keatinge-Clay, A.T., 2007. A tylosin ketoreductase reveals how chirality is determined in polyketides. Chem. Biol. 14, 898–908. https://doi.org/10.1016/j.chembiol.2007.07.009.

Keatinge-Clay, A.T., Stroud, R.M., 2006. The structure of a ketoreductase determines the organization of the β-carbon processing enzymes of modular polyketide synthases. Structure 1993 14, 737–748. https://doi.org/10.1016/j.str.2006.01.009.

Keller, S., Vargas, C., Zhao, H., Piszczek, G., Brautigam, C.A., Schuck, P., 2012. High-precision isothermal titration calorimetry with automated peak-shape analysis. Anal. Chem. 84, 5066–5073. https://doi.org/10.1021/ac3007522.

Klaus, M., D'Souza, A.D., Nivina, A., Khosla, C., Grininger, M., 2019. Engineering of chimeric polyketide synthases using SYNZIP docking domains. ACS Chem. Biol. 14, 426–433. https://doi.org/10.1021/acschembio.8b01060.

Knobloch, J.E., Shaklee, P.N., 1997. Absolute molecular weight distribution of low-molecular-weight heparins by size-exclusion chromatography with multiangle laser light scattering detection. Anal. Biochem. 245, 231–241. https://doi.org/10.1006/abio.1996.9984.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., Svergun, D.I., 2003. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. J. Appl. Crystallogr. 36, 1277–1282. https://doi.org/10.1107/S0021889803012779.

Kosol, S., Gallo, A., Griffiths, D., Valentic, T.R., Masschelein, J., Jenner, M., de Los Santos, E.L.C., Manzi, L., Sydor, P.K., Rea, D., Zhou, S., Fülöp, V., Oldham, N.J., Tsai, S.C., Challis, G.L., Lewandowski, J.R., 2019. Structural basis for chain release from the enacyloxin polyketide synthase. Nat. Chem. 11, 913–923. https://doi.org/doi: 10.1038/s41557-019-0335-5.

Kozin, M.B., Svergun, D.I., 2001. Automated matching of high- and low-resolution structural models. J. Appl. Crystallogr. 34, 33–41. https://doi.org/10.1107/S0021889800014126.

Kuo, J., Lynch, S.R., Liu, C.W., Xiao, X., Khosla, C., 2016. Partial in vitro reconstitution of an orphan polyketide synthase associated with clinical cases of Nocardiosis. ACS Chem. Biol. 11, 2636–2641. https://doi.org/10.1021/acschembio.6b00489.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., Thornton, J.M., 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486. https://doi.org/10.1007/bf00228148.

Longhi, S., Receveur-Bréchot, V., Karlin, D., Johansson, K., Darbon, H., Bhella, D., Yeo, R., Finet, S., Canard, B., 2003. The C-terminal domain of the measles virus nucleoprotein is intrinsically disordered and folds upon binding to the C-terminal moiety of the phosphoprotein. J. Biol. Chem. 278, 18638–18648. https://doi.org/10.1074/jbc.M300518200.

Lupas, A., Van Dyke, M., Stock, J., 1991. Predicting coiled coils from protein sequences. Science 252, 1162–1164. https://doi.org/10.1126/science.252.5009.1162.

Madeira, F., Park, Y.M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A.R.N., Potter, S.C., Finn, R.D., Lopez, R., 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res. 47, W636–W641. https://doi.org/10.1093/nar/gkz268.

Mahenthiralingam, E., Song, L., Sass, A., White, J., Wilmot, C., Marchbank, A., Boaisha, O., Paine, J., Knight, D., Challis, G.L., 2011. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic Burkholderia ambifaria genomic island. Chem. Biol. 18, 665–677. https://doi.org/10.1016/j.chembiol.2011.01.020.

Masschelein, J., Sydor, P.K., Hobson, C., Howe, R., Jones, C., Roberts, D.M., Yap, Z.L., Parkhill, J., Mahenthiralingam, E., G. L. Challis, 2019. A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis. Nat. Chem. 11, 913–923. https://doi.org/doi: 10.1038/s41557-019-0335-5.

McGuffin, L.J., Bryson, K., Jones, D.T., 2000. The PSIPRED protein structure prediction server. Bioinformatics 16, 404–405. https://doi.org/10.1093/bioinformatics/16.4.404.

Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K., de Bruijn, I., Chooi, Y.H., Claesen, J., Coates, R.C., Cruz-Morales, P., Duddela, S., Düsterhus, S., Edwards, D.J., Fewer, D.P., Garg, N., Geiger, C., Gomez-Escribano, J.P., Greule, A., Hadjithomas, M., Haines, A.S., Helfrich, E.J.N., Hillwig, M.L., Ishida, K., Jones, A.C., Jones, C.S., Jungmann, K., Kegler, C., Kim, H.U., Kötter, P., Krug, D., Masschelein, J., Melnik, A.V., Mantovani, S.M., Monroe, E.A., Moore, M., Moss, N., Nützmann, H.-W., Pan, G., Pati, A., Petras, D., Reen, F.J., Rosconi, F., Rui, Z., Tian, Z., Tobias, N.J., Tsunematsu, Y., Wiemann, P., Wyckoff, E., Yan, X., Yim, G., Yu, F., Xie, Y., Aigle, B., Apel, A.K., Balibar, C.J., Balskus, E.P., Barona-Gómez, F., Bechthold, A., Bode, H.B., Borriss, R., Brady, S.F., Brakhage, A.A., Caffrey, P., Cheng, Y.-Q., Clardy, J., Cox, R.J., De Mot, R., Donadio, S., Donia, M.S., van der Donk, W.A., Dorrestein, P.C., Doyle, S., Driessen, A.J.M., Ehling-Schulz, M., Entian, K.-D., Fischbach, M.A., Gerwick, L., Gerwick, W.H., Gross, H., Gust, B., Hertweck, C., Höfte, M., Jensen, S.E., Ju, J., Katz, L., Kaysser, L., Klassen, J.L., Keller, N.P., Kormanec, J., Kuipers, O.P., Kuzuyama, T., Kyrpides, N.C., Kwon, H.-J., Lautru, S., Lavigne, R., Lee, C.Y., Linquan, B., Liu, X., Liu, W., Luzhetskyy, A., Mahmud, T., Mast, Y., Méndez, C., Metsä-Ketelä, M., Micklefield, J., Mitchell, D.A., Moore, B.S., Moreira, L.M., Müller, R., Neilan, B.A., Nett, M., Nielsen, J., O'Gara, F., Oikawa, H., Osbourn, A., Osburne, M.S., Ostash, B., Payne, S.M., Pernodet, J.-L., Petricek, M., Piel, J., Ploux, O., Raaijmakers, J.M., Salas, J.A., Schmitt, E.K., Scott, B., Seipke, R.F., Shen, B., Sherman, D.H., Sivonen, K., Smanski, M.J., Sosio, M., Stegmann, E., Süssmuth, R.D., Tahlan, K., Thomas, C.M., Tang, Y., Truman, A.W., Viaud, M., Walton, J.D., Walsh, C.T., Weber, T., van Wezel, G.P., Wilkinson, B., Willey, J.M., Wohlleben, W., Wright, G.D., Ziemert, N., Zhang, C., Zotchev, S.B., Breitling, R., Takano, E., Glöckner, F.O., 2015. Minimum information about a biosynthetic gene cluster. Nat. Chem. Biol. 11, 625–631. https://doi.org/10.1038/nchembio.1890.

Mészáros, B., Erdos, G., Dosztányi, Z., 2018. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. Nucleic Acids Res. 46, W329–W337. https://doi.org/10.1093/nar/gky384.

Mollica, L., Bessa, L.M., Hanoulle, X., Jensen, M.R., Blackledge, M., Schneider, R., 2016. Binding mechanisms of intrinsically disordered proteins: theory, simulation, and experiment. Front. Mol. Biosci. 3, 52. https://doi.org/10.3389/fmolb.2016.00052.

Musiol, E.M., Härtner, T., Kulik, A., Moldenhauer, J., Piel, J., Wohlleben, W., Weber, T., 2011. Supramolecular templating in kirromycin biosynthesis: the acyltransferase KirCII loads ethylmalonyl-CoA extender onto a specific ACP of the trans-AT PKS. Chem. Biol. 18, 438–444. https://doi.org/10.1016/j.chembiol.2011.02.007.

Nooren, I.M.A., Thornton, J.M., 2003. Structural characterisation and functional significance of transient protein-protein interactions. J. Mol. Biol. 325, 991–1018.

Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C., Piel, J., 2008. Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. Nat. Biotechnol. 26, 225–233. https://doi.org/10.1038/nbt1379.

Ozbabacan, S.E. A., Engin, H.B., Gursoy, A., Keskin, O., 2011. Transient protein-protein interactions. Protein Eng. Des. Sel. 24, 635–648. https://doi.org/10.1093/protein/gzr025.

Rambo, R.P., Tainer, J.A., 2013. Accurate assessment of mass, models and resolution by small-angle scattering. Nature 496, 477–481. https://doi.org/10.1038/nature12070.

Receveur-Bréchot, V., Durand, D., 2012. How random are intrinsically disordered proteins? A small angle scattering perspective. Curr. Protein Pep. Sci. 13, 55–75.

Richter, C.D., Nietlispach, D., Broadhurst, R.W., Weissman, K.J., 2008. Multienzyme docking in hybrid megasynthetases. Nat. Chem. Biol. 4, 75–81. https://doi.org/10.1038/nchembio.2007.61.

Sahu, D., Bastidas, M., Lawrence, C.W., Noid, W.G., Showalter, S.A., 2016. Assessing coupled protein folding and binding through temperature-dependent isothermal titration calorimetry. Methods Enzymol. 567, 23–45. https://doi.org/10.1016/bs.mie.2015.07.032.

Sambrook, J., Russell, D., 2001. Molecular Cloning: A Laboratory Manual (Third Edition), Cold Spring Harbor Laboratory Press, New York.

Shammas, S.L., Crabtree, M.D., Dahal, L., Wicky, B.I.M., Clarke, J., 2016. Insights into coupled folding and binding mechanisms from kinetic studies. J. Biol. Chem. 291, 6689–6695. https://doi.org/10.1074/jbc.R115.692715.

Shen, Y., Bax, A., 2013. Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. J. Biomol. NMR 56, 227–241. https://doi.org/10.1007/s10858-013-9741-y.

Staunton, J., Weissman, K.J., 2001. Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. 18, 380–416.

Straight, P.D., Fischbach, M.A., Walsh, C.T., Rudner, D.Z., Kolter, R., 2007. A singular enzymatic megacomplex from Bacillus subtilis. Proc. Natl. Acad. Sci. U. S. A. 104, 305–310. https://doi.org/10.1073/pnas.0609073103.

Svergun, D.I., 1999. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. Biophys. J. 76, 2879–2886. https://doi.org/10.1016/S0006-3495(99)77443-6.

Svergun, D.I., 1992. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. J. Appl. Crystallogr. 25, 495–503. https://doi.org/10.1107/S0021889892001663.

Svergun, D.I., Koch, M.H.J., Timmins, P.A., May, R.P., 2013. Small angle X-ray and neutron scattering from solutions of biological macromolecules. OUP Oxford.

Tang, Y., Kim, C.-Y., Mathews, I.I., Cane, D.E., Khosla, C., 2006. The 2.7-Angstrom crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. U. S. A. 103, 11124–11129. https://doi.org/10.1073/pnas.0601924103.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S., 2010. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538. https://doi.org/10.1126/science.1188308.

Thompson, J.D., Linard, B., Lecompte, O., Poch, O., 2011. A comprehensive benchmark study of multiple sequence alignment methods: current challenges and future perspectives. PloS One 6, e18093. https://doi.org/10.1371/journal.pone.0018093.

Volkov, V.V., Svergun, D.I., IUCr, 2003. Uniqueness of ab initio shape determination in small-angle scattering. J. Appl. Crystallogr. 36, 860–864, https://doi.org/10.1107/S0021889803000268.

Walker, J.M. (Ed.), 2005. The Proteomics Protocols Handbook. Humana Press. https://doi.org/10.1385/1592598900.

Wang, Y., Marrero, M.C., Medema, M.H., Dijk, A.D.J. van, 2020. Coevolution-based prediction of protein-protein interactions in polyketide biosynthetic assembly lines. Bioinformatics. In press. doi: 10.1093/bioinformatics/btaa595

Weber, T., Laiple, K.J., Pross, E.K., Textor, A., Grond, S., Welzel, K., Pelzer, S., Vente, A., Wohlleben, W., 2008. Molecular analysis of the kirromycin biosynthetic gene cluster revealed β-alanine as precursor of the pyridone moiety. Chem. Biol. 15, 175–188. https://doi.org/10.1016/j.chembiol.2007.12.009.

Weissman, K.J., 2015. Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. Nat. Prod. Rep. 33, 203–230. https://doi.org/10.1039/c5np00109a

Weissman, K.J., 2009. Introduction to polyketide biosynthesis. Methods Enzymol. 459, 3–16. https://doi.org/10.1016/S0076-6879(09)04601-1.

Weissman, K.J., 2006. The structural basis for docking in modular polyketide biosynthesis. Chembiochem 7, 485–494. https://doi.org/10.1002/cbic.200500435.

Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., Skiniotis, G., 2014. Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature 510, 560–564. https://doi.org/10.1038/nature13409.

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. Chem. Biol. 20, 1340–1351. https://doi.org/10.1016/j.chembiol.2013.09.015.

Wu, N., Cane, D.E., Khosla, C., 2002. Quantitative analysis of the relative contributions of donor acyl carrier proteins, acceptor ketosynthases, and linker regions to intermodular transfer of intermediates in hybrid polyketide synthases. Biochemistry 41, 5056–5066. https://doi.org/10.1021/bi012086u.

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 Chem. Biol. 11, 2466–2474. https://doi.org/10.1021/acschembio.6b00345.

Zheng, J., Fage, C.D., Demeler, B., Hoffman, D.W., Keatinge-Clay, A.T., 2013. The missing linker: a dimerization motif located within polyketide synthase modules. ACS Chem. Biol. 8, 1263–1270. https://doi.org/10.1021/cb400047s.

Figure captions

Fig. 1. The enacyloxin biosynthetic pathway. The biosynthesis occurs in two stages: assembly of the core by a hybrid polyketide synthase (PKS) (solid spheres)/nonribosomal peptide synthetase (NRPS) (hollow spheres), followed by elaboration of the released chain by a series of post-assembly line enzymes. The PKS portion of the pathway encompasses seven large subunits (Bamb_5925–5919, boxed), the first six of which belong to the *cis*-AT PKS class, and the last of which is a *trans*-AT PKS (Mahenthiralingam et al., 2011). Each subunit contains one or more functional modules (one loading

and 10 chain extension), which are themselves composed of catalytic and acyl carrier protein (ACP) domains. As determined in this work, communication across the PKS intersubunit interfaces is mediated by a collection of docking domains belonging to at least two different types: 1-related and 2-related (DD types were assigned based on structure elucidation at high resolution (in bold), or detailed biophyiscal and bioinformatics analysis), while a type 3 DD (or SLiM–β-hairpin (Kosol et al., 2019)) was recently shown to operate at the PCP (Bamb_5927)/C (Bamb_5915) interface involved in chain off-loading. Key: MT, *C*-methyl transferase; GNAT, Gcn5-related *N*-acetyl transferase; KS, ketosynthase (the ° indicates a condensation-inactive version which serves as an ACP-transacylase); AT, acyl (malonyl) transferase; DH, dehydratase; KR, ketoreductase; PCP, peptidyl carrier protein; C, condensation. ΔpI: the difference in calculated pIs (Bjellqvist et al., 1994, 1993; Walker, 2005) for the identified ^{C}DDs and ^{N}DDs acting at each interface.

Fig. 2. Docking domains (DDs) in modular PKSs. (A) The four structural classes of DDs characterised to date from *cis*-AT PKSs (Broadhurst et al., 2003; Buchholz et al., 2009; Whicher et al., 2013), are shown in the presence of typical up- and downstream flanking domains. The C-terminal DDs (^{C}DDs) are illustrated in blue, and the ^{N}DDs in red. Class 3 is notably present in hybrid PKS/NRPS systems (where the NDD is homodimeric) (Richter et al., 2008), as well as pure NRPSs (where the ^{N}DD is monomeric) (Hacker et al., 2018). (B) Classes of DDs identified to date in *trans*-AT PKS systems. The 4 α-helix bundle type is formed by two α-helices contributed by each of the ^{C}DD and ^{N}DD (Dorival et al., 2016). The DHD domains (for 'dehydratase (DH) docking') (shown in thick blue lines) which are present at KS/DH junctions, are intrinsically disordered regions (IDRs) which directly contact the downstream DH domain (Jenner et al., 2018). The ^{N}DD of the four α-helix bundle type of docking domain (shown in red) is also an IDR. Key: ACP, acyl carrier protein; KS, ketosynthase; PCP, peptidyl carrier protein; C, condensation.

Fig. 3. Analysis of independent and induced folding of DDs by CD. (A) CD analysis of the discrete ^{C}DDs and ^{N}DDs from each interface (the investigated construct is indicated). The % α-helix calculated with the CDNN software (Böhm et al., 1992) is indicated. (B) Differential CD analysis of the DD pairs. In each case, a combined spectrum was obtained by simultaneously analysing the two isolated DDs (both at 100 μM). Comparison with the spectrum of a mixture of the two docking domains, again at an overall concentration of 100 μM, was used to identify increased structuration. The % α-helix calculated with the CDNN software (Böhm et al., 1992) under the two conditions is indicated.

Fig. 4. Characterisation of the 5925 ^{C}DD/5924 ^{N}DD and 5920 ^{C}DD/5919 ^{N}DD interactions by NMR. (A) Saturation experiment with ^{15}N-labeled 5925 ^{C}DD in the presence of unlabelled 5924 ^{N}DD (concentration ratios (^{C}DD:^{N}DD): (1) 400 μM:0 μM; (2) 200 μM:240 μM; (3) 100 μM:400 μM; and, (4) 60 μM:600 μM). (B) Saturation experiment with ^{15}N-labeled 5924 ^{N}DD in the presence of unlabelled 5925 ^{C}DD (concentration ratios (^{N}DD:^{C}DD): (1) 400 μM:0 μM; (2) 300 μM:350 μM; and, (3) 76 μM:600 μM). (C) Saturation experiment with ^{15}N-labelled 5919 ^{N}DD in the presence of unlabelled 5920 ^{C}DD (concentration ratios (^{N}DD:^{C}DD): (1) 150 μM:0 μM; (2) 150 μM:300 μM; (3) 200 μM:1000 μM; (4) 200 μM:4000 μM).

Fig. 5. NMR structures of 5924 ^{N}DD and the 5925 ^{C}DD–5924 ^{N}DD complex. (A) Cartoon representation of the 5924 ^{N}DD dimer, with one monomer coloured in blue and the other in light blue (residues indicated with primes). (B) Interactions between the protein chains. Top and side views of the residues (represented as sticks) involved in dimer formation. Basic residues are shown in blue, polar in light green, aromatic in purple and aliphatic in yellow. (C) Comparison of dimer topology between CurH ^{N}DD and 5924 ^{N}DD. The 5924 ^{N}DD dimer is represented as in (A). The CurH ^{N}DD chains are coloured in orange and green. To emphasise the overall difference in architecture, one monomer of CurH ^{N}DD is superimposed on one monomer of 5924 ^{N}DD. (D) Cartoon representation of the 5925 ^{C}DD–5924 ^{N}DD dimer. 5925 ^{C}DD is shown in red (dark and light for the two monomers), 5924 ^{N}DD in blue, and the $(G_3S)_3$ linker in beige. The α-helices present in 5925 ^{C}DD are indicated as α1 and α2, and those in 5924 ^{N}DD as αA and αB. The primes indicate the second monomer of each DD. (E) Interactions between protein chains in the 5925 ^{C}DD–5924 ^{N}DD dimer. Residues involved in hydrophobic interactions (top view) and in polar interactions (bottom view) are shown as sticks. Basic residues are coloured in blue, polar in light green, methionine in dark green, aromatic in purple and aliphatic in yellow. (F) Comparison of the mode of ^{C}DD/^{N}DD interaction between CurG ^{C}DD–CurH ^{N}DD and 5925 ^{C}DD–5924 ^{N}DD. The figure shows a monomer of CurH ^{N}DD (orange) interacting with a monomer of CurG ^{C}DD (green), superimposed on monomers of 5925 ^{C}DD and 5924 ^{N}DD, as shown in (D). For clarity of presentation, certain linkers in the structures have been removed.

Fig. 6. NMR structure of the 5920 ^{C}DD–5919 ^{N}DD complex. (A) Cartoon representation of the 5920 ^{C}DD–5919 ^{N}DD fusion construct, with 5920 ^{C}DD shown in yellow and 5919 ^{N}DD in green. (B) Hydrophobic interactions between 5920 ^{C}DD and 5919 ^{N}DD. Residues are shown as sticks. (C) Polar interactions between 5920 ^{C}DD and 5919 ^{N}DD. Residues are shown as sticks. In (B) and (C), aliphatic residues are coloured in yellow, methionine in green, basic in blue and acidic in red. For clarity of presentation, linkers in panels (B) and (C) have been removed.

Table 1. Summary of findings on the enacyloxin docking domains (^{C}DDs = grey columns, ^{N}DDs = green columns) relative to known DD classes (1a, 1b, 2 and 4HB).

^{C}DD	No. α-helices	Length docking motif (aa)[a]	IDR	Quaternary structure	pI[c]	Docking affinity K_d (μM)[d]	pI[c]	Quaternary structure	IDR	Length docking motif (aa)[a]	No. α-helices	^{N}DD
Type 1a	3	12 (1 α-helix)	n.d.	Monomer		n.d.		Homodimer	n.d.	26	1	Type 1a
Type 1b	4	13 (2 α-helices)	n.d.	Monomer		45 ± 22		Homodimer	n.d.	25	1	Type 1b
Type 2	2	22–24 (2 α-helices)	n.d.	Monomer		16 ± 5; 4 ± 2		Homodimer	n.d.	αA: 10–18 αB: 22 (32–38 total)	2	Type 2
4HB	2	25 (2 α-helices)	N	Monomer		5.8 ± 0.2		Monomer	Y	25	2	4HB
5925[e]	2	17	Y	Monomer	3.8	1.1 ± 0.1	9.7	Homodimer	N	31	2	5924
5924	1[b]	22	Y	Monomer	4.6	None	10.6	Homodimer	N	34	2[b]	5923
5923	1+[b]	54	Y	Monomer	5.0	26	5.4	Homodimer	N	32	1[b]	5922
5922	2 (1)[b]	70 (21)	Y	Monomer	4.1	57 (90)	9.7	Monomer	N	31	1[b]	5921
5921	2[b]	20	Y	Monomer	3.6	8.5	6.4	Homodimer	N	31	1[b]	5920
5920	1	11 (1 α-helix)	Y	Monomer	3.7	43	10.2	Homodimer	N	27	1	5919

[a]Based on solved NMR or crystal structure for Type 1a (Broadhurst et al., 2003), 1b (Buchholz et al., 2009), 2 (Whicher et al., 2014), 4HB (Dorival et al., 2016; Zeng et al., 2016), 5925 and 5920, while for 5924–5921, the experimentally-determined maximal functional length of the domain is indicated; in the case of 5922 ^{C}DD, the full-length domain spans 80 residues, but the C-terminal 21 aa (data shown in parentheses) also interact with partner 5921 ^{N}DD.

[b]Predicted using PsiPRED (Buchan and Jones, 2019; McGuffin et al., 2000).

[c]Predicted using Compute pI (Bjellqvist et al., 1994, 1993; Walker, 2005).

[d]For the type 1b domains (Buchholz et al., 2009), K_d represents the affinity between ACP-^{C}DD and ^{N}DD; for the type 2 domains (Whicher et al., 2014), between ACP-^{C}DD and ^{N}DD-KS-AT (CurG/CurH and CurK/CurL, respectively); for all other pairs, affinity was measured for interaction between the discrete DDs.

[e]The color coding of the enacyloxin DDs indicates definitive class assignments (5925/5924 (violet, type 2-related) and 5920/5919 (yellow, type 1-related, though truncated to one α-helix)) and putative class assignments (5924/5923 (type 2-related), and 5923/5922, 5922/5921 and 5921/5920 (type 1-related)).

n.d. = not determined; IDR = intrinsically disordered region.