Published in final edited form as:

ACS Chem Biol. 2016 September 16; 11(9): 2466–2474. doi:10.1021/acscchembio.6b00345.

Portability and Structure of the Four-Helix Bundle Docking Domains of *trans*-Acyltransferase Modular Polyketide Synthases

Jia Zeng[†], Drew T. Wagner[†], Zhicheng Zhang[‡], Luisa Moretto[†], Janci D. Addison[†], and Adrian T. Keatinge-Clay^{†,‡}

[†]Department of Molecular Biosciences, The University of Texas at Austin, Austin, Texas 78712, United States

[‡]Department of Chemistry, The University of Texas at Austin, Austin, Texas 78712, United States

Abstract

The polypeptides of multimodular polyketide synthases self-assemble into biosynthetic factories. While the docking domains that mediate the assembly of *cis*-acyltransferase polyketide synthase polypeptides are well-studied, those of the more recently discovered *trans*-acyltransferase polyketide synthases have just started to be described. Located at the C- and N-termini of many polypeptides, these 25-residue, two-helix, pseudosymmetric motifs noncovalently connect domains both between and within modules. Domains expressed with their natural, cognate docking motifs formed complexes stable to size-exclusion chromatography with 1–10 μ M dissociation constants as measured by isothermal titration calorimetry. Deletion and swapping experiments demonstrate portability of the docking motifs. A 1.72 Å-resolution structure of the N-terminal portion of the macrolactin synthase polypeptide MlnE shows an uncomplexed N-terminal docking motif to be preorganized in the conformation it assumes within the docking domain complex.

Graphical abstract

Correspondence to: Adrian T. Keatinge-Clay.

Supporting Information: The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscchem-bio.6b00345.

Accession Codes: The coordinates and structure factors for ^NDD_{MlnD}+MlnKR6 have been deposited (PDB Code 5D2E).

Notes: The authors declare no competing financial interest.

Multimodular polyketide synthases (PKSs) are multi-enzyme factories that synthesize bioactive, complex molecules from simple acyl-CoA precursors.^{1–3} They are generally comprised of several polypeptides that noncovalently self-assemble into megadalton structures. A single synthase can contain more than 100 distinct domains, and the bacillaene synthase has been observed to accrete into an ~100 MDa organelle.^{4,5} While protein–protein interactions are known to mediate the ordered assembly of PKS polypeptides, how they do so remains a subject of active investigation.^{6,7}

Modules capable of extending a polyketide intermediate minimally harbor two domains: a carbon–carbon bond-forming ketosynthase (KS) and an acyl carrier protein (ACP). Modules from *cis*-AT PKSs also include an acyltransferase (AT) specific for the acyl-CoA to be added to the growing polyketide, while those from *trans*-AT PKSs rely on separately encoded ATs.^{2,3} Modules may also contain processing enzymes such as a methyltransferase (MT), a ketoreductase (KR), a dehydratase (DH), and an enoylreductase (ER). Polypeptides containing as many as nine modules (e.g., the 16 990-residue MlsA1 of the mycolactone synthase) are known.⁸

At the termini of *cis*-AT PKS polypeptides, docking motifs help mediate the ordered linkage of modules (defined as starting with a KS and ending with an ACP).^{9,10} They are present C-terminal to the ACP of an upstream polypeptide and N-terminal to the KS of a downstream polypeptide but have not been observed within a module to split it between polypeptides. The first two helices of the three-helix C-terminal docking motif and the long N-terminal docking motif helix are thought to mediate polypeptide homodimerization, while the third helix of the C-terminal docking motif and the coiled-coil of the N-terminal docking motif have been observed to associate into a four-helix bundle thought to bring partner polypeptides together. These C- and N-terminal docking motifs are largely portable and have been employed to aid in the physical association of engineered polypeptides.^{11,12}

How the polypeptides of the recently discovered *trans*-AT PKSs assemble is just starting to be described. In contrast to *cis*-AT PKSs, which are almost always disconnected between modules, *trans*-AT PKSs are most commonly disconnected within modules. Thus, junctions appear not only between ACP and KS, but apparently between any two domains (e.g., KS/DH, KS/KR, DH/KR, KR/MT). A pair of docking motifs (^CDD and ^NDD, where DD stands for “docking domain”) has been shown to mediate the association of the virginiamycin PKS polypeptides VirA and VirFG to connect the fifth module of the synthase with the sixth.¹³ Isothermal titration calorimetry measurements with constructs containing these docking motifs showed an ~6 μ M dissociation constant, and an NMR structure of ^CDD_{VirA} and ^NDD_{VirFG} (linked by seven residues) reveals the four-helix bundle nature of the complexed DD.

Sequence inspection of the termini of *trans*-AT polypeptides thought to physically associate revealed that the 25-residue pseudosymmetric ^CDD and ^NDD mediate not only intermodular but also intramodular polypeptide disconnections. Four of these docking motif pairs are present in the macrolactin synthase, one at a KS/KR disconnection and three at DH/KR disconnections (Figure 1). Terminal domains expressed with their docking motifs were observed to form complexes stable to size-exclusion chromatography. While deletion

of ^CDDs resulted in the loss of this interaction, swapping either ^CDDs or ^NDDs resulted in the formation of the anticipated complexes. A 1.72 Å-resolution structure of an ^NDD and KR from the N-terminus of the macrolactin synthase polypeptide MlnE reveals the ^NDD to be preorganized for complex formation.

Results

Docking Bioinformatics

Bioinformatic investigation of the terminal regions of *trans*-AT PKSs that could serve as docking motifs indicates the dominant mode through which polypeptides are physically ordered.^{2,3} Most polypeptides appear to contain split domains at their termini capable of associating with a split domain on the terminus of a neighboring polypeptide to form a single, well-folded domain.¹⁴ Domain splitting within *trans*-AT PKSs seemingly occurs not only within the boundaries of the enzymatic domains but also within the recently described DD, an ~50-residue docking domain comprised of four helices.¹³

Split enzymatic domains are observable within *trans*-AT PKSs on a sequence level (Figure 2a). A KS from the misakinolide synthase is split into its catalytic body and its flanking subdomain;¹⁵ these regions form an extensive interface within an intact KS.¹⁶ A DH from an uncharacterized synthase in *Geobacter uraniireducens* is split after its first ~20 residues;² these residues contain an “HPLL” motif that fills a hydrophobic pocket within an intact DH.¹⁷ A KR from the disorazol synthase is split after its first ~10 residues;¹⁸ these residues contain a β-strand that integrates into the β-sheets of both KR subdomains.¹⁹

Split DDs are also observable within *trans*-AT PKSs on a sequence level, connecting such domains as KS to DH, KS to KR, DH to KR, ACP to KS, ACP to ER, and even ACP to the nonribosomal peptide synthetase (NRPS) condensation (C) domain (Figures 2b and 3a).²⁰ When aligned with ClustalX and represented with WebLogo, the consensus residues of the ^CDD and ^NDD become apparent (Figure 3b).^{21,22}

Docking Motif Characteristics

Both ^CDD and ^NDD are comprised of 25 residues (based on the first and last residues always present in both motifs, assuming that the N-terminal methionine is cleaved from ^NDD) and contain remarkably similar features (Figure 3). The ^CDD_{VirA}/^NDD_{VirFG} complex (PDB code 2N5D) reveals that each motif possesses two amphipathic helices connected by a four-residue linker (positions 12–15).¹³ Several residues (positions 2, 5, 6, 9, 14, 16, 19, 22, 23) contribute to the hydrophobic core of each motif (a hydrophobic residue in position 26 of ^CDD may also contribute), with the loop residue in position 14 (usually a leucine or isoleucine) helping to hydrophobically connect the two helices. The second helices of both motifs are frequently capped by the residue at position 15 (D/S/T in ^CDD, S/T in ^NDD). While ^CDDs and ^NDDs share many attributes, they also possess features that distinguish them from one another. The helices of ^NDD are more closely packed, apparently owing to the close hydrophobic interactions between residues at positions 5, 22, and 23 (position 5 in ^NDD is predominantly a β-branched isoleucine or valine, whereas in ^CDD it is predominantly a leucine or methionine). Within the DD complex, equivalent residues at

position 23 are differentially oriented, with the side chain of the ^NDD residue closer to the hydrophobic center of the domain. A principle hydrophobic contact with the residue in position 23 of ^NDD is with the residue in position 2 of ^CDD, which is commonly a bulky hydrophobic residue in ^CDD in contrast to the diverse residues located in position 2 of ^NDD. In general, the N-terminal residues of ^CDD contribute more to the interface than those of ^NDD, while the C-terminal residues of ^NDD contribute more to the interface than those of ^CDD.

To learn more about how ^CDD and ^NDD complement one another, the 28 ^CDD/^NDD pairs in Figure 3b were homology-modeled using the program SCWRL with the ^CDD_{VirA}/^NDD_{VirFG} complex (PDB 2N5D) as the template (Figure S1).²³ Since only side chain and not backbone positions are optimized by SCWRL, it provides a readout for whether the ^CDD/^NDD pairs are complementary. Few clashes occurred, and the most frequent clash from position 2 of ^NDD (e.g., an arginine that the SCWRL program attempts to fit in the hydrophobic core) is resolvable through a slight adjustment of the backbone.

Complex Formation

The seven polypeptides of the macrolactin synthase make six ordered connections (MlnB/MlnC, MlnC/MlnD, MlnD/MlnE, MlnE/MlnF, MlnF/MlnG, and MlnG/MlnH) (Figure 1). Four out of six of these connections are DD-mediated (the other two are from split type A dehydrating bimodules, which must rely on another type of docking interaction since DD motifs are not present).² To more closely study DD-mediated connections, the domains at the polypeptide termini along with their native DD motifs were cloned, expressed in *Escherichia coli* BL21(DE3), and purified. The histidine tag that facilitates purification was positioned at the N-terminus of C-terminal domains and at the C-terminus of N-terminal domains so as not to interfere with the DD motifs or their interaction with one another. MlnDH3+^CDD_{MlnB} was the only construct that could not be solubly expressed.

Each of the constructs MlnDH4+^CDD_{MlnC}, ^NDD_{MlnD}+MlnKR4, MlnKS6+^CDD_{MlnD}, ^NDD_{MlnE}+MlnKR6, MlnDH9+^CDD_{MlnF}, and ^NDD_{MlnG}+MlnKR9 gave rise to a single peak in size-exclusion chromatograms (Figures 4a and S2). The elution of MlnKS6+^CDD_{MlnD} peaked in fraction 9, MlnDH4+^CDD_{MlnC} and ^NDD_{MlnG}+MlnKR9 in fractions 11 and 12, and ^NDD_{MlnD}+MlnKR4, ^NDD_{MlnE}+MlnKR6, and ^NDD_{MlnG}+MlnKR9 in fractions 13 and 14. While the elution of DH-containing constructs before KR-containing constructs is unexpected (~40 vs ~60 kDa, respectively), this does not necessarily indicate a difference in their oligomerization states.

When cognate pairs (MlnDH4+^CDD_{MlnC} and ^NDD_{MlnD}+MlnKR4; MlnKS6+^CDD_{MlnD} and ^NDD_{MlnE}+MlnKR6; and MlnDH9+^CDD_{MlnF} and ^NDD_{MlnG}+MlnKR9) were incubated together (~50 μM each) prior to size-exclusion chromatography, a new peak with a faster elution time was observed in each case, indicating formation of a relatively stable complex (Figures 4a and S2). SDS/PAGE analysis revealed that both species are present in these complexes (readily observed through the faster migration of the KR-containing constructs). The noncognate pairs did not show any complex formation, with each construct migrating as if it had been run individually. Isothermal titration calorimetry supported these findings, with significant affinity for cognates (1.8 μM for MlnDH4+^CDD_{MlnC} and ^NDD_{MlnD}+MlnKR4,

0.8 μM for MlnKS6+^CDD_{MlnD} and ^NDD_{MlnE}+MlnKR6, and 9.0 μM for MlnDH9+^CDD_{MlnF} and ^NDD_{MlnG}+MlnKR9) and insignificant affinity for noncognates (combining MlnDH4+^CDD_{MlnC} with ^NDD_{MlnG}+MlnKR9 did not release more heat than the reference cell) (Figure S3).

To investigate whether ^CDDs are necessary for docking, they were deleted from three constructs. Without their ^CDD neighbors, constructs MlnDH4, MlnKS6, and MlnDH9 did not form complexes with ^NDD_{MlnD}+MlnKR4, ^NDD_{MlnE}+MlnKR6, or ^NDD_{MlnG}+MlnKR9, respectively (Figures 4b and S4). To investigate whether ^CDDs are sufficient for docking to their cognate ^NDDs, swapping experiments were performed. Thus, MlnDH4+^CDD_{MlnF}, MlnDH9+^CDD_{MlnC}, and MlnDH9+^CDD_{MlnD} were generated. Each of the swapped constructs docked with its anticipated partner (Figures 4c,d and S5). The construct ^NDD_{MlnC}+MlnKR9 was generated to investigate whether ^NDDs are also portable. It formed the anticipated complex with MlnDH4+^CDD_{MlnC} but not with the noncognate MlnDH9+^CDD_{MlnF} (Figure S6).

Structure of an Uncomplexed ^NDD

The 1.72 Å-resolution crystal structure of the KR from the sixth module of the macrolactin synthase (MlnKR6) and the neighboring ^NDD_{MlnE} (^NDD_{MlnE}+MlnKR6; PDB code 5D2E) reveals the structure of an uncomplexed ^NDD and shows the types of interactions that can be made between an uncomplexed DD motif and its neighboring domain (Figure 5a and Table 1). Every residue of ^NDD_{MlnE} is visible in the electron density maps, although they possess higher B-factors than those of MlnKR6 (71 vs. 37 Å²). Its two amphipathic helices and the four residues connecting them are nearly in the same conformation observed for ^NDD_{VirFG} in the ^CDD_{VirA}/^NDD_{VirFG} NMR structure, indicating a high level of structural similarity between an ^NDD in the absence of its cognate ^CDD and an ^NDD complexed with its cognate ^CDD (PDB 2N5D).¹³ ^NDD_{MlnE} makes contact with MlnKR6 principally *via* its second helix (Figure 5b). The side chain of E24 (position 18) forms a salt bridge with K295, the side chain of E27 (position 21) caps helix αG , the carbonyl of K30 (position 24) forms a hydrogen bond with the side chain of K34 (this residue, located on the linker connecting ^NDD_{MlnE} and MlnKR6, also forms a salt bridge with E473), and the carbonyl of S31 (position 25) forms a hydrogen bond with the backbone NH of K34. The side chain of L32 (position 26) interacts with hydrophobic residues on both helices of ^NDD_{MlnE}.

From the available structures, a model of the docking interaction between the MlnD and MlnE polypeptides was constructed (the ^CDD_{MlnD}/^NDD_{MlnE} homology model was superposed on the ^NDD_{MlnE}+MlnKR6 crystal structure through the alignment of ^NDD_{MlnE}; C _{α} rmsd = 1.16 Å) (Figure 5c,d). No clashes between ^NDD_{MlnD} and MlnKR6 were formed since the hydrophobic portion of ^NDD_{MlnE} that interfaces with ^CDD_{MlnD} is directed away from MlnKR6.

Discussion

In many respects, the recently discovered *trans*-AT PKSs appear quite alien from their more well-characterized cousins, the *cis*-AT PKSs. A profound difference between these systems is the nature of the disconnections made between neighboring polypeptides. The

disconnections of *cis*-AT PKSs are located neatly between modules, and the associated docking motifs are very similar to one another. In contrast, the disconnections of *trans*-AT PKSs are positioned more frequently within modules than between modules, and several classes of docking motifs exist.

Split domains have recently been shown to play a role in the ordered assembly of *trans*-AT PKS polypeptides.^{13,14} While instances of split canonical PKS domains such as KS, DH, and KR can be observed on the sequence level based on known boundaries and structures, the most common split domain is the recently discovered, four-helix bundle DD (Figures 2 and 3). With its low-micromolar dissociation constant, it is capable of helping mediate the association of neighboring polypeptides.¹³ The homology models of the ^CDD/^NDD complexes presented here suggest that, although surface residues with complementary chemistry are often in position to interact across the interface, shape complementarity within the hydrophobic core contributes most toward docking specificity (Figures 3b and S1). The small size of the complexed DD (~6 kDa) makes it a versatile universal junction for both intermodular and intramodular disconnections, since it is unlikely to disrupt the overall architecture or function of the megadalton-sized synthases containing them.

In contrast to each of the previously characterized docking motifs from modular PKSs, DD motifs are not dependent on homodimerization.¹³ The N-terminal docking motifs of *cis*-AT PKSs are dimeric coiled-coils connected to dimeric KSs (PDB codes 2HG4 and 4MZ0).^{24,25} Another class of docking motifs observed in PKS/NRPS hybrid synthases is represented by the N-terminal 60-residues of the second polypeptide of the tubulysin synthase, which, as a homodimer, is thought to provide binding sites for the ~10-residue tails of the first polypeptide.²⁶ The communication-mediating (COM) motifs of NRPSs (~25 residues following a C-terminal epimerization domain and ~15 residues preceding an N-terminal condensation domain) are monomeric, although the crystal structure of a surfactin NRPS module indicates that a portion of the condensation domain aids the N-terminal COM motif in forming a binding site for the α -helical, C-terminal COM motif.^{27–29}

At least the bacillaene *trans*-AT PKS is known to form a megacomplex.⁵ By electron microscopy, it is observable as a dense, ~100 MDa mass associated with the plasma membranes of *Bacillus subtilis* cells. The lateral interactions of the megacomplex are hypothesized to be formed by three-helix LINKS motifs that extend from the KS flanking subdomains to join neighboring polypeptide homodimers in register with one another.⁷ The lateral contacts mediated by LINKS motifs and vertical contacts mediated by DD motifs (^CDD_{PKSM} and ^NDD_{PKSN} are present in the bacillaene synthase from *B. subtilis*) provide a framework for understanding how biosynthetic “organelles” are formed through the self-assembly of their polypeptide components.

The four-helix DD may provide an evolutionary resource for *trans*-AT PKSs. The ability of DDs to mediate intramodular disconnections in addition to intermodular disconnections suggests that they can facilitate the movement of enzymatic domains into and out of modules. In this way, DDs may have helped generate the greater diversity of module types in *trans*-AT PKSs compared with *cis*-AT PKSs. A potential benefit from the insertion of a genetic element encoding ^CDD and ^NDD (as short as 160 bases) into a gene encoding a large

polypeptide is that the large polypeptide would be split into two smaller polypeptides, which may fold more efficiently. The majority of polypeptides that associate with one another *via* four-helix docking domains are encoded on neighboring genes. When they are interrupted by other genes, the inserted genes often encode enzymes that operate on polyketide intermediates bound to the genetically interrupted module (e.g., β -branching enzymes between the genes encoding ElaK/ElaO, TaiE/TaiK, VirA/VirFG).²

The level of independence possessed by DD motifs remains to be determined. They could be aided by neighboring domains and physically associate with those domains within assembled synthases. The $^{\text{NDD}}_{\text{MlnE}} + \text{MlnKR6}$ structure shows several interactions between $^{\text{NDD}}_{\text{MlnE}}$ and MlnKR6 that could help preorganize $^{\text{NDD}}_{\text{MlnE}}$ for its interaction with $^{\text{CDD}}_{\text{MlnD}}$. If DD motifs can mediate the ordered assembly of polypeptides similar to how sticky ends mediate the specific association of DNA duplexes, they will find employment as tools for chemical and synthetic biologists.^{14,30,31} DD motifs will certainly be utilized in the engineering of *trans*-AT PKSs and could even be fused to fluorescent proteins or electron microscopy tags to enable the visualization of *trans*-AT PKSs within intact cells.³² The characterization of the four-helix docking motifs has been architecturally and functionally informative, yet it is only the first of several classes of docking motif pairs within *trans*-AT PKSs to be described.

Methods

Materials

Isopropyl β -D-1-thiogalactopyranoside (IPTG) was purchased from Carbosynth. All other reagents, including DNA primers, were purchased from Sigma-Aldrich. Ni-Nitrilotriacetic acid (NTA) affinity resin was purchased from Qiagen. Protein concentration devices (Amicon Ultra-15 centrifugal filter units, 30 000 molecular weight cutoff) were purchased from Millipore.

Plasmid Construction

The constructs $\text{MlnDH4} + ^{\text{CDD}}_{\text{MlnC}}$, $^{\text{NDD}}_{\text{MlnD}} + \text{MlnKR4}$, $\text{MlnKS6} + ^{\text{CDD}}_{\text{MlnD}}$, $^{\text{NDD}}_{\text{MlnE}} + \text{MlnKR6}$, $\text{MlnDH9} + ^{\text{CDD}}_{\text{MlnF}}$, and $^{\text{NDD}}_{\text{MlnG}} + \text{MlnKR9}$ were amplified by PCR using *Bacillus amyloliquefaciens* FZB42 genomic DNA as the template (Table S1). The DNA encoding constructs from the C-terminal ends of polypeptides was ligated between the *NheI* and *XhoI* sites in pET28b, while the DNA encoding constructs from the N-terminal ends of polypeptides were ligated between the *NdeI* and *XhoI* sites of pET21a. For removing and swapping $^{\text{CDD}}$ s, the last residue of the enzymatic domain was chosen to be four residues after the last residue of the enzyme that is anticipated to be well-folded based on known crystal structures; thus, the ensuing residues containing the $^{\text{CDD}}$ s were deleted or swapped (Table S2). To append $^{\text{CDD}}$ s to enzymatic domains, the DNA encoding the $^{\text{CDD}}$ was ligated into the *XhoI* site of plasmids encoding only the enzymatic domains (insert orientation was checked by sequencing). To fuse $^{\text{NDD}}_{\text{MlnD}}$ with MlnKR9, the DNA encoding MlnKR9 was ligated between the *NheI* and *XhoI* sites of pET21a and then the DNA encoding $^{\text{NDD}}_{\text{MlnD}}$ was ligated between the *NdeI* and *NheI* sites.

Protein Expression and Purification

Plasmids were transformed into *E. coli* BL21(DE3). Cells were inoculated into 6 L of Luria–Bertani medium containing 50 mg/L kanamycin or ampicillin at 37 °C, grown to an OD₆₀₀ of 0.5, and induced with 1 mM IPTG. After 16 h at 15 °C, cells were collected by centrifugation, resuspended in lysis buffer [0.5 M NaCl and 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES) (pH 7.5)], and sonicated. After centrifugation at 20 000g for 30 min, the lysate was poured over a column of nickel-NTA resin equilibrated with lysis buffer, which was then washed with 15 mL of lysis buffer containing 10 mM imidazole and eluted with 3 mL of lysis buffer containing 150 mM imidazole. Eluted protein was flash frozen in liquid nitrogen and stored at –80 °C until needed. For crystallization screening, each construct was further purified by size-exclusion chromatography with a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES (pH 7.5).

Complex Formation Assays

Pairs of constructs were combined in an ~1:1 molar ratio (~50 μ M of each protein) in a volume of 1 mL and injected onto a Superdex 200 gel filtration column (GE Healthcare Life Sciences) equilibrated with 150 mM NaCl, 10 mM HEPES (pH 7.5) with a flow rate of 0.4 mL/min. Protein elution was monitored by absorbance at 280 nm, and fractions were collected every 0.7 mL after 8 mL had eluted. Individual domains were also run (~50 μ M in a 1 mL volume). Fractions 8–16 from each run were analyzed through Coomassie-stained SDS-PAGE gels (loading 10 μ L of each fraction per lane).

Isothermal Titration Calorimetry

An isothermal titration calorimeter (ITC200, MicroCal Inc.) was employed to measure the affinity (K_d), stoichiometry (n), and apparent enthalpy change (ΔH°) of potential docking pairs. Samples of protein and were prepared in HEPES (10 mM HEPES, 150 mM NaCl, pH 7.5). Samples were spin-filtered and loaded into the cell and the syringe. A stirring speed of 300 rpm at a temperature of 25 °C was used for the 29 injections of 1.2 μ L each lasting 1.2 s. The first injection was set to a smaller volume (0.4 μ L in 0.8 s) than the following ones to check for background dilution heats prior to the actual run. Titration experiments were performed in triplicate to show reproducibility. Control experiments confirmed that the heats of dilution caused by the proteins being titrated into buffer were negligible. The traces were analyzed using ORIGIN, version 7.0 (MicroCal, Inc.) with a model that assumed a single binding site. The heats of dilution for the sample in the cell were subtracted from the final trace before integration with respect to time. The K_d and stoichiometry values obtained from the traces were averaged across three runs and a standard deviation was calculated (Figure S3).

Crystallization and Structure Determination

Crystals of ^NDD_{MlnE}+MlnKR6 were grown by sitting vapor diffusion at 22 °C by mixing 1 μ L of protein solution (~28 mg mL⁻¹ in 150 mM NaCl, 10 mM HEPES, pH 7.5) with 1 μ L of 11% PEG3350, 0.24 M sodium malonate (pH 7.0). Crystals were briefly soaked in a cryoprotecting buffer [15% (v/v) glycerol in the reservoir solution] before being flash frozen

in liquid nitrogen. Diffraction data, collected at ALS beamline 5.0.2, were processed with iMosflm and scaled with SCALA from the CCP4 suite.^{33,34} The structure was solved by molecular replacement with PhaserMR using another KR from a *trans*-AT PKS (PDB 4J1S) as the search model.³⁵ The solution contains one monomer per asymmetric unit. The model was refined with Coot and Refmac5 (Table 1).^{36,37}

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Financial support was provided by the National Institutes of Health (Grant GM106112) and the Welch Foundation (Grant F-1712). Instrumentation and technical assistance for this work were provided by the Macromolecular Crystallography Facility at the University of Texas at Austin, with financial support from the College of Natural Sciences, the Office of the Executive Vice President and Provost, and the Institute for Cellular and Molecular Biology. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences, and the Howard Hughes Medical Institute. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

1. Keatinge-Clay AT. The structures of type I polyketide synthases. *Nat Prod Rep.* 2012; 29:1050–73. [PubMed: 22858605]
2. Piel J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat Prod Rep.* 2010; 27:996–1047. [PubMed: 20464003]
3. Helfrich EJ, Piel J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat Prod Rep.* 2016; 33:231–316. [PubMed: 26689670]
4. Laureti L, Song L, Huang S, Corre C, Leblond P, Challis GL, Aigle B. Identification of a bioactive 51-membered macrolide complex by activation of a silent polyketide synthase in *Streptomyces ambofaciens*. *Proc Natl Acad Sci U S A.* 2011; 108:6258–63. [PubMed: 21444795]
5. Straight PD, Fischbach MA, Walsh CT, Rudner DZ, Kolter R. A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc Natl Acad Sci U S A.* 2007; 104:305–10. [PubMed: 17190806]
6. Weissman KJ, Muller R. Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem.* 2008; 9:826–848. [PubMed: 18357594]
7. Gay DC, Wagner DT, Meinke JL, Zogzas CE, Gay GR, Keatinge-Clay AT. The LINKS motif zippers *trans*-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex. *J Struct Biol.* 2016; 193:196–205. [PubMed: 26724270]
8. Stinear TP, Mve-Obiang A, Small PL, Frigui W, Pryor MJ, Brosch R, Jenkin GA, Johnson PD, Davies JK, Lee RE, Adusumilli S, Garnier T, Haydock SF, Leadlay PF, Cole ST. Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci U S A.* 2004; 101:1345–9. [PubMed: 14736915]
9. Broadhurst RW, Nietlispach D, Wheatcroft MP, Leadlay PF, Weissman KJ. The structure of docking domains in modular polyketide synthases. *Chem Biol.* 2003; 10:723–731. [PubMed: 12954331]
10. Buchholz TJ, Geders TW, Bartley FE 3rd, Reynolds KA, Smith JL, Sherman DH, et al. Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem Biol.* 2009; 4:41–52. [PubMed: 19146481]
11. Gokhale RS, Khosla C. Role of linkers in communication between protein modules. *Curr Opin Chem Biol.* 2000; 4:22–27. [PubMed: 10679375]
12. Menzella HG, Reid R, Carney JR, Chandran SS, Reisinger SJ, Patel KG, Hopwood DA, Santi DV. Combinatorial polyketide biosynthesis by *de novo* design and rearrangement of modular polyketide synthase genes. *Nat Biotechnol.* 2005; 23:1171–6. [PubMed: 16116420]

13. Dorival J, Annaïval T, Risser F, Collin S, Roblin P, Jacob C, Gruez A, Chagot B, Weissman KJ. Characterization of intersubunit communication in the virginiamycin trans-acyl transferase polyketide synthase. *J Am Chem Soc.* 2016; 138:4155. [PubMed: 26982529]
14. Shekhawat SS, Ghosh I. Split-protein systems: beyond binary protein-protein interactions. *Curr Opin Chem Biol.* 2011; 15:789–97. [PubMed: 22070901]
15. Ueoka R, Uria AR, Reiter S, Mori T, Karbaum P, Peters EE, Helfrich EJ, Morinaka BI, Gugger M, Takeyama H, Matsunaga S, Piel J. Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat Chem Biol.* 2015; 11:705–12. [PubMed: 26236936]
16. Gay DC, Gay G, Axelrod AJ, Jenner M, Kohlhaas C, Kampa A, Oldham NJ, Piel J, Keatinge-Clay AT. A close look at a ketosynthase from a trans-acyltransferase modular polyketide synthase. *Structure.* 2014; 22:444–51. [PubMed: 24508341]
17. Keatinge-Clay AT. Crystal structure of the erythromycin polyketide synthase dehydratase. *J Mol Biol.* 2008; 384:941–53. [PubMed: 18952099]
18. Kopp M, Irschik H, Pradella S, Müller R. Production of the tubulin destabilizer disorazol in *Sorangium cellulosum*: biosynthetic machinery and regulatory genes. *ChemBioChem.* 2005; 6:1277–86. [PubMed: 15892181]
19. Keatinge-Clay AT, Stroud RM. The structure of a ketoreductase determines the organization of the beta-carbon processing enzymes of modular polyketide synthases. *Structure.* 2006; 14:737–48. [PubMed: 16564177]
20. Keating TA, Marshall CG, Walsh CT, Keating AE. The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol.* 2002; 9:522–526. [PubMed: 12055621]
21. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal W and Clustal X version 2.0. *Bioinformatics.* 2007; 23:2947–2948. [PubMed: 17846036]
22. Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: A sequence logo generator. *Genome Res.* 2004; 14:1188–1190. [PubMed: 15173120]
23. Canutescu AA, Shelenkov AA, Dunbrack RL Jr. A graph theory algorithm for protein side-chain prediction. *Protein Sci.* 2003; 12:2001–2014. [PubMed: 12930999]
24. Tang Y, Kim CY, Mathews II, Cane DE, Khosla C. The 2.7-Ångström crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc Natl Acad Sci U S A.* 2006; 103:11124–9. [PubMed: 16844787]
25. Whicker JR, Smaga SS, Hansen DA, Brown WC, Gerwick WH, Sherman DH, Smith JL. Cyanobacterial polyketide synthase docking domains: a tool for engineering natural product biosynthesis. *Chem Biol.* 2013; 20:1340–51. [PubMed: 24183970]
26. Richter CD, Nietlispach D, Broadhurst RW, Weissman KJ. Multienzyme docking in hybrid megasynthetases. *Nat Chem Biol.* 2008; 4:75–81. [PubMed: 18066054]
27. Hahn M, Stachelhaus T. Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc Natl Acad Sci U S A.* 2006; 103:275–280. [PubMed: 16407157]
28. Tanovic A, Samel SA, Essen LO, Marahiel MA. Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science.* 2008; 321:659–63. [PubMed: 18583577]
29. Hur GH, Meier JL, Baskin J, Codelli JA, Bertozzi CR, Marahiel MA, Burkart MD. Crosslinking studies of protein-protein interactions in nonribosomal peptide biosynthesis. *Chem Biol.* 2009; 16:372–381. [PubMed: 19345117]
30. Fletcher JM, Boyle AL, Bruning M, Bartlett GJ, Vincent TL, Zaccari NR, Armstrong CT, Bromley EH, Booth PJ, Brady RL, Thomson AR, Woolfson DN. A basis set of de novo coiled-coil peptide oligomers for rational protein design and synthetic biology. *ACS Synth Biol.* 2012; 1:240–50. [PubMed: 23651206]
31. Thompson KE, Bashor CJ, Lim WA, Keating AE. SYNZIP protein interaction toolbox: in vitro and in vivo specifications of heterospecific coiled-coil interaction domains. *ACS Synth Biol.* 2012; 1:118–29. [PubMed: 22558529]

32. Shu X, Lev-Ram V, Deerinck TJ, Qi Y, Ramko EB, Davidson MW, Jin Y, Ellisman MH, Tsien RY. A genetically encoded tag for correlated light and electron microscopy of intact cells, tissues, and organisms. *PLoS Biol.* 2011; 9:e1001041. [PubMed: 21483721]
33. Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr Sect D Biol Crystallogr.* 2011; 67:271–81. [PubMed: 21460445]
34. Winn MD, Ballard CC, Cowtan KD, Dodson EJ, Emsley P, Evans PR, Keegan RM, Krissinel EB, Leslie AG, McCoy A, McNicholas SJ, Murshudov GN, Pannu NS, Potterton EA, Powell HR, Read RJ, Vagin A, Wilson KS. Overview of the CCP4 suite and current developments. *Acta Crystallogr Sect D Biol Crystallogr.* 2011; 67:235–42. [PubMed: 21460441]
35. Piasecki SK, Zheng J, Axelrod AJ, Detelich ME, Keatinge-Clay AT. Structural and functional studies of a trans-acyltransferase polyketide assembly line enzyme that catalyzes stereoselective α - and β -ketoreduction. *Proteins Struct Funct Genet.* 2014; 82:2067–77. [PubMed: 24634061]
36. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr.* 2010; 66:486–501. [PubMed: 20383002]
37. Vagin AA, Steiner RS, Lebedev AA, Potterton L, McNicholas S, Long F, Murshudov GN. REFMAC5 dictionary: organisation of prior chemical knowledge and guidelines for its use. *Acta Crystallogr Sect D Biol Crystallogr.* 2004; 60:2184–2195. [PubMed: 15572771]

Figure 1.

Macrolactin *trans*-AT PKS, split between seven polypeptides (MlnB–MlnH). The ordered assembly of *trans*-AT PKS polypeptides is only just starting to be understood. The recently discovered four-helix docking domains (DDs) serve as universal joints at the MlnB/MlnC, MlnC/MlnD, MlnD/MlnE, and MlnF/MlnG disconnections. How the MlnE/MlnF and MlnG/MlnH disconnections that lie within split type A dehydrating bimodules are joined remains unknown. Note that each of the polypeptide disconnections is within a module, in stark contrast to *cis*-AT PKSs. The module to which an enzyme belongs is indicated with numbers.

Figure 2.

Split domains that help assemble trans-AT PKSs. (A) Canonical PKS domains, such as KS, DH, and KR, are sometimes split between polypeptides. A KS from the misakinolide synthase is divided into its catalytic body and flanking subdomain, which share a large interface within an intact KS. A DH from an uncharacterized synthase in *Geobacter uraniireducens* is divided after the “HPLL” motif, which fills a hydrophobic pocket within an intact DH. A KR from the disorazol synthase is divided after its first KR strand (β 1), which integrates into the β -sheets of both KR subdomains. (B) The four-helix DD is divided between its second and third helices. They are apparently capable of linking any two domains, either between or within modules. Regions anticipated to be well-folded based on solved structures are colored. MisC/MisD, misakinolide, *Candidatus entothionella*, AKQ22699/AKQ22698; Gura3096/Gura3095, uncharacterized, *Geobacter uraniireducens*, WP_049818939/WP_011939927; DszA/DszB, disorazole, *Sorangium cellulosum*, AAY32964/AAY32965; Diff/G, diffidin, *Bacillus amyloliquefaciens*, WP_039063294/WP_039063293; ElaJ/ElaK/ElaQ/ElaR, elansolid, *Chitinophaga sancti*, AEC04356/AEC04357/AEC04363/AEC04364; TstDEF/TstGH, thailanstatin, *Burkholderia thailandensis*, AGN11881/AGN11882; MlnB/MlnC/MlnD/MlnE/MlnF/MlnG, macrolactin, *Bacillus amyloliquefaciens*, ABS73797/ABS73798/ABS73799/ABS73800/ABS73801/ABS73802; RizD/RizE, rhizopodin, *Stigmatella aurantiaca* Sg a15, CCA89328/CCA89329; OzmH/OzmJ, oxazolomycin, *Streptomyces bingchengensis* BCW-1, ADI12770/ADI12769; SiaF/SiaG, antibiotic SIA7248, *Streptomyces sp. A7248*, AFS33449/AFS33450; VirA/VirFG, virginiamycin, *Streptomyces virginiae*, BAF50727/WP_033220283.

Figure 3. Motifs of the four-helix docking domain. (A) The complexed docking domain that joins the VirA and VirFG polypeptides in the virginiamycin synthase was solved by NMR (PDB 2N5D).¹³ Its 25-residue, pseudosymmetric ^CDD and ^NDD components are rotated ~90° out to the left and right sides for comparison. For both ^CDD and ^NDD, the hydrophobic loop residue at position 14 helps mediate the hydrophobic connection between the two amphipathic helices. (B) The ClustalX sequence alignments and WebLogo graphic reveal the similarities and differences of the docking motifs. DDs mediate the juxtaposition of many domains (e.g., KS/DH, KS/KR, DH/KR, ACP/KS, ACP/ER, and ACP/C). DifF/G, difficidin, *Bacillus amyloliquefaciens*, WP_039063294/WP_039063293; ElaJ/ElaK/ElaO/ElaQ/ElaR, elansolid, *Chitinophaga sancti*, AEC04356/ AEC04357/AEC04361/AEC04363/AEC04364; SorB/SorC, sorangicin, *Sorangium cellulosum*, ADN68477/ADN68478; TstDEF/TstGH, thailanstatin, *Burkholderia thailandensis*, AGN11881/AGN11882; MlnB/MlnC/MlnD/MlnE/MlnF/MlnG, macrolactin, *Bacillus amyloliquefaciens*, ABS73797/ABS73798/ABS73799/ABS73800/ABS73801/ABS73802; RizD/RizE, rhizopodin, *Stigmatella aurantiaca* Sg a15, CCA89328/ CCA89329; AtcE/AtcF, anthracimycin, *Streptomyces* sp. T676, CTQ34882/CTQ34883; ChiB/ChiC, chivosazol, *Sorangium cellulosum*, AAY89049/AAY89050; OzmH/OzmJ, oxazolomycin, *Streptomyces bingchenggensis* BCW-1, ADI12770/ADI12769; PtzC/PtzD, patellazoles, *Candidatus*

Endolissoclinum faulkneri, WP_025300871/WP_025300870; RhiC/RhiD, rhizoxin, *Paraburkholderia rhizoxinica*, WP_013435481/ WP_013435480; SiaF/SiaG, antibiotic SIA7248, *Streptomyces sp. A7248*, AFS33449/AFS33450; TaiE/TaiK/TaiL/TaiM, thailandamide, *Burkholderia thailandensis*, WP_060970107/WP_043296830/ WP_043295724/WP_041223765; VirA/VirFG, virginiamycin, *Streptomyces virginiae*, BAF50727/WP_033220283; BryA/BryB/BryC, bryostatin, ABM63537/ABM63527/ ABM63528; EtnH/EtnI, etnangien, *Sorangium cellulosum*, WP_012235823/ WP_044965062; Ta-1/TaO, myxovirescin, *Myxococcus xanthus*, WP_011553945/ WP_020480755; Bat1/Bat2, batumin, *Pseudomonas batumici*, KIH86004/KIH86005; KirAII/KirAIII, kirromycin, *Streptomyces collinus*, WP_020937656/WP_020937655; BaeM/BaeN, bacillaene, *Bacillus amyloliquefaciens*, WP_043867146/WP_043867147; PksM/PksN, bacillaene, *Bacillus subtilis*, WP_014476861/WP_014476862.

Figure 4.

Polypeptide association mediated by DD motifs. Each Coomassie-stained SDS/PAGE gel shows fractions 8–16 from size-exclusion chromatography of either individual constructs or pairs of constructs. (A) Each pair of constructs containing cognate DD motifs apparently formed complexes (most easily observed through the faster migration of the $N^{DD}+KR$ construct compared with the N^{DD} and KR constructs run individually). Each pair of constructs containing noncognate DD motifs did not show complex formation. (B) The removal of C^{DD} s resulted in the loss of complex formation (each construct migrated on the size-exclusion column as if it had been run individually), indicating that they are necessary in mediating an association with a polypeptide containing a cognate N^{DD} . (C) When C^{DD} s were swapped between constructs, they were sufficient to mediate the association with a polypeptide containing a cognate N^{DD} . Gels showing complex formation are boxed in red. (D) Representative size-exclusion chromatograms showing constructs not possessing matching docking motifs migrating as if they had been run individually as well as constructs possessing matching docking motifs migrating more quickly.

Figure 5.

Structure of a docking motif and KR from the N-terminus of a *trans*-AT PKS polypeptide. (A) The 1.72 Å-resolution crystal structure of NDD_{MlnE} +MlnKR6 reveals the conformation of an NDD in the absence of its cognate CDD . (B) Stereodiagram shows several contacts between NDD_{MlnE} and MlnKR6. Whether these help orient NDD_{MlnE} to complex with CDD_{MlnD} is unclear. (C) The interaction between the MlnD and MlnE polypeptides was modeled by superposing a homology model of CDD_{MlnD}/NDD_{MlnE} (using the CDD_{VirA}/NDD_{VirFG} structure as the template with the program SCWRL) on NDD_{MlnE} +MlnKR6. (D) A stereodiagram shows the overlay of NDD_{MlnE} (crystal structure, PDB 5D2E) with the backbone of NDD_{VirFG} (NMR structure, PDB 2N5D) (C_{α} rmsd = 1.16 Å).

Table 1
Parameters for the Crystal Structure of $\text{N}^{\text{DD}}\text{MlnE}+\text{MlnKR6}$

Data Collection	
space group	<i>I</i> 222
cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	92.63, 110.93, 139.31
<i>α</i> , <i>β</i> , <i>γ</i> (deg)	90.00, 90.00, 90.00
resolution (Å)	50–1.72 (1.77–1.72)
<i>R</i> _{merge}	0.063
<i>I</i> / <i>σI</i>	18.85 (2.14)
completeness (%)	89.5 (76.06)
redundancy	7.0 (5.9)
Refinement	
resolution (Å)	50–1.72 (1.77–1.72)
no. reflns	68,897 (4032)
<i>R</i> _{work}	0.183
<i>R</i> _{free}	0.216
no. atoms	4295
protein	3855
ligand/ion	67
water	373
<i>B</i> -factors (Å ²)	39.0
protein	31.0
ligand/ion	39.1
water	44.7
rms deviations	
bond lengths (Å)	0.019
bond angles (deg)	1.98