Chemistry & Biology, Vol. 10, 723–731, August, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/S1074-5521(03)00156-X

The Structure of Docking Domains in Modular Polyketide Synthases

R. William Broadhurst,[1] Daniel Nietlispach, Michael P. Wheatcroft, Peter F. Leadlay, and Kira J. Weissman[1,*]
Department of Biochemistry
University of Cambridge
80 Tennis Court Road
Cambridge CB2 1GA
United Kingdom

Summary

Polyketides from actinomycete bacteria provide the basis for many valuable medicines, so engineering genes for their biosynthesis to produce variant molecules holds promise for drug discovery. The modular polyketide synthases are particularly amenable to this approach, because each cycle of chain extension is catalyzed by a different module of enzymes, and the modules are arranged within giant multienzyme subunits in the order in which they act. Protein-protein interactions between terminal docking domains of successive multienzymes promote their correct positioning within the assembly line, but because the overall complex is not stable in vitro, the key interactions have not been identified. We present here the NMR solution structure of a 120 residue polypeptide representing a typical pair of such domains, fused at their respective C and N termini: it adopts a stable dimeric structure which reveals the detailed role of these (predominantly helical) domains in docking and dimerization by modular polyketide synthases.

Introduction

Medically important complex polyketides are biosynthesized on giant polyketide synthase (PKS) multienzymes [1] which function as molecular assembly lines [1–4]. Each extension module contains a ketosynthase (KS) domain which catalyzes C-C bond formation, an acyl carrier protein (ACP) domain, and appropriate additional domains which introduce the extender unit and ensure the correct degree of reduction of the resulting intermediate. Typical PKS subunits are tightly homodimeric [5, 6] and contain between one and six modules each [7]. They are thought to associate with other multienzyme subunits through contacts at their C and N termini to form the overall PKS complexes [6, 8]. For example, the 6-deoxyerythronolide B synthase (DEBS) which assembles the polyketide core of erythromycin A contains three multienzyme subunits DEBS 1, DEBS 2, and DEBS 3 each housing two extension modules (Figure 1A) [1, 2, 9]. Biosynthesis of the full-length chain therefore requires two intermodular transfers between ACP and KS domains that are also interprotein transfers (for example, between the C terminus of DEBS 1 and the N terminus of DEBS 2) (Figure 1A).

Recognition that the N termini of PKS multienzymes that house extender modules (i.e., the regions N-terminal of the KS domain) contain regularities in their amino acid sequence typical of amphipathic parallel α-helical coiled coils [10], led to the proposal that these N termini are involved in specific coiled-coil interactions that stabilize PKS homodimeric assemblies [7]. More recent studies [11] have highlighted the potential role of these regions as "linkers" interacting with partner "linker" regions at the extreme C termini of the previous PKS multienzyme. These linker regions are referred to here as "docking domains," given that they adopt a specific three-dimensional fold, as discussed below. Khosla and colleagues have reported that docking domain partners can be substituted by other such partners without impairing biological function of the PKS [12, 13], and that they can also mediate acyl chain transfer between some domains and modules that do not normally cooperate with each other [14, 15]. Furthermore, they have proposed that intersubunit protein-protein recognition is mediated by interactions between helices [11].

Although the interface between successive multienzymes likely also involves ACP and KS domains [15, 16], it is clear that the intermolecular docking-domain interaction is central to an understanding of the structural basis for discrimination between potential partners and, therefore, to attempts to improve hybrid PKSs. However, it has not been obvious how to study this interaction directly, because PKS multienzymes bind even their correct partners weakly, at least in vitro [6, 17, 18]. Cysteine cross-linking has recently been used [19], but there is no guarantee that such cross-linked structures will reflect the native complex. A new approach was suggested by previous work showing that genetic fusion of subunits of a dimeric protein substantially enhanced its stability [20]. Crucially, there is direct evidence that polyketide synthase activity is preserved in vivo after fusion of the termini of PKS multienzymes [21, 22] (C. Olano et al., submitted).

Results

Expression of Docking Domains

Given the results of the functional studies, we chose to investigate polypeptides representing the fused docking domains of DEBS 1 and DEBS 2 and of DEBS 2 and DEBS 3, respectively, as well as four polypeptides representing the native docking domains as controls (unfused C termini of DEBS 1 and DEBS 2 and unfused N termini of DEBS 2 and DEBS 3). These docking domains are typical of those found in a large number of modular PKS multienzymes, some of which are shown in Figure 2. This analysis supports the presence of three conserved helical regions in the C-terminal docking domains (typically comprising 80–100 amino acid residues C-terminal of the ACP domain, Figure 2A) and a further conserved

*Correspondence: kjw21@cus.cam.ac.uk
[1] These authors contributed equally to this work.

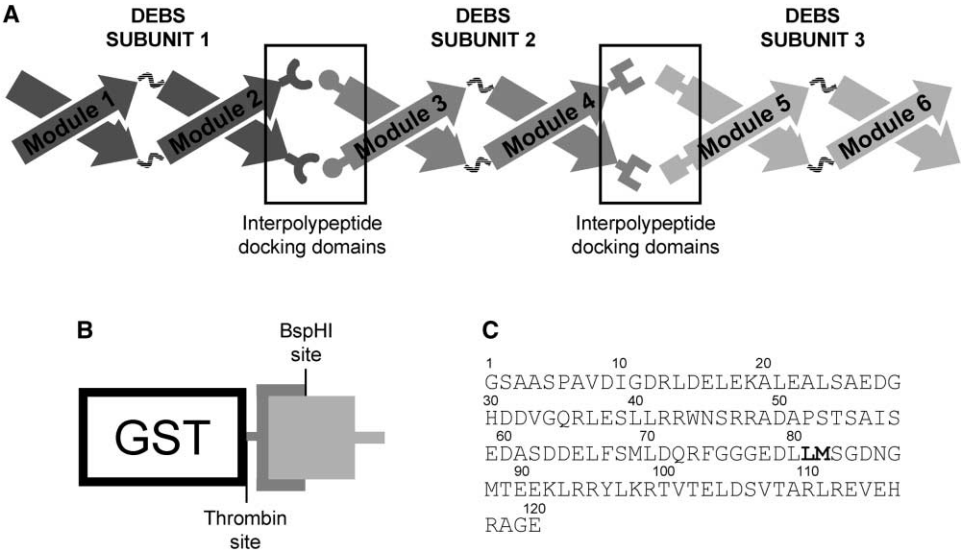

Figure 1. Schematic Organization of the Erythromycin-Producing Polyketide Synthase

(A) The erythromycin PKS, 6-deoxyerythronolide B synthase (DEBS), comprises six extension modules distributed between three multienzyme subunits. Each subunit is homodimeric, with the individual polypeptides twisted around each other in a "double-helical" structure. Protein-protein interactions are mediated, in part at least, by "docking domains" (typically 30–90 residues) at the ends of the subunits.
(B) The C-terminal docking domain from DEBS 2 was fused to the N-terminal docking domain of DEBS 3 through an engineered BspHI site. The resulting Dock 2-3 protein was expressed as a C-terminal glutathione S-transferase fusion, and the GST removed by limited proteolysis.
(C) Sequence of Dock 2-3 when cleaved from GST. LM (in bold) indicates the introduced BspHI restriction site.

helical region in the N-terminal partner domains (generally 30–40 residues N-terminal of the KS domain, Figure 2B).

The docking domain polypeptides were cloned into pGEX4T-3 for expression as C-terminal translational fusions with glutathione S-transferase (GST). Domain boundaries were chosen to include regions from the end of the highly conserved ACP domain of the upstream module to the start of the highly conserved KS domain of the downstream module. The six docking domain constructs were expressed in *E. coli* and purified by affinity chromatography using glutathione agarose.

All six GST-fused proteins were obtained in good yield, but thrombolytic cleavage to remove GST revealed that unpartnered docking domains were generally more sensitive to proteolytic degradation than their fused "C+N" docking domain counterparts. For example, when the unpartnered C terminus of DEBS 2 was cleaved from GST, it was cut at an internal site 23 residues from the end to give a protein of 5.9 kDa (mass determined by liquid chromatography/mass spectrometry) instead of the expected 8.5 kDa (Figure 3A). However, when the C terminus of DEBS 2 is fused to the N terminus of DEBS 3, this site is protected and the protein can be isolated intact (Figure 3B). At this stage, we selected the fused docking domains of DEBS 2 and 3 (termed Dock 2-3) for further study (Figures 1B and 1C) because Dock 2-3 was the most stable protein under the thrombolytic conditions. Dock 2-3 was purified to homogeneity using anion exchange chromatography followed by gel filtration.

Analytical ultracentrifugation at equilibrium yielded a molecular weight for Dock 2-3 of 29,920 Da (calculated: 26,740 Da) [6]. Although this value indicates the presence of dimer and possibly minor quantities of higher order structures, there was no evidence by NMR for aggregation of Dock 2-3, when the protein was present at an approximately 20-fold higher concentration. Circular dichroism analysis of Dock 2-3 showed that it was highly α-helical in character, and stable to thermal denaturation (T_m is 56°C). This evidence that Dock 2-3 had a well-folded homodimeric structure encouraged us to investigate its solution structure by NMR.

Analysis of Dock 2-3 by NMR

In [^{1}H, ^{15}N]-HSQC spectra of Dock 2-3 each residue contributes a single resonance, suggesting that the dimeric assembly is symmetrical with a 2-fold rotational axis. Elements of secondary structure were identified using H^α, C^α, C^β, and C' chemical shifts and patterns of short and medium range nuclear Overhauser effects (NOEs). Consistent with sequence-based predictions (Figure 2) residues 1–80, which mimic the C terminus of DEBS 2, contain three α helices (1: residues 11–24; 2: 31–49; and 3: 64–75), while residues 83–120, which mimic the DEBS 3 N terminus, contain a single longer α helix (4: residues 89–115). A ^{13}C-separated NOESY experiment revealed numerous long-range NOE connections between helices 1 and 2 (group A) and between helices 3 and 4 (group B), but no contacts between these two groups. The ^{15}N relaxation properties of backbone amide sites demonstrated that the linker between helices 2 and 3 is highly dynamic and that the apparent overall rotational correlation times of residues in the two groups differ significantly (10.1 $\pm$ 0.7 ns for group A; 11.9 $\pm$ 0.7 ns for group B, data not shown). These results indicate that Dock 2-3 consists of two dimeric structured domains (A and B) that undergo independent rotational diffusion

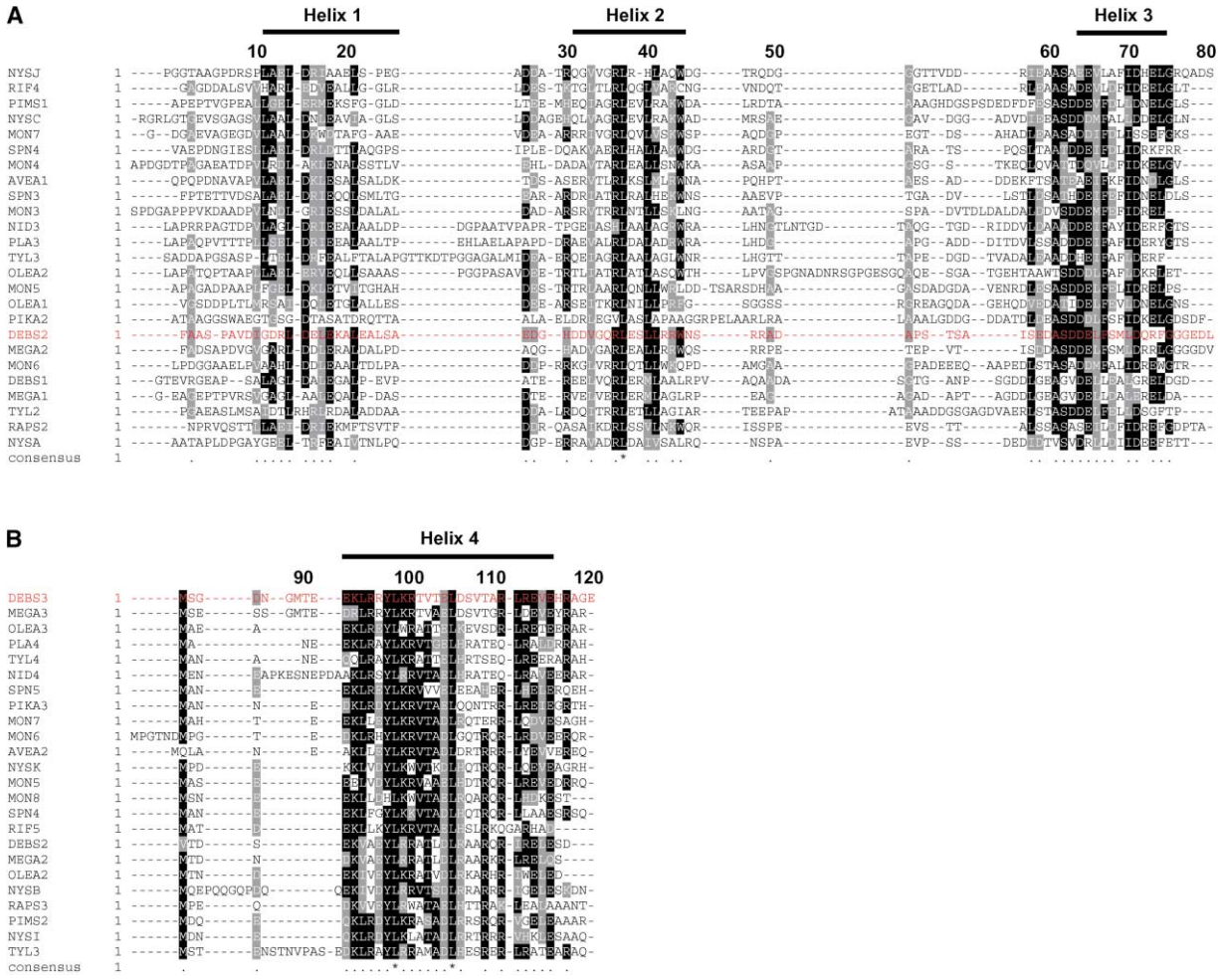

Figure 2. Sequence Alignments of Docking Domains from Polyketide Synthases

(A) Multiple alignment of C-terminal docking domains (the end of DEBS 2 is shown in red) demonstrates that three distinct regions of homology are shared among a wide range of PKSs (• indicates similarity and * identity). These regions coincide with the extent of α-helical regions as determined by NMR (indicated by overhead bars). The numbering corresponds to residues in Dock 2-3.

(B) Multiple alignment of the N-terminal docking domain partners of the C termini shown in (A) (the beginning of DEBS 3 is shown in red). AVEA, avermectin; PIKA, pikromycin; MON, monensin; SPN, spinosyn; NID, niddamycin; PLA, platenolide; TYL, tylosin; OLEA, oleandomycin; PIMS, pimiracin; NYS, nystatin; RIF, rifamycin; DEBS, erythromycin; and MEGA, megalomicin, in each case followed by the subunit number. This figure was produced using Boxshade 3.21.

and are connected by flexible tethers. Solution structures of the A and B domains (residues 1–60 and 61–120, respectively) were therefore calculated separately using NOE distance restraints, ϕ and ψ dihedral angle restraints derived by TALOS from N, H^{α}, C^{α}, C^{β}, and C′ chemical shifts [23] and hydrogen bond restraints in the α-helical regions, as detailed in Table 1. Figure 4 shows that the A and B domain ensembles are well defined in regions of the protein backbone that contain regular secondary structure, but poorly restrained elsewhere.

Overall Structure of Dock 2-3

Domain A (Figures 4A and 4B) contains an unusual intertwined four α helix bundle formed by helices 1, 2, 1′, and 2′. Within each monomer, the two helices are connected by a short loop (residues 25–30). The crossing angles between helices 1 and 1′ (20.8°) and between 2 and 2′ (20.4°) are very close to the ideal value of 20° predicted by the classic "ridges in grooves" model for packing within four-helix bundles [24]. A similar fold acts as a dimerization motif in the diabetes-associated transcriptional activator hepatocyte nuclear factor-1α (HNF-1α) [25]. We therefore propose that this portion of the docking domain structure also operates as a dimerization element, stabilizing DEBS 2 at its C terminus [7].

The A and B domains are separated by a long, highly mobile loop of 14 residues, which is poorly defined in solution. This loop varies in length among PKS docking domains and shows only low sequence conservation (Figure 2). The B domain comprises helices 3 and 3′, 4, and 4′. The later two helices, which correspond to the N terminus of DEBS 3, form seven turns of a parallel coiled-coil dimer (Figures 4C and 4D). This coiled coil could mediate subunit dimerization at the N terminus of the PKS homodimer as predicted previously [7]. Helices 3 and 3′ bind to opposite sides of this coiled coil, con-

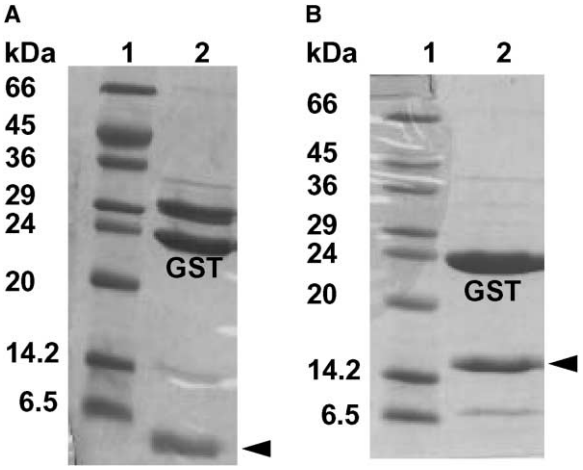

Figure 3. SDS-PAGE Analysis of Docking Domains

(A) Thrombin-catalyzed cleavage of the C-terminal docking domain of DEBS 2 fused to GST (lane 2) yields not the intact domain with an expected MW 8.46 kDa, but instead, a smaller band at 5.90 kDa (indicated by the arrow), corresponding to the first 55 residues of the 78 residue domain. (Lane 1, molecular weight markers).

(B) Limited proteolysis with thrombin of Dock 2-3 (lane 2) yields primarily the intact protein (expected MW 13.2 kDa), indicated by the arrow.

tacting both helices 4 and 4′ and forming a parallel four-helix bundle; a similar topology is seen in the Myc family of basic/helix-loop-helix/zipper transcription factor dimers [24]. The crossing angles between helices 3 and 4 and 3 and 4′ are 54.7° and 37.1°, respectively. The point of fusion between docking domain partners lies in the middle of the loop between helices 3 and 4 (Figure 4D), in an unstructured region of the protein, confirming the functional evidence [21, 22] (C. Olano et al., submitted) that fusion is unlikely to have introduced significant distortion into the structure. A schematic model for the overall structure of Dock 2-3 is shown in Figure 5.

The Dimerization Interfaces

The hydrophobic core of the "X-type" four-helix bundle [25] (domain A) is formed by seven leucine side chains (residues 14, 17, 21, 24, 37, 40, and 41), as well as by I10, A20, V33, and W44 (Figure 1C). Residues with hydrophobic side chains are generally well conserved at these positions in docking domains (Figure 2A), which suggests that all A domains of this type adopt the same fold. The Trp residue at the end of helix 2 serves to cap the hydrophobic core [26]. The side chains of R13 and E16′ (helices 1 and 1′) and E22 and R47′ (helices 1 and 2′) can form surface salt bridges which may make a favorable contribution to dimerization [27]. Sequence alignments (Figure 2A) show that in contrast to DEBS 2, the residue at position 13 is typically negatively charged (E or D) and that position 16 is usually occupied by R; in most cases, a salt bridge is not predicted to occur between residues 22 and 47′. These observations imply that the hydrophobic interactions are the most significant for dimerization.

Table 1. Experimental Restraints and Structural Statistics

	Residues 1–60		Residues 61–120	
Number of NOE restraints				
Intramolecular and unambiguous	398		402	
Intermolecular and unambiguous	101		91	
Subunit ambiguous	1212		1113	
Number of hydrogen bond restraints	42		40	
Number of dihedral angle restraints	60		78	
	<SA>	<SA>$_c$	<SA>	<SA>$_c$
Precision of coordinates				
Rmsd of backbone atoms (Å)	0.23 ± 0.06[a]	0.17[a]	0.33 ± 0.14[b]	0.20[b]
Rmsd of all heavy atoms (Å)	0.80 ± 0.08[a]	0.77[a]	0.87 ± 0.11[b]	0.73[b]
Rms deviations from experimental restraints				
NOE distances (Å)	0.089 ± 0.002	0.087	0.079 ± 0.001	0.079
Dihedral angles (°)	0.33 ± 0.03	0.30	0.23 ± 0.06	0.26
Rms deviations from idealized geometry				
Bonds (Å)	0.0053 ± 0.0002	0.0053	0.0037 ± 0.0001	0.0037
Angles (°)	0.57 ± 0.01	0.57	0.49 ± 0.01	0.49
Impropers (°)	0.70 ± 0.01	0.69	0.54 ± 0.01	0.51
Final energy $E_{L\text{-}J}$ (kJ/mol)[c]	−100.1 ± 42.2	−163.0	−115.3 ± 48.6	−130.8
Location of residues in Ramachandran analysis				
Most favored regions	77.0%[d]	81.8%[d]	96.0%[e]	98.8%[e]
Additionally allowed regions	17.8%[d]	15.9%[d]	3.1%[e]	1.2%[e]
Generously allowed regions	2.7%[d]	0.0%[d]	0.9%[e]	0.0%[e]
Disallowed regions	2.6%[d]	2.3%[d]	0.0%[e]	0.0%[e]

<SA> is the average root-mean-square (rms) deviation for the ensemble; <SA>$_c$ is the value for the structure that is closest to the mean. Rmsd, rms deviation. Confidence intervals are SD.

[a] Computed over residues 6–50.

[b] Computed over residues 64–75 and 89–115.

[c] The Lennard-Jones potential was not used at any stage in the refinement.

[d] Computed over residues 4–52.

[e] Computed over residues 62–77 and 87–117.

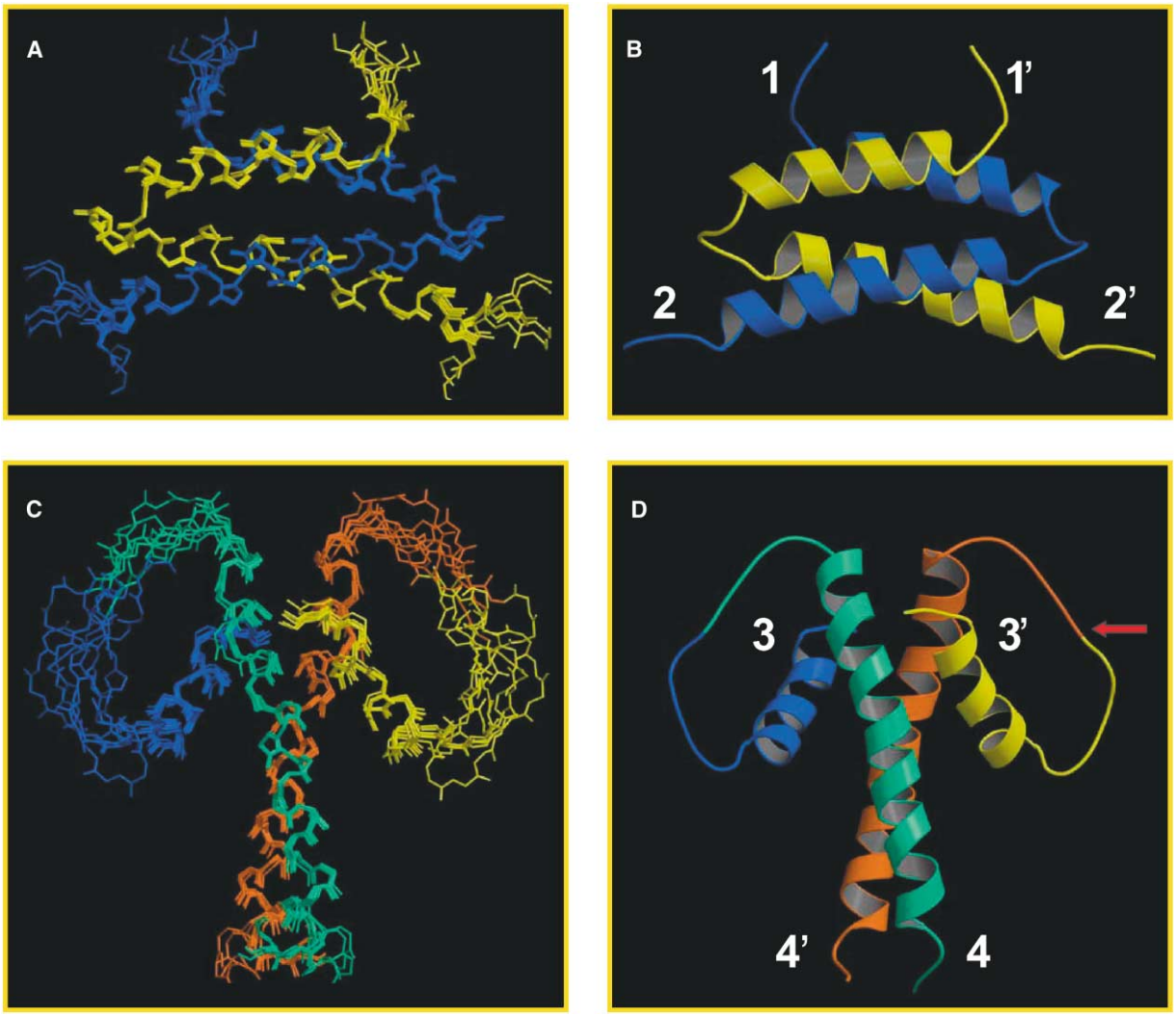

Figure 4. Solution Structure of Dock 2-3

(A) Overlay of the backbone (C^{α} trace) of domain A (residues 4–52). Domain A forms an intertwined, antiparallel four-helix bundle which stabilizes the DEBS 2 homodimer.

(B) Representation of the structure closest to the mean in the same orientation as that in (A).

(C) Overlay of the backbone (C^{α} trace) of domain B (residues 61–118). The third helix of the C-terminal docking domain of DEBS 2 docks against the coiled-coil formed by the N-terminal docking domain of DEBS 3, creating a parallel four-helix bundle. This docking interaction mediates, in part, the critical protein-protein recognition between the DEBS 2 and 3 subunits.

(D) Representation of the structure closest to the mean in the same orientation as that in (C). A red arrow marks the point of fusion.

In domain B, the interface of the coiled coil is formed by four leucine side chains (93, 97, 104, and 111), a threonine (100), and two valines (107 and 114). Of these, the leucines are most highly conserved among docking domains (Figure 2B), the residue at position 100 is usually A or V, and those at 107 and 114 are typically hydrophobic. These strong sequence similarities suggest that the coiled-coil dimerization motif is a conserved structural element among N-terminal docking domains.

The structure of the coiled coil also allows formation of an intrahelical i to i+3 salt bridge between residues R110 and E113. This ion pair may stabilize a segment of the helix that triggers coiled-coil formation, as has been found for the yeast transcriptional factor GCN4 [28, 29]. It is consistent with this idea that reversal of the charge of the corresponding arginine in the N terminus of module 3 in DEBS 2, through mutagenesis to glutamate, was recently reported to reduce the activity of the downstream module [19].

The Docking Interface

The interaction between the C-terminal docking domain of DEBS 2 and the N-terminal docking domain of DEBS 3 is limited to domain B. The primary determinant of docking is a set of conserved hydrophobic interactions between helix 3 (or 3′) and the coiled-coil created by helices 4 and 4′, which together form the core of the four-helix bundle (Figure 6A); the total surface area buried by the docking interaction is 1100 ± 80 $Å^2$, divided equally between the two sites (calculated using NACCESS [30]).

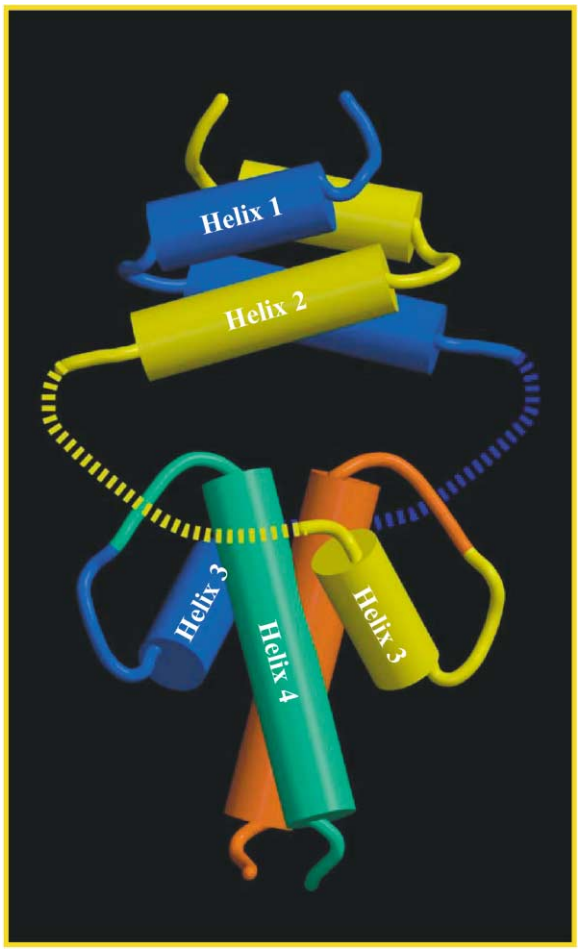

Figure 5. Structure of Dock 2-3

Schematic diagram of the structure of Dock 2-3, where α helices have been represented by cylinders and loops by lines. Although NMR analysis shows that the overall structure of Dock 2-3 is symmetrical, it does not provide any information about the relative orientations of domains A and B. The representation shown here is only one possible model for the overall structure of Dock 2-3.

The interface between helix 3 and the coiled coil is composed of F67, L70, and F74 on helix 3 and Y96, L′97, T100, V′101, and L′104 on helices 4 and 4′. Positions 67 and 74 are either F or L in many docking domains, while position 70 is usually I or L. In helix 4, Y96, L97, and L104 are very highly conserved, while position 100 is usually hydrophobic (A or V) and 101 is generally T. These hydrophobic interactions therefore seem to be a shared component in the docking between modules in many different PKSs.

In addition to the hydrophobic interface, two partially buried salt bridges located at the ends of helices 3 and 3′ may play a role in stabilizing the docking interaction (Figure 6B). The first involves D64 on helix 3 which is in range of K92 on helix 4 (and again for helices 3′ and 4′). The charges of side chains at both sites are highly conserved in this group of docking domains (Figure 2).

The second specific interaction is between the side chains of amino acids 73 and 105. In the case of DEBS, this salt bridge appears to play a critical role in discrimination among the subunits. In the DEBS 2-3 docking interaction, R73 is matched with D105, while in the corresponding interaction between DEBS 1 and 2, E73 would be matched with R105, with R108 possibly making an additional salt bridge. Clearly, misdocking of DEBS 2 against DEBS 2, or DEBS 1 against DEBS 3, would be disfavored by the repulsive ionic interactions that would result at these positions. Fos-Jun heterodimer formation [31] provides a well-documented precedent that such destabilizing interactions between partially buried residues can mediate specific oligomerization.

Discussion

The structure of Dock 2-3 is consistent with the idea that all the PKS docking domains of Figure 2 adopt a very similar fold when they form specific complexes with their cognate partners. However, a significant minority of PKS docking domains show lower sequence similarity to those in Figure 2 at critical residues (J. Garcia-Bernardo, S. Kent, and K.J.W., unpublished data), and it remains to be determined by experiment whether these adopt a different structure.

The structure reveals that the docking domains as defined here appear to play important roles both in the docking of PKS subunits and in the stabilization of PKS homodimers, and also shows that both sets of protein-protein interactions involve interhelical contacts. Within the DEBS docking domains, individual amino acids are suitably positioned to create unfavorable interactions between like charges if misdocking of PKS multienzyme subunits occurs. This could account, at least in part, for the observed specificity of acyl transfer between DEBS multienzymes in vivo which allows the production of (essentially) a single polyketide product. However, for many of the docking domain partners shown in Figure 2, the key residues are identical within several multienzymes of the same PKS and so an analogous mechanism cannot be invoked to control misdocking. In these cases, additional protein-protein interactions involving ACP and/or KS surface residues are highly likely to be involved in discriminating between the correct and the incorrect docking partners.

A recent study used the results of engineered cysteine cross-linking [19] to propose that docking in modular PKS multienzymes involves an antiparallel heterodimeric arrangement of a single helix from the C-terminal domain with a single helix from the N-terminal domain. No evidence was adduced for homodimerization of the docking elements. Our structure, however, reveals that eight helices are required and that the docking domains clearly also play a role in stabilizing the dimeric PKS structure [6]. The authors also proposed, on the basis of their model, that a charged residue should participate in a key docking interaction, but mutation of this site had no effect on chain transfer. In fact, the mutated residue corresponds to a surface-exposed site in the A domain (residue 38) and so is not directly involved in docking. The experimentally determined structure of the DEBS Dock 2-3 protein presented here should provide a clearer basis for interpreting future mutagenesis and engineering experiments that aim to analyze (and opti-

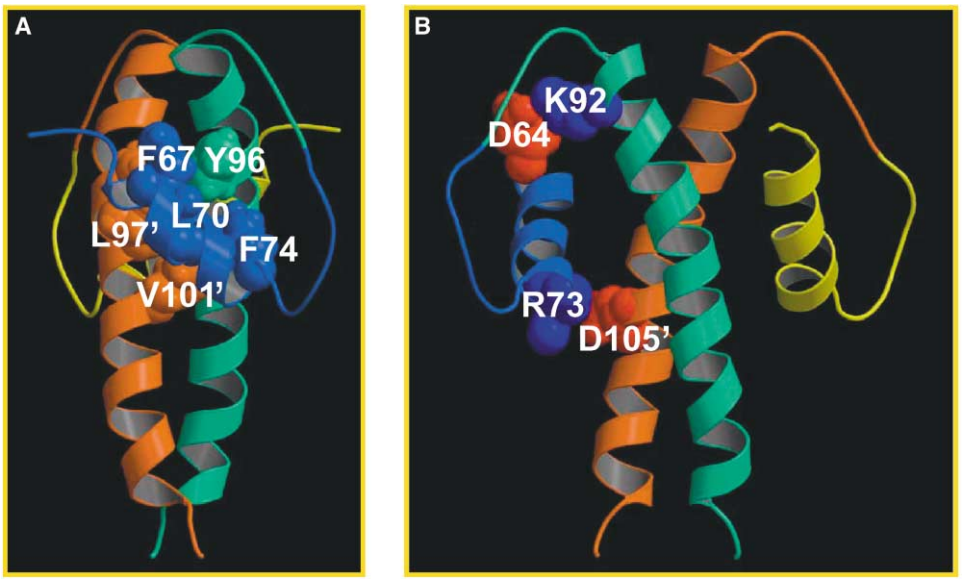

Figure 6. Residues Involved in Docking between DEBS 2 and DEBS 3

(A) The parallel four-helix bundle is held together by a series of hydrophobic interactions between helix 3 and 3′ and the coiled coil formed by helices 4 and 4′.

(B) Partially buried salt bridges at the ends of helices 3 and 3′ may play a role in determining the specificity of docking in the DEBS system.

mize) the protein-protein interfaces between successive PKS multienzymes.

The C- and N-terminal docking domains together are functionally equivalent [6] to the *intra*protein linkers of only 20–30 residues which mediate intermodular transfer within the same PKS multienzymes (for example, between DEBS modules 1 and 2). It remains unclear in either case how ACP and KS domains are brought close enough together to permit direct acyl transfer across the modular interface. The flexible linker between helices 2 and 3 of Dock 2-3 suggests that docking domains may be highly mobile elements within the PKS, which directly promote chain transfer [11]. Further structural studies of fused PKS docking domains (including portions of adjacent ACP or KS domains) are likely to provide additional insight into these key interface interactions.

Significance

The modular polyketide synthases (PKSs) are responsible for the biosynthesis of a large proportion of clinically important drugs. The growing polyketide chain is transferred between correctly ordered modules of fatty acid-synthase-related activities, distributed between three or more giant, homodimeric subunits. One of the most crucial yet least understood aspects of catalysis by these "assembly-line" multienzymes is the structural basis for the correct end-to-end "docking" between the subunits, because the interaction between the docking domains of even correct PKS partners is weak, at least in vitro. We report here on a structural model for this docking domain complex, obtained by fusing the docking domain partners together via their respective C and N termini. We have determined the NMR solution structure of one such "docking domain" fusion protein, a model for the interaction between DEBS 2 and DEBS 3 in the erythromycin polyketide synthase.

The structure contains two separate four α helix bundles with different topologies, which together mediate not only specific docking interactions but also promote dimerization of each homodimer. Sequence alignment of large numbers of docking domains from other PKSs makes it likely that they adopt similar three-dimensional structures. At least for DEBS, individual amino acid sidechains have been identified which might contribute to the destabilization of misdocked partner subunits and thus influence the specificity of polyketide chain growth. The experimental approach reported here, which had been validated previously by functional studies on fused DEBS proteins, should be applicable to a wide range of polyketide synthases, nonribosomal peptide synthetases, and mixed systems. Our results also suggest new strategies for optimizing intermodular transfer in the biosynthesis of novel polyketides by hybrid PKSs.

Experimental Procedures

Design of Expression Constructs

The unfused docking domains were amplified by PCR as BamHI-EcoRI fragments from plasmid pIB023 (I.U. Böhm, personal communication) containing the genes for DEBS 1–3, as follows: the C terminus of DEBS 1 (83 residues) (primers: 5′-ATA TA*G GAT CCA* CCG AGG TCC GGG GG-3′ [forward] and 5′-ATT C*GA ATT C*TC AAT CGC CGT CGA GC-3′ [reverse]), the N terminus of DEBS 2 (32 residues) (primers: 5′-ATA TA*G GAT CCA* CTG ACA GCG AGA AG-3′ [forward] and 5′-ATT C*GA ATT C*TC AGT CGG ATT CCA GC-3′ [reverse]), the C terminus of DEBS 2 (78 residues) (primers: 5′-ATA TA*G GAT CCG* CGG CCT CAC CGG CG-3′ [forward] and 5′-ATT C*GA ATT C*TC ACA GGT CCT CTC CCC C-3′ [reverse]), and the N terminus of DEBS 3 (38 residues) (primers: 5′-ATA TA*G GAT CCA*

GCG GTG ACA ACG GCA TGA-3′ [forward] and 5′-ATT C*GA ATT C*TC ACT CAC CGG CCC GGT GC-3′ [reverse]). The C terminus of DEBS 1 fused to the N terminus of DEBS 2 was amplified by PCR as a BamHI-EcoRI fragment from plasmid pCMS32 (primers: 5′-ATA TA*G GAT CC*A CCG AGG TCC GGG GG-3′ [forward] and 5′-ATT C*GA ATT C*TC AGT CGG ATT CCA GC-3′ [reverse]), and the C terminus of DEBS 2 fused to the N terminus of DEBS 3 was similarly amplified from plasmid pCMS50 (primers: 5′-ATA TA*G GAT CC*G CGG CCT CAC CGG CG-3′ [forward] and 5′-ATT C*GA ATT C*TC ACT CAC CGG CCC GGT GC-3′ [reverse]) [23]. In both pCMS32 and pCMS50, the docking domains had been joined using an engineered BspHI site. The N-terminal methionine was not included in the sequences of the N-terminal docking domains, as it is known to be removed in vivo from DEBS proteins [17]. All six genes were cloned into BamHI/EcoRI-digested pGEX4T-3 (Pharmacia) and the DNA sequences confirmed by sequencing.

Expression and Purification of GST-Fusion Proteins

The six docking domain constructs as their GST-fusion proteins were expressed in *E. coli* BL21-CodonPlus (RP) (Stratagene) in LB medium; protein expression was induced at an A_{600} of 0.8, and the cultures were grown for a further 5 hr at 37°C. The cells were lysed by sonication and the cell debris removed by centrifugation at 4°C. Glutathione sepharose beads (Sigma) equilibrated with PBS buffer were added to the lysate and incubated for 45 min at 4°C with agitation. The beads were then washed with copious PBS. The GST fusion proteins were eluted with 10 mM glutathione in 50 mM Tris-HCl (pH 8.2) containing 10% glycerol.

Preparation of Labeled Dock 2-3

Dock 2-3 was expressed and purified as described above, but the cells were grown in ^{15}N- or ^{13}C,^{15}N-labeled rich growth medium for *E. coli* (OD2) (Silantes). The GST was removed by digestion with thrombin (50 U mg^{-1}, 5 hr, 21°C) in 15–20 mM Tris-HCl, 300 mM NaCl, and 5 mM $CaCl_2$, and the reaction quenched by the addition of proteolysis inhibitors (EDTA, protease inhibitor cocktail [Roche] and 4-(2-aminoethyl)-benzenesulfonyl fluoride hydrochloride [AEBSF]).

Purification of Dock 2-3

Dock 2-3 was purified on a HiTrap Q HP ion exchange column: the thrombolytic cleavage reaction was applied to the column in 14% buffer B, and Dock 2-3 eluted using a gradient of 14%–80% buffer B (buffer A: 10 mM HCl, 20% glycerol, 1 mM DTT, 1 mM EDTA, pH adjusted with imidazole to 6.8; buffer B: as A, but with 1 M NaCl). Dock 2-3 was then purified to homogeneity by gel filtration on a Superdex 75 (16/60 HiLoad) column (Pharmacia) in 100 mM NaH_2PO_4 (pH 6.5). The protein was then spin concentrated to ~1 mM (Millipore Ultrafree or Amicon Ultra, 5 kDa MW cut off) and sodium azide and proteolysis inhibitors were added for long-term storage (AEBSF, protease inhibitor cocktail [Roche]).

Preparation of [^{12}C, ^{14}N], [^{13}C, ^{15}N]-Mixed-Labeled Dock 2-3

Dock 2-3 was expressed in both LB and Silantes ^{13}C,^{15}N-labeled media, and the labeled and unlabeled dimers respectively purified to homogeneity in 100 mM NaH_2PO_4, pH 6.5, as described above. Equimolar amounts of labeled and unlabeled materials were combined (2 mg ml^{-1} protein), heated to 90°C for 10 min, and cooled slowly to room temperature. The mixture was then concentrated to ~1 mM and inhibitors of proteolysis added.

Liquid Chromatography-Mass Spectrometry Analysis

The fusion of GST to the DEBS 2 C terminus was subjected to thrombolytic cleavage (20 U mg^{-1}, 1 hr, 23°C) in 15 mM Tris-HCl, 300 mM NaCl, and 5 mM $CaCl_2$. A sample was analyzed by SDS-PAGE while the remainder was analyzed by liquid chromatography-mass spectrometry (C18 column, 5%–75% acetonitrile/0.1% trifluoroacetic acid over 35 min) on a ThermoFinnigan LCQ.

Circular Dichroism

CD spectra and denaturation profiles of Dock 2-3 were collected on an Aviv Circular Dichroism Spectrometer Model 215 at protein concentrations of 0.2 and 2 mg ml^{-1} (2 and 0.2 mm path length quartz cuvette, respectively). Spectra were acquired between 250 and 185 nm with a 0.5 nm bandwith. For thermal stability experiments, the CD signal at 222 nm was monitored at 1°C intervals from 4°C to 95°C. Upon completion of the folding transition, the temperature was reduced to monitor refolding of the protein.

Equilibrium Ultracentrifugation

Sedimentation equilibrium experiments on Dock 2-3 were performed using a Beckman Optima XLi analytical centrifuge equipped with absorbance optics, an An 60Ti rotor, and three sample cells each containing a two-channel carbon-filled epon centerpiece. Samples were run at 20°C at 15,000, 20,000, or 25,000 rpm and reached equilibrium after approximately 16 hr. An average of five final scans was taken of each cell at 280 nm, with a step size of 0.001 cm (using the step mode). Samples (100 μl, protein concentration from 0.3–1.0 mg ml^{-1}) were centrifuged against buffer blanks. Solvent density ($\rho = 1.00875$) was determined using the SEDNTERP program. The partial specific volume of the protein ($v_{25} = 0.7174$) was calculated as described previously and adjusted to the correct temperature ($v_{20} = 0.7153$) [32]. Data were fitted to models for a single homogeneous species or for an oligomer of a single species.

NMR Spectroscopy

Spectra were recorded at 25°C on Bruker DRX500 and DRX800 spectrometers. The protein backbone and side-chain resonances of Dock 2-3 were assigned by standard triple-resonance NMR techniques [33] using a [100% ^{13}C, ^{15}N]-labeled sample. NOEs were identified using three-dimensional ^{13}C- and ^{15}N-separated NOESY spectra. Intermolecular contacts were obtained from a $^{13}C/^{15}N$ X-filtered NOESY experiment on the [^{12}C, ^{14}N]/[^{13}C, ^{15}N] mixed-labeled sample [34]. All spectra were processed using the AZARA suite of programs (W. Boucher, personal communication) and analyzed with ANSIG [35]. Structure diagrams were prepared using MOLSCRIPT [36] and Raster3D [37].

Structure Determination

Structures were calculated from extended templates by simulated annealing using CNS version 1.0 [38], with manual screening of ambiguous restraints. To generate the final NOE tables, ten iterations were calculated, each using 40 structures. In the last round 40 structures of the A domain were calculated, from which 8 with the lowest energy, no NOE violations greater than 0.5 Å, and no dihedral angle violations greater than 5° were selected for the final ensemble. For the B domain, 100 structures were calculated in the last round and 7 selected for the final ensemble, using the same criteria. The ensembles were assessed using PROCHECK-NMR [39].

Acknowledgments

We thank S. Kent and Dr. J. Garcia-Bernardo for compiling docking domain sequences, Dr. H. Hong for LC-MS analysis, Dr. M. Moncrieffe for advice and assistance, and Professor J. Staunton FRS for helpful discussions. This work was supported by a grant from the BBSRC. K.J.W. is a Royal Society Dorothy Hodgkin Fellow.

Received: May 6, 2003
Revised: June 17, 2003
Accepted: June 23, 2003
Published: August 22, 2003

References

1. Cortés, J., Haydock, S.F., Roberts, G.A., Bevitt, D.J., and Leadlay, P.F. (1990). An unusually large multifunctional polypeptide in the erythromycin-producing polyketide synthase of *Saccharopolyspora-erythraea*. Nature *348*, 176–178.
2. Donadio, S., Staver, M.J., McAlpine, J.B., Swanson, S.J., and Katz, L. (1991). Modular organization of genes required for complex polyketide biosynthesis. Science *252*, 675–679.
3. Staunton, J., and Weissman, K.J. (2001). Polyketide biosynthesis: a millennium review. Nat. Prod. Rep. *18*, 380–416.
4. Rawlings, B.J. (2001). Type I polyketide biosynthesis in bacteria (Part B). Nat. Prod. Rep. *18*, 231–281.
5. Aparicio, J.F., Caffrey, P., Marsden, A.F.A., Staunton, J., and

Leadlay, P.F. (1994). Limited proteolysis and active-site studies of the first multienzyme component of the erythromycin-producing polyketide synthase. J. Biol. Chem. *269*, 8524–8528.

6. Staunton, J., Caffrey, P., Aparicio, J.F., Roberts, G.A., Bethell, S.S., and Leadlay, P.F. (1996). Evidence for a double-helical structure for modular polyketide synthases. Nat. Struct. Biol. *3*, 188–192.
7. Aparicio, J.F., Molnár, I., Schwecke, T., König, A., Haydock, S.F., Khaw, L.E., Staunton, J., and Leadlay, P.F. (1996). Organization of the biosynthetic gene cluster for rapamycin in *Streptomyces hygroscopicus*: analysis of the enzymatic domains in the modular polyketide synthase. Gene *169*, 9–16.
8. Cane, D.E., and Walsh, C.T. (1999). The parallel and convergent universes of polyketide synthases and nonribosomal peptide synthetases. Chem. Biol. *6*, R319–R325.
9. Bevitt, D.J., Cortés, J., Haydock, S.F., and Leadlay, P.F. (1992). 6-deoxyerythronolide B synthase-2 from *Saccharopolyspora erythraea*—cloning of the structural gene, sequence analysis and inferred domain structure of the multifunctional enzyme. Eur. J. Biochem. *204*, 39–49.
10. Lupas, A. (1997). Predicting coiled-coil regions in proteins. Curr. Opin. Chem. Biol. *7*, 388–393.
11. Gokhale, R.S., and Khosla, C. (2000). Role of linkers in communication between protein modules. Curr. Opin. Chem. Biol. *4*, 22–27.
12. Tsuji, S.Y., Cane, D.E., and Khosla, C. (2001). Selective protein-protein interactions direct channeling of intermediates between polyketide synthase modules. Biochemistry *40*, 2326–2331.
13. Wu, N., Tsuji, S.Y., Cane, D.E., and Khosla, C. (2001). Assessing the balance between protein-protein interactions and enzyme-substrate interactions in the channeling of intermediates between polyketide synthase modules. J. Am. Chem. Soc. *123*, 6465–6474.
14. Gokhale, R.S., Tsuji, S.Y., Cane, D.E., and Khosla, C. (1999). Dissecting and exploiting intermodular communication in polyketide synthases. Science *284*, 482–485.
15. Wu, N., Cane, D.E., and 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.
16. Ranganathan, A., Timoney, M., Bycroft, M., Cortés, J., Thomas, I.P., Wilkinson, B., Kellenberger, L., Hanefeld, U., Galloway, I.S., Staunton, J., and Leadlay, P.F. (1999). Knowledge-based design of bimodular and trimodular polyketide synthases based on domain and module swaps: a route to simple statin analogues. Chem. Biol. *6*, 731–741.
17. Caffrey, P., Bevitt, D.J., Staunton, J., and Leadlay, P.F. (1992). Identification of DEBS-1, DEBS-2 and DEBS-3, the multienzyme polypeptides of the erythromycin-producing polyketide synthase from *Saccharopolyspora erythraea*. FEBS Lett. *304*, 225–228.
18. Gokhale, R.S., Hunziker, D., Cane, D.E., and Khosla, C. (1999). Mechanism and specificity of the terminal thioesterase domain from the erythromycin polyketide synthase. Chem. Biol. *6*, 117–125.
19. Kumar, P., Li, Q., Cane, D.E., and Khosla, C. (2003). Intermodular communication in modular polyketide synthases: structural and mutational analysis of linker mediated protein-protein recognition. J. Am. Chem. Soc. *125*, 4097–4102.
20. Liang, H., Sandberg, W.S., and Terwilliger, T.C. (1993). Genetic fusion of subunits of a dimeric protein substantially increases its stability and rate of folding. Proc. Natl. Acad. Sci. USA *90*, 7010–7014.
21. McDaniel, R., Kao, C.M., Hwang, S.J., and Khosla, C. (1997). Engineered intermodular and intramodular polyketide synthase fusions. Chem. Biol. *4*, 667–674.
22. Squire, C.M., Goss, R.J.M., Hong, H., Leadlay, P.F., and Staunton, J. (2003). Catalytically active tetramodular DEBS fusion proteins. Chem. Bio. Chem., in press.
23. Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR *13*, 289–302.
24. Branden, C., and Tooze, J. (1999). Introduction to Protein Structure (New York: Garland Publishing).
25. Rose, R.B., Endrizzi, J.A., Cronk, J.D., Holton, J., and Alber, T. (2000). High-resolution structure of the HNF-1α dimerization domain. Biochemistry *39*, 15062–15070.
26. Hill, R.B., and DeGrado, W.F. (1998). Solution structure of α_2D, a native-like *de novo* designed protein. J. Am. Chem. Soc. *120*, 1138–1145.
27. Fersht, A. (1999). Structure and Mechanism in Protein Science (New York: W.H. Freeman and Company).
28. Kammerer, R.A., Schulthess, T., Landwehr, R., Lustig, A., Engel, J., Aebi, U., and Steinmetz, M.O. (1998). An autonomous folding unit mediates the assembly of two-stranded coiled coils. Proc. Natl. Acad. Sci. USA *95*, 13419–13424.
29. Kammerer, R.A., Jaravine, V.A., Frank, S., Schulthess, T., Landwehr, R., Lustig, A., García-Escheverría, C., Alexandrescu, A.T., Engel, J., and Steinmetz, M.O. (2001). An intrahelical salt bridge within the trigger site stabilises the GCN4 leucine zipper. J. Biol. Chem. *276*, 13685–13688.
30. Hubbard,S.J., and Thornton, J.M. (1993) "NACCESS," Computer Program, Department of Biochemistry and Molecular Biology, University College London.
31. Glover, J.N.M., and Harrison, S.C. (1995). Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. Nature *373*, 257–261.
32. Laue, T.M., Shah, B.D., Ridgeway, T.M., and Pelletier, S.L. (1992). Analytical Ultracentrifugation in Biochemistry and Polymer Science, S.E. Harding, A.J. Rowe, and J.C. Horton, eds. (Cambridge, UK: Royal Society of Chemistry), pp 90–125.
33. Ferentz, A.E., and Wagner, G. (2000). NMR: a multifaceted approach to macromolecular structure. Q. Rev. Biophys. *33*, 29–65.
34. Zwahlen, C., Legault, P., Vincent, S.J.F., Greenblatt, J., Konrat, R., and Kay, L.E. (1997). Methods for measurement of intermolecular NOEs by multinuclear NMR spectroscopy: application to a bacteriophage λ N-peptide/boxB RNA complex. J. Am. Chem. Soc. *119*, 6711–6721.
35. Kraulis, P.J., Domaille, P.J., Campbell-Burk, S.L., van Aken, T., and Laue, E.D. (1994). Solution structure and dynamics of Ras p21GDP determined by heteronuclear 3 and 4-dimensional NMR spectroscopy. Biochemistry *33*, 3515–3531.
36. Kraulis, P.J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. *24*, 946–950.
37. Merritt, E.A., and Bacon, D.J. (1997). Raster3D: photorealistic molecular graphics. Methods Enzymol. *277*, 505–524.
38. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. Acta. Crystallogr. D *54*, 905–921.
39. Laskowski, R.A., MacArthur, M.W., and Thornton, J.M. (1998). Validation of protein models derived from experiment. Curr. Opin. Struct. Biol. *8*, 631–639.

Accession Numbers

The coordinates for the A and B domains of the Dock 2-3 protein have been deposited in the Protein Data Bank under ID codes 1PZQ and 1PZR, respectively. The BMRB accession number for the chemical shift data is 5885.