

Mechanism of intersubunit ketosynthase–dehydratase interaction in polyketide synthases

Matthew Jenner^{1,2*}, Simone Kosol¹, Daniel Griffiths¹, Panward Prasongpholchai¹, Lucio Manzi³, Andrew S Barrow³, John E Moses³, Neil J Oldham³, Józef R Lewandowski¹ & Gregory L Challis^{1,2*}

Modular polyketide synthases (PKSs) produce numerous structurally complex natural products that have diverse applications in medicine and agriculture. PKSs typically consist of several multienzyme subunits that utilize structurally defined docking domains (DDs) at their N and C termini to ensure correct assembly into functional multiprotein complexes. Here we report a fundamentally different mechanism for subunit assembly in *trans*-acyltransferase (*trans*-AT) modular PKSs at the junction between ketosynthase (KS) and dehydratase (DH) domains. This mechanism involves direct interaction of a largely unstructured docking domain (DD) at the C terminus of the KS with the surface of the downstream DH. Acyl transfer assays and mechanism-based crosslinking established that the DD is required for the KS to communicate with the acyl carrier protein appended to the DH. Two distinct regions for binding of the DD to the DH were identified using NMR spectroscopy, carbene footprinting, and mutagenesis, providing a foundation for future elucidation of the molecular basis for interaction specificity.

Modular PKSs assemble structurally diverse carbon frameworks via a succession of covalently tethered intermediates¹. Due to their modular architecture, which is responsible for a logical series of enzymatic chain elongation and modification reactions, such PKSs are frequently likened to molecular production lines². These fascinating molecular machines usually consist of several multienzyme subunits, and must undergo noncovalent self-assembly to produce functional megasynthases with molecular weights typically in the megadalton range^{3,4}.

Three domains are required for chain elongation in a PKS module: an acyl carrier protein (ACP) domain, which is post-translationally modified via attachment of a coenzyme A–derived phosphopantetheine (ppant) ‘arm’ to a conserved Ser residue; an acyltransferase (AT) domain, which loads an (alkyl)malonyl extender unit onto the ACP domain’s ppant thiol; and a ketosynthase (KS) domain, which utilizes a conserved active site Cys residue to receive the growing polyketide chain from the ppant thiol of the ACP domain in the upstream module and joins it to the extender unit via a decarboxylative Claisen condensation⁵. The α - and β -carbon atoms in the resulting β -ketoester can be modified by optional catalytic domains, including ketoreductases (KRs), dehydratases (DHs), enoylreductases, and C- and O-methyltransferases (MTs)^{6,7}.

Two phylogenetically distinct classes of modular PKSs, known as *cis*-AT and *trans*-AT, have been identified⁸. In the former, an AT domain is typically found in every module, whereas in the latter the modules lack such domains and a single standalone AT loads a malonyl extender unit onto each ACP domain⁸. The archetypal example of a *cis*-AT PKS is the 6-deoxyerythronolide B synthase (DEBS), which assembles the polyketide precursor of erythromycin A⁹. The three subunits of this well-studied system self-assemble via N- and/or C-terminal docking domains (DDs)^{10–12}, which are portable and can be exploited to produce engineered systems with unnatural subunit combinations¹³.

In contrast to their *cis*-AT counterparts, the mechanism(s) by which subunits self-assemble in *trans*-AT PKSs are less well understood. A key difference between these two types of assembly lines is that, unlike those in *cis*-AT PKSs, the junctions between subunits

in *trans*-AT systems often occur within a module. This gives rise to various types of split modules (for example, with junctions between KS/KR, KS/DH, KR/MT and DH/KR domains)¹⁴. Recently, DDs in the *trans*-AT PKSs responsible for the biosynthesis of virginiamycin and macrolactin have been reported^{15,16}. These consist of approximately 25 residue appendages at the N and C termini of subunits that form a four-helix bundle upon docking with a cognate partner and have been found at ACP/KS, DH/KR and KS/KR junctions. The complexes formed by such DDs were found to have dissociation constants in the 0.8–10 μ M range, and a high degree of selectivity for cognate over noncognate partners was observed.

Despite these advances, the mechanism of communication across KS/DH junctions in *trans*-AT PKSs remains obscure. To address this problem, we elected to study the GbnD4 and GbnD5 subunits of the gladiolin PKS, which have KS and DH domains at their C and N termini, respectively. Gladiolin is a macrolide antibiotic with promising activity against drug-resistant strains of *Mycobacterium tuberculosis* and negligible toxicity toward mammalian cells that we recently discovered as a metabolite of *Burkholderia gladioli* BCC0238, a clinical isolate from a cystic fibrosis patient¹⁷. Here we show that a largely unstructured DD at the C terminus of the GbnD4 KS domain interacts directly with the GbnD5 DH domain, revealing a new paradigm for subunit communication across KS/DH junctions in PKS assembly lines.

RESULTS

Identifying putative docking domains at KS/DH junctions

Inspection of the sequences of KS/DH junctions in *trans*-AT PKSs suggested that, unlike other types of catalytic domains at subunit interfaces, the DH domains lack an N-terminal DD. On the other hand, the KS domains have an approximately 40–80-residue unannotated region fused to their C termini, which could function as DDs that interact directly with the downstream DH domain (Fig. 1a). These putative DH docking (DHD) domains were found to be present at most, but not all, KS/DH junctions in *trans*-AT PKSs for which the metabolic product is known (Supplementary Fig. 1). A profile hidden Markov model based on the sequences of

¹Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK. ²Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, UK. ³School of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK. *e-mail: M.Jenner@warwick.ac.uk or G.L.Challis@warwick.ac.uk

the putative DHD domains in these systems was developed¹⁸, but it was found to perform poorly in the identification of additional examples, consistent with the low level of conservation observed for these regions.

Secondary structure analysis of the aligned DHD domain sequences using the PsiPred and IUPred servers indicated a high degree of disorder with the propensity for β -strand formation in two distinct regions^{19,20}. The DHD domains are predicted to have flexible or context-dependent dynamics, with most order parameters, S^2 , below 0.8 ($S^2 = 1$ indicates complete rigidity; $S^2 = 0$ indicates unrestricted motion) in both the presence and the absence of the corresponding KS domain (Supplementary Fig. 2). Regions of the DHD domains with higher order parameters appear to coincide with amino acids that are predicted to be involved in binding events by the web server ANCHOR²¹. Despite the lack of sequence conservation, the predicted binding sites seemingly occur in similar two-site patterns for all DHD domains (Supplementary Fig. 3). In addition, DHD domains possess a notably low isoelectric point (pI), which may facilitate association with the downstream DH domain via electrostatic interactions. Interestingly, phylogenetic analysis of all *trans*-AT PKS DH domains revealed that those predicted to be interaction partners for DHD domains belong to a discrete clade (Supplementary Fig. 4).

Acyl transfer across a *trans*-AT PKS KS/DH junction

To investigate the role of DHD domains in communication across KS/DH junctions, we overproduced the C-terminal ACP-KS-DHD tri-domain of GbnD4 and the N-terminal DH-ACP di-domain of GbnD5 (Fig. 1b) in *Escherichia coli* as N-terminal His₆ fusion proteins, which were purified to near-homogeneity using immobilized metal-ion affinity chromatography (IMAC) (Supplementary Fig. 5). The identity of the purified proteins was confirmed by ESI-Q-TOF-MS analyses (Supplementary Fig. 6). The gladiolin PKS contains 20 KS domains, of which 17 are predicted to catalyze chain elongation and 3 appear to function as transacylases that shuttle particular chain elongation intermediates between ACP domains¹⁷. In most other *trans*-AT PKSs, a conserved active site His residue, which plays an essential role in chain elongation, is lacking from transacylating KS domains. However, this is not the case for the gladiolin PKS, and it is unclear whether the KS domain at the C terminus of GbnD4 catalyzes chain elongation or transacylation. We thus sought to establish whether the KS domain in the ACP-KS-DHD tri-domain construct is active by investigating whether it can catalyze transfer of an acetyl group from the upstream ACP domain to the corresponding domain in the DH-ACP construct. To do this, we exploited the inability of *E. coli* to convert ACP domains of PKSs into their phosphopantetheinylated holo form. Sfp, a substrate-tolerant phosphopantetheinyl transferase, was used to transfer the S-acetyl-ppant moiety of acetyl-CoA onto the ACP domain of the ACP-KS-DHD construct (Supplementary Fig. 7). The DH-ACP di-domain was converted to its holo form by treatment with Sfp and coenzyme A (Supplementary Fig. 7) and then incubated with a two-fold excess of the acetylated holo-ACP-KS-DHD tri-domain. Separation and analysis of the protein mixture using UHPLC-ESI-Q-TOF-MS showed that 39.9% ($\sigma^2 = 1.34$) of the DH-ACP di-domain was acetylated (Fig. 2a and Supplementary Fig. 8). The level of DH-ACP di-domain acetylation was much lower (5.9%, $\sigma^2 = 0.39$) in an analogous experiment employing an ACP-KS-DHD variant, in which the active site Cys residue of the KS domain had been mutated to Ala, confirming that the KS domain catalyzes the acetyl transfer reaction (Fig. 2a).

We next investigated whether the C-terminal DHD domain of GbnD4 and the N-terminal DH domain of GbnD5 play a role in the KS-catalyzed transacylation reaction. A GbnD4 apo-ACP-KS(Δ DHD) di-domain, lacking the C-terminal DHD domain, was overproduced, purified and converted to its acetylated holo form

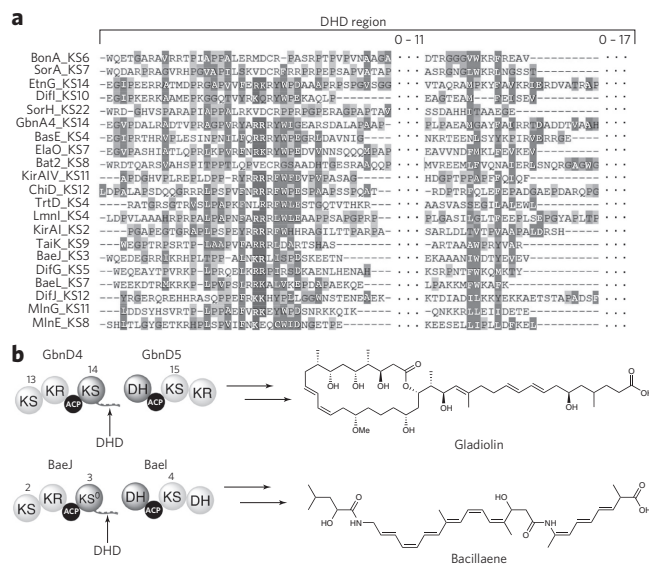


Figure 1 | Bioinformatics analysis of DHD domains and examples from the gladiolin and bacillaene PKSs.

(a) Multiple sequence alignment of DHD domains identified by manual inspection of KS/DH boundaries in *trans*-AT PKSs for which the metabolic product is known. Localized conservation is observed in the N-terminal region, but the C-terminal region is poorly conserved. Dotted sections indicate residues between the N- and C-terminal regions of the DHD domain, and the number of residues located in these regions is indicated above. Sequences from the following pathways were used: Bae, bacillaene (*Bacillus amyloliquefaciens*); Bas, basiliskamide (*Bacillus laterosporus* PE36); Bat, batumin/kalimantacin (*Pseudomonas fluorescens*); Dfn, difficidin (*B. amyloliquefaciens*); Etn, etnangien (*Sorangium cellulosum* So ce56); Ela, elansolid (*Chitinophaga pinensis*); Gbn, gladiolin (*B. gladioli*); Kir, kirromycin (*Streptomyces collinus*); Lnm, leinamycin (*Streptomyces atroolivaceus*); Mln, macrolactin (*B. amyloliquefaciens*); Mmp, mupirocin (*P. fluorescens*); Sor, sorangicin (*S. cellulosum*); Tai, thailandamide (*Burkholderia thailandensis* E264); Trt, tartrolon (*Teredinibacter turnerae* T7901). (b) Examples of DHD domains in the gladiolin (top) and bacillaene (bottom) PKSs and structures of the metabolites produced by these assembly lines.

using Sfp and acetyl-CoA (Supplementary Figs. 5–7). Negligible acetylation of the holo-ACP-DH di-domain was observed when it was incubated with the acetylated holo-ACP-KS(Δ DHD) di-domain (Fig. 2a). Similarly, the GbnD5 ACP domain, which lacks the upstream DH domain, was overproduced, purified and phosphopantetheinylated using Sfp and coenzyme A (Supplementary Figs. 5–7). A small amount of acetyl transfer was observed (6.6%; $\sigma^2 = 0.63$) when the GbnD5 ACP domain was incubated with the acetylated GbnD4 holo-ACP-KS-DHD tri-domain (Supplementary Fig. 9). These data demonstrate that both the DHD domain and the DH are required for efficient acyl transfer across the GbnD4–GbnD5 interface.

Crosslinking of the ACP-KS-DHD-DH-ACP complex

Type II fatty acid synthase ACPs post-translationally modified with an analog of ppant in which the terminal thiol has been replaced by *trans*- β -chloroacrylamide have been shown to react efficiently with the active site Cys residue of cognate KSs to generate covalently crosslinked complexes²². Such an approach has also been shown to be effective for crosslinking ACP and KS domains from the DEBS *cis*-AT PKS²³. The extent of crosslinking in these experiments correlates with the strength of the interaction between the proteins. We exploited this mechanism-based crosslinker to further explore the interaction between the GbnD4

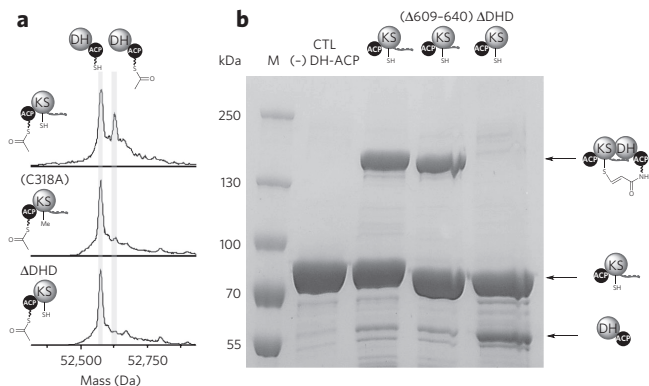


Figure 2 | Acyl transfer and protein crosslinking assays demonstrate that DHD domains play a key role in communication across KS/DH interfaces.

(a) Deconvoluted mass spectra of GbnD5 holo-DH-ACP following incubation with GbnD4 Ac-ACP-KS-DHD (top), Ac-ACP-KS(C318A)-DHD (middle) and Ac-ACP-KS(Δ DHD) (bottom). Transfer of the acetyl group onto the holo-DH-ACP di-domain results in a +42 Da mass shift. The data show that an acyl group can be transferred across the subunit interface (top) and that the KS active site Cys residue plays an important role in this process (middle). No acetylation of the holo-DH-ACP di-domain is observed when the DHD domain is removed (bottom). (b) SDS-PAGE (6%) analysis of crosslinking reactions between the GbnD5 DH-ACP di-domain loaded with the chloroacrylamide-terminated ppant analog and the GbnD4 ACP-KS-DHD tri-domain (left), the GbnD4 ACP-KS-DHD(Δ 609–640) truncated tri-domain (middle), and the ACP-KS(Δ DHD) di-domain (right). Efficient formation of a crosslinked complex (~130 kDa) is observed for the ACP-KS-DHD tri-domain (left) and GbnD4 ACP-KS-DHD(Δ 609–640) truncated tri-domain (middle), but not the ACP-KS(Δ DHD) di-domain (right). CTL, control; M, molecular weight markers.

ACP-KS-DHD tri-domain and the GbnD5 DH-ACP di-domain. The *trans*- β -chloroacrylamide-containing pantetheine analog was synthesized and then enzymatically converted to the corresponding coenzyme A analog, which was used to post-translationally modify the ACP domain of the apo-DH-ACP construct through Sfp (Supplementary Fig. 10) using an established procedure²³. Incubation of the resulting crosslinker-modified DH-ACP di-domain with the GbnD4 apo-ACP-KS-DHD tri-domain resulted in efficient formation of a covalent complex (Fig. 2b). No such complex was observed when the crosslinker-modified DH-ACP di-domain was omitted from the reaction or when the wild-type apo-ACP-KS-DHD tri-domain was replaced by the C318A mutant (Fig. 2b and Supplementary Fig. 11). Similarly, incubation of the crosslinker-modified GbnD5 DH-ACP di-domain with the ACP-KS(Δ DHD) di-domain yielded no covalent complex (Fig. 2b), confirming the importance of the DHD domain for communication across the GbnD4–GbnD5 interface.

Bioinformatics analyses indicate that DHD domains contain two separate DH-binding regions (Supplementary Fig. 3). To explore the contribution made by each of these regions to communication across the junction between KS and DH domains, we produced a GbnD4 ACP-KS-DHD(Δ 609–640) construct. In this construct, the C-terminal DH-binding region of the DHD domain was removed, but the N-terminal DH-binding region remained intact. Incubation of the crosslinker-modified DH-ACP di-domain with the ACP-KS-DHD(Δ 609–640) construct resulted in a similar amount of covalent complex formation to that observed with the ACP-KS-DHD tri-domain (Fig. 2b). This indicates that the N-terminal DH-binding region plays a key role in the formation of a complex between the DHD and DH domains. A modest reduction in the amount of acetyl transfer was also observed (31.1%; $\sigma^2 = 0.16$), providing additional

support for this hypothesis. Further experiments will be required to clarify the role played by the C-terminal DH-binding region, which shows a lower degree of conservation than the N-terminal region (Fig. 1a).

Characterization of the DHD domain by NMR spectroscopy

The GbnD4 DHD domain was overproduced as an N-terminal His₆ fusion protein and purified via IMAC, allowing its structural and dynamic properties to be investigated using NMR spectroscopy, which is an ideal tool for atomic resolution characterization of largely unstructured proteins^{24–26}. Sharp signals with narrow dispersion in the ¹H dimension of ¹H–¹⁵N HSQC spectra indicated that the DHD domain adopts a mostly disordered structure. Signals resulting from 83% of the backbone heavy atoms were assigned using a standard set of triple resonance experiments on a uniformly ¹³C- and ¹⁵N-labeled sample. Some of the residues were found to give rise to a second, weaker set of signals, suggesting the presence of multiple conformations that slowly interconvert (Fig. 3a). The slow rate of exchange may be due, at least in part, to the relatively low temperature at which the spectra were acquired. The chemical shifts of signals resulting from the major conformational isomer were used to calculate secondary structure propensity scores, allowing detection of regions of the DHD domain with a tendency to form α -helices or β -strands. No propensity for secondary structure formation was found for the minor conformer. Only very small propensities that barely reached the threshold for any type of secondary structure element were identified for the major conformer (Supplementary Fig. 12). Consistent with this, the experimental chemical shift values were in good agreement with those calculated for a random coil conformation of the DHD domain (Supplementary Fig. 12). Moreover, a ¹H–¹H NOESY experiment indicated a lack of long-range contacts, suggesting that the DHD domain does not adopt a defined tertiary structure. Finally, the circular dichroism spectrum of the DHD domain was consistent with a largely unstructured protein, indicating only a modest propensity for β -strand formation (Supplementary Fig. 13).

Additional insight into the structural properties of the Gbn4 DHD domain was provided by backbone dynamics measurements. As is typical for an intrinsically disordered protein, steady state ¹H–¹⁵N NOE, ¹⁵N spin-lattice (*R*₁) and spin-spin (*R*₂) relaxation rates revealed that the DHD domain has high local mobility on the fast picosecond-to-nanosecond timescale (Supplementary Fig. 14). Whereas the low overall ¹H–¹⁵N NOE values for the DHD domain show that it is generally very flexible, Val19 to Glu30 and, to a lesser extent, Ala49 to Arg60 gave higher ¹H–¹⁵N NOE values, indicative of reduced mobility on the fast time scale in these regions. A similar trend, manifested by higher *R*₂ rates in the Val19–Glu30 and Ala49–Arg60 regions, was also observed in the ¹⁵N relaxation data (Supplementary Fig. 14).

Probing the interaction between the DHD and DH domains

To investigate the nature of the interaction between the DHD and DH domains, we examined how the 2D BEST-TROSY-¹⁵N-HSQC spectrum of the uniformly ¹⁵N-labeled GbnD4 DHD domain changes when increasing concentrations of the GbnD5 DH-ACP di-domain are added. Several signals were observed to decrease in intensity without any notable change in chemical shift, and no new signals appeared in the spectrum (Fig. 3a), suggesting a slow-to-intermediate exchange process and a dissociation constant in the low micromolar range. Interestingly, the residues of the DHD domain for which we observed the strongest decrease in signal intensity as a function of DH-ACP concentration were localized in the Val19–Glu30 and Ala49–Arg60 regions (Fig. 3b,c). This suggests that two distinct sites of the DHD domain bind to the DH-ACP di-domain. Unfortunately, detection of any structural or conformational changes in the DHD domain upon binding to

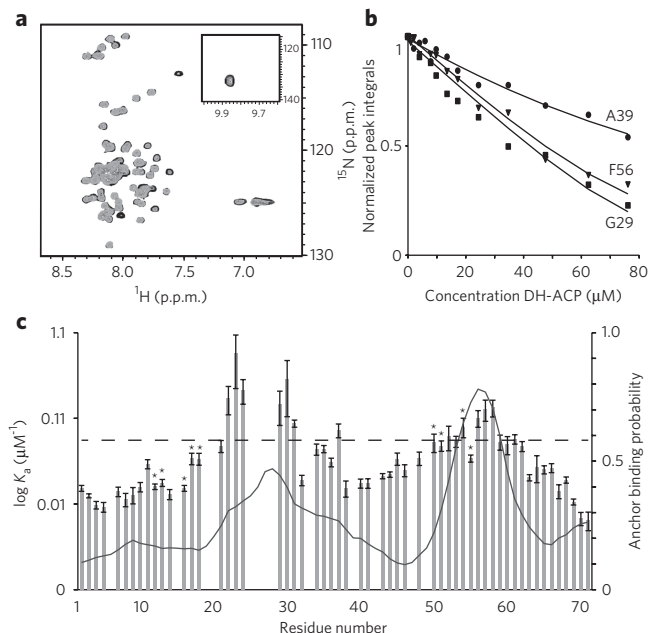


Figure 3 | NMR titrations reveal two regions of DHD domains involved in the interaction with DH-ACP di-domains. (a) Overlaid 2D BEST-TROSY- ^{15}N -HSQC spectra of U- ^{15}N -labeled GbnD4 DHD domain titrated with unlabeled GbnD5 apo-DH-ACP di-domain. The spectrum in black shows the DHD domain alone. The gray spectrum shows the DHD domain in the presence of an equivalent amount of the GbnD5 DH-ACP di-domain. As the concentration of the GbnD5 DH-ACP di-domain was increased, several signals were observed to decrease in intensity, suggestive of slow to intermediate exchange with the binding affinity in the lower micromolar range. (b) Experimental data points and individual fits of normalized signal integrals for selected residues located in the first (Gly29, squares) and second (Phe56, triangles) DH-ACP-interacting regions of the DHD domain, and the region between them (Ala39, circles). (c) $\log K_s$ values calculated from signal integral values for individual residues of the DHD domain. Residues with values above the threshold (dotted line, defined in Online Methods) are involved in an interaction with the DH-ACP di-domain. These data agree well with the binding site probabilities predicted by the ANCHOR web server (solid line). Asterisks indicate residues for which a second set of weaker resonances was observed in the spectra. The error bars correspond to two s.d. of 1,000 iterations of fits using Monte Carlo error analysis.

the DH-ACP di-domain was precluded by the large size of the complex, which caused broadening of the signals in the NMR spectrum beyond the limit of detection. Because of the slow-to-intermediate rate of exchange and the two-site interaction of the DHD domain with the DH-ACP di-domain, it was only possible to estimate the dissociation constant of the complex. The rate of signal intensity decay as a function of DH-ACP di-domain concentration yields apparent dissociation constants on the residue level (Fig. 3b). Compared to the average of these apparent dissociation constants for the full DHD domain (approximately 40 μM), the average values for the Val19–Glu30 and Ala49–Arg60 regions suggest that these bind to the DH-ACP di-domain with substantially higher affinities (average apparent dissociation constants in the 7–15 μM range).

Mapping of interaction sites by carbene footprinting

To establish the site(s) of interaction of the GbnD4 DHD domain with the GbnD5 DH domain, we employed a recently developed carbene footprinting methodology²⁷. This approach involves labeling of the solvent-exposed surface of a protein with a highly reactive

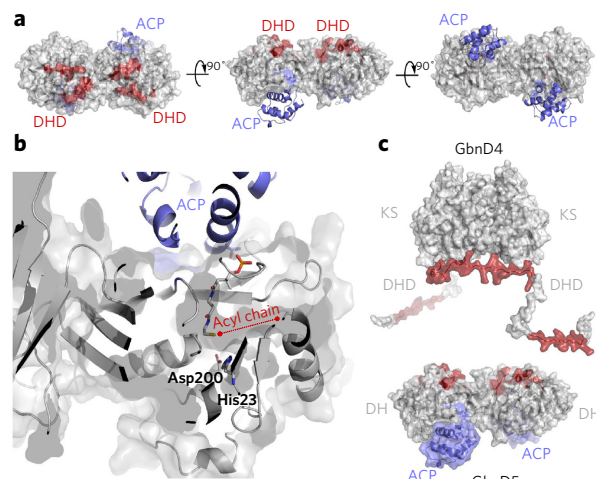


Figure 4 | Mapping DHD and ACP domain interaction sites on the DH domain. (a) Identification of the interaction sites of the GbnD4 DHD domain and GbnD5 ACP domain on the dimeric GbnD5 DH domain using carbene footprinting. Regions masked by the DHD and ACP domains are highlighted in red and blue, respectively, on a homology model of GbnD5 DH domain (gray). Residue constraints derived from the footprinting experiments were used to dock the ACP domain (blue) onto the DH domain. Residues used to constrain the docking calculations are as follows: DH: Leu31, Arg61, Leu298, Val299, Asp300; ACP: Leu394, Ala395, Leu396. The model shows adjacent, but distinct, binding sites for the DHD and ACP domains. These place the GbnD4 KS domain and the GbnD5 ACP domain in close proximity, facilitating efficient communication across the subunit interface. (b) Predicted location of the ppant arm in the model of the GbnD5 DH-ACP di-domain complex based on the X-ray crystal structure of a crosslinked complex of the ACP and DH from the *E. coli* fatty acid synthase (PDB ID 4KEH). The docking site for the ACP domain is situated near the entrance to the active site of the DH domain. This allows the thiol of the ppant arm to sit adjacent to the catalytic Asp and His residues in the active site of the DH domain. A putative binding pocket for the acyl chain of the substrate sits to the right of the ppant thiol. (c) Overall model for the interaction between the GbnD4 KS-DHD (gray/red) and GbnD5 DH-ACP (gray/blue) di-domains, both of which form homodimers. The red patches on the surface of the DH and DHD domains indicate mutual interaction sites.

carbene followed by proteolytic digestion and LC–MS/MS analysis. When this experiment is carried out in the presence and absence of an interaction partner, differential labeling is observed, allowing the site of interaction to be identified.

LC–MS analysis of tryptic digests of the DHD domain and DH-ACP di-domain resulted in 70% and 52% sequence coverage, respectively (Supplementary Fig. 15). Differential labeling of five tryptic fragments from the DH domain was observed when carbene footprinting was carried out in the presence and absence of the DHD domain (Supplementary Fig. 16). MS/MS analysis of these differentially labeled tryptic fragments allowed near-residue level identification of the DHD domain's binding site (Supplementary Fig. 17). On the other hand, differential labeling of only a single tryptic fragment of the DHD domain was observed (Supplementary Fig. 16), and MS/MS analysis showed that the masked region overlaps with the C-terminal DH-interacting region of the DHD domain (Ala49–Arg60) identified by NMR spectroscopic analysis (Supplementary Fig. 17). Unfortunately, no labeling data was available for the other DH-interacting region identified by the NMR experiments, Val19–Glu30, because of the presence of three contiguous Arg residues in the sequence.

We also sought to map the interaction of the ACP domain with the DH domain in the GbnD5 DH-ACP di-domain. To do this, the DH domain was overproduced as an N-terminal His₆ fusion protein in *E. coli* and purified using IMAC (Supplementary Fig. 5). Thirty-nine percent sequence coverage was observed in LC-MS analyses of tryptic digests of the ACP domain (Supplementary Fig. 15). Three tryptic fragments from the DH domain and one fragment from the ACP domain were found to be differentially labeled when the data from carbene footprinting of the DH-ACP di-domain were compared to the data from the standalone DH and ACP domains (Supplementary Fig. 18). As before, the interaction interface between the DH and the ACP was mapped to near-residue level using MS/MS (Supplementary Fig. 19).

The results of these experiments were visualized on a homology model of the GbnD5 DH domain, revealing distinct binding sites for the DHD and ACP domains. These sites are located adjacent to each other, allowing simultaneous interaction of the DHD and ACP domains with the DH domain (Fig. 4a). This brings the GbnD4 KS domain into close proximity of the GbnD5 ACP domain, facilitating acyl transfer or chain elongation across the subunit interface. The GbnD5 ACP domain also needs to be capable of delivering the acyl group attached to its ppant thiol into the active site of the DH domain. Docking simulations using residue constraints derived from the results of the carbene footprinting experiments place the conserved Ser residue that undergoes phosphopantetheinylation in the ACP domain at the entrance to the DH active site (Fig. 4b). Taken together, these data suggest that a single ACP binding site is sufficient to allow the ppant arm to access the active sites of both the GbnD4 KS domain and the GbnD5 DH domain (Fig. 4c).

Probing the specificity of the DHD-DH interaction

To investigate whether the GbnD4 DHD domain can interact with DH domains from other systems, we overproduced and purified the N-terminal DH-ACP di-domain from the BaeL subunit of the bacillaene PKS (Supplementary Fig. 5) and examined its ability to communicate with the GbnD4 ACP-KS-DHD tri-domain using the acyl transfer and crosslinking assays (Fig. 5). The extent of acyl transfer and crosslinking for the noncognate proteins was substantially reduced in comparison to that observed for the cognate pair (Fig. 5a), congruent with the notion that a degree of specificity in the interaction between DHD and DH domain pairs is required for the correct assembly of *trans*-AT PKS subunits, especially in systems that contain multiple KS-DHD/DH-ACP junctions within a single assembly line (Supplementary Fig. 20). In addition, we examined whether the GbnD4 ACP-KS-DHD(Δ 609-640) construct can interact with the DH-ACP di-domain from BaeL. However, a covalent complex was not formed when the chloroacrylamide-modified DH-ACP di-domain was incubated with this construct (Fig. 5b), indicating that the N-terminal DH-binding region of the DHD domain and/or the KS domain are important determinants of interaction specificity.

DISCUSSION

The mechanism illuminated here for intersubunit communication across KS/DH junctions differs markedly from that reported for other types of junction in *trans*-AT PKSs^{15,16}, which involve mutually compatible DDs at both the N- and C termini of the interacting subunits. Indeed, to the best of our knowledge, utilization of only a single DD at the C terminus of the upstream subunit to interact directly with a catalytic domain at the N terminus of the downstream subunit represents a fundamentally new mechanism for ensuring productive subunit interaction in assembly line multienzymes.

Unlike other types of DDs in *trans*-AT (and other) PKSs, the DHD domain is largely unstructured (note, however, that the four-helix bundle DD at the N terminus of VirG also appears to show elements of disorder¹⁵). Despite this, the dissociation constant estimated for

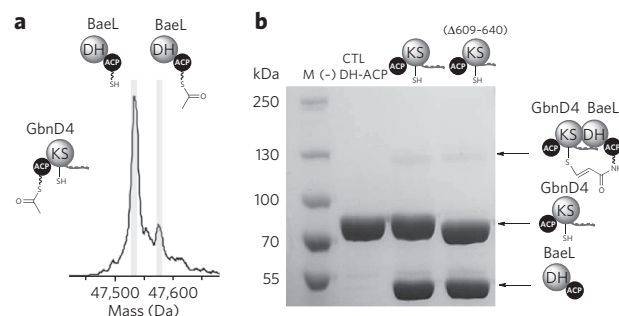


Figure 5 | DHD domains interact selectively with their cognate DH domain partners. (a) Deconvoluted mass spectrum of the BaeL holo-DH-ACP di-domain following incubation with the GbnD4 Ac-ACP-KS-DHD tri-domain. In comparison to the GbnD4 holo-DH-ACP di-domain, only low levels of acetylation can be observed, suggesting that the GbnD4 DHD and the BaeL DH domains interact weakly. (b) SDS-PAGE (6%) analysis of the crosslinking reaction between the GbnD4 ACP-KS-DHD tri-domain and GbnD4 ACP-KS-DHD(Δ 609-640) truncated tri-domain with the BaeL DH-ACP di-domain loaded with the β -chloroacrylamido ppant analog. Only trace amounts of the crosslinked complex (~130 kDa) can be observed, providing further evidence for a weak interaction between noncognate DHD and DH domains.

the DHD domain from its DH domain partner is in the low micromolar range, which is similar to that reported for other types of DD in both *cis*- and *trans*-AT PKSs^{10,15,16}. The occurrence of intrinsically disordered regions in prokaryotic proteins is not uncommon, but is more common in eukaryotes^{28,29}, wherein they play important roles in intracellular signaling and the ordered assembly of macromolecular complexes³⁰. Such regions evolve faster than globular regions³¹, and are not only subject to enhanced Darwinian selection³², but also tolerate a greater range of mutations, allowing faster adjustments to specificity³³. The short and disordered nature of the two DH-binding regions of the DHD domain suggests that they should be classified as 'short linear motifs' (SLiMs)³⁴. The utilization of SLiMs to mediate intersubunit communication in *trans*-AT PKSs may facilitate the evolution of novel biosynthetic pathways³⁵.

The carbene footprinting analyses show that the GbnD4 DHD domain interacts with two distinct regions on the surface of the GbnD5 DH domain. These lie on opposite sides of a hydrophobic cavity, which, in conventional DH domains (i.e., those internal to a subunit) in both *cis*- and *trans*-AT PKSs, is occupied by a ~20 residue N-terminal region that is lacking in DH domains from KS/DH junctions³⁶. A conserved HPLL motif in the N-terminal appendage of conventional DH domains occludes the region corresponding to the binding site for the DHD domain on the surface of DH domains associated with KS/DH junctions. On the basis of structural studies of a DH domain from a KS/DH junction in another *trans*-AT PKS system, this recently led others to speculate that this region may provide a binding site for a putative DD at the C terminus of the upstream subunit (Supplementary Fig. 21)³⁷.

Bioinformatics analyses suggest that the two regions in the GbnD4 DHD domain that have been identified to interact with the GbnD5 DH domain are likely to be a general feature of all DHD domains (Supplementary Fig. 3). Higher levels of conservation are observed for the N-terminal region, consistent with the observation that it is critical for complex formation (Fig. 2b).

In conclusion, complementary data from a range of structural and biochemical techniques have afforded unparalleled insights into the mechanism of communication across a ubiquitous type of subunit interface in *trans*-AT PKSs. These insights could ultimately facilitate future efforts to create hybrid assembly lines incorporating

subunits from different *trans*-AT PKS systems that are capable of producing radically new chemotypes.

Received 24 February 2017; accepted 15 November 2017; published online 8 January 2018

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

References

- Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed. Engl.* **48**, 4688–4716 (2009).
- Keatinge-Clay, A.T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
- Richter, C.D., Nietlispach, D., Broadhurst, R.W. & Weissman, K.J. Multienzyme docking in hybrid megasynthetases. *Nat. Chem. Biol.* **4**, 75–81 (2008).
- Weissman, K.J. & Müller, R. Protein-protein interactions in multienzyme megasynthetases. *ChemBioChem* **9**, 826–848 (2008).
- Fischbach, M.A. & Walsh, C.T. Assembly-line enzymology for polyketide and nonribosomal peptide antibiotics: logic, machinery, and mechanisms. *Chem. Rev.* **106**, 3468–3496 (2006).
- Keatinge-Clay, A.T. Stereocontrol within polyketide assembly lines. *Nat. Prod. Rep.* **33**, 141–149 (2016).
- Wagner, D.T. *et al.* α -Methylation follows condensation in the gephyronic acid modular polyketide synthase. *Chem. Commun. (Camb.)* **52**, 8822–8825 (2016).
- Helfrich, E.J.N. & Piel, J. Biosynthesis of polyketides by *trans*-AT polyketide synthases. *Nat. Prod. Rep.* **33**, 231–316 (2016).
- Katz, L. The DEBS paradigm for type I modular polyketide synthases and beyond. *Methods Enzymol.* **459**, 113–142 (2009).
- Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F. & Weissman, K.J. The structure of docking domains in modular polyketide synthases. *Chem. Biol.* **10**, 723–731 (2003).
- Gokhale, R.S. & Khosla, C. Role of linkers in communication between protein modules. *Curr. Opin. Chem. Biol.* **4**, 22–27 (2000).
- Buchholz, T.J. *et al.* Structural basis for binding specificity between subclasses of modular polyketide synthase docking domains. *ACS Chem. Biol.* **4**, 41–52 (2009).
- Kuo, J., Lynch, S.R., Liu, C.W., Xiao, X. & Khosla, C. Partial *in vitro* reconstitution of an orphan polyketide synthase associated with clinical cases of nocardiosis. *ACS Chem. Biol.* **11**, 2636–2641 (2016).
- Gulder, T., Freeman, M. & Piel, J. The catalytic diversity of multimodular polyketide synthases: natural product biosynthesis beyond textbook assembly rules. *Top. Curr. Chem.* http://dx.doi.org/10.1007/128_2010_113 (2011).
- Dorival, J. *et al.* Characterization of intersubunit communication in the virginiamycin *trans*-acyl transferase polyketide synthase. *J. Am. Chem. Soc.* **138**, 4155–4167 (2016).
- Zeng, J. *et al.* Portability and structure of the four-helix bundle docking domains of *trans*-acyltransferase modular polyketide synthases. *ACS Chem. Biol.* **11**, 2466–2474 (2016).
- Song, L. *et al.* Discovery and biosynthesis of gladiolin: a *Burkholderia gladioli* antibiotic with promising activity against *Mycobacterium tuberculosis*. *J. Am. Chem. Soc.* **139**, 7974–7981 (2017).
- Eddy, S.R. Profile hidden Markov models. *Bioinformatics* **14**, 755–763 (1998).
- McGuffin, L.J., Bryson, K. & Jones, D.T. The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404–405 (2000).
- Dosztányi, Z., Csizmok, V., Tompa, P. & Simon, I. IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics* **21**, 3433–3434 (2005).
- Dosztányi, Z., Mészáros, B. & Simon, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**, 2745–2746 (2009).
- Worthington, A.S., Rivera, H., Torpey, J.W., Alexander, M.D. & Burkart, M.D. Mechanism-based protein cross-linking probes to investigate carrier protein-mediated biosynthesis. *ACS Chem. Biol.* **1**, 687–691 (2006).

- Kapur, S. *et al.* Mechanism based protein crosslinking of domains from the 6-deoxyerythronolide B synthase. *Bioorg. Med. Chem. Lett.* **18**, 3034–3038 (2008).
- Kosol, S., Contreras-Martos, S., Cedeño, C. & Tompa, P. Structural characterization of intrinsically disordered proteins by NMR spectroscopy. *Molecules* **18**, 10802–10828 (2013).
- Jensen, M.R., Ruigrok, R.W. & Blackledge, M. Describing intrinsically disordered proteins at atomic resolution by NMR. *Curr. Opin. Struct. Biol.* **23**, 426–435 (2013).
- Uversky, V.N. Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J.* **282**, 1182–1189 (2015).
- Manzi, L. *et al.* Carbene footprinting accurately maps binding sites in protein-ligand and protein-protein interactions. *Nat. Commun.* **7**, 13288 (2016).
- Pavlović-Lažetić, G.M. *et al.* Bioinformatics analysis of disordered proteins in prokaryotes. *BMC Bioinformatics* **12**, 66 (2011).
- Vucetic, S. *et al.* DisProt: a database of protein disorder. *Bioinformatics* **21**, 137–140 (2005).
- Wright, P.E. & Dyson, H.J. Intrinsically disordered proteins in cellular signalling and regulation. *Nat. Rev. Mol. Cell Biol.* **16**, 18–29 (2015).
- Lin, Y.S., Hsu, W.L., Hwang, J.K. & Li, W.H. Proportion of solvent-exposed amino acids in a protein and rate of protein evolution. *Mol. Biol. Evol.* **24**, 1005–1011 (2007).
- Nilsson, J., Grahn, M. & Wright, A.P.H. Proteome-wide evidence for enhanced positive Darwinian selection within intrinsically disordered regions in proteins. *Genome Biol.* **12**, R65 (2011).
- Bellay, J. *et al.* Bringing order to protein disorder through comparative genomics and genetic interactions. *Genome Biol.* **12**, R14 (2011).
- Neduvu, V. & Russell, R.B. Peptides mediating interaction networks: new leads at last. *Curr. Opin. Biotechnol.* **17**, 465–471 (2006).
- Gitlin, L., Hagai, T., LaBarbera, A., Solovey, M. & Andino, R. Rapid evolution of virus sequences in intrinsically disordered protein regions. *PLoS Pathog.* **10**, e1004529 (2014).
- Keatinge-Clay, A. Crystal structure of the erythromycin polyketide synthase dehydratase. *J. Mol. Biol.* **384**, 941–953 (2008).
- Wagner, D.T. *et al.* Structural and functional trends in dehydrating bimodules from *trans*-acyltransferase polyketide synthases. *Structure* **25**, 1045–1055 e2 (2017).

Acknowledgments

This research was supported by grants from the BBSRC (BB/L021692/1 to G.L.C. and BB/L022761/1 to J.R.L.). The Bruker MaXis II instrument used in this study was funded by the BBSRC (BB/M017982/1). N.J.O., J.E.M., L.M. and A.S.B. thank the University of Nottingham for funding. The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007–2013)/ERC Grant Agreement 639907 (to J.R.L.). G.L.C. is the recipient of a Wolfson Research Merit Award from the Royal Society (WM130033). We thank L. Song for assistance with LC-MS analyses.

Author contributions

M.J. and G.L.C. designed the study, and all authors contributed to performing the research as follows: M.J. and D.G. carried out bioinformatics analyses; M.J. overproduced and purified all recombinant proteins, and conducted acyl transfer assays and LC-MS analyses of intact proteins; P.P. synthesized the *trans*- β -chloroacrylamide-containing pantetheine analog, and P.P. and M.J. carried out the crosslinking experiments; S.K. and J.R.L. conducted the NMR and CD experiments, and analyzed the data; A.S.B. and J.E.M. synthesized the diazirine reagent used in the footprinting experiments; L.M. and N.J.O. carried out the footprinting experiments and analyzed the data; M.J., S.K., J.R.L. and G.L.C. wrote the paper, and all authors contributed to revision of the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the [online version of the paper](#). Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.J. or G.L.C.

ONLINE METHODS

Profile hidden Markov model. Amino acid sequences corresponding to putative DHD domains lying downstream of *trans*-AT KS domains at KS/DH junctions were excised manually and aligned using ClustalX³⁸. Using this alignment file, a profile hidden Markov model (pHMM) was built from these sequences using the hmmbuild function of HMMER 3.1b2 (ref. 18), with default parameters.

Gene synthesis, cloning and mutagenesis. The following constructs were purchased from Epoch Life Sciences: pET24a_ACP-KS-DHD, pET24a_ACP-KS(Δ DHD) (residues 1–588 of the tri-domain construct; **Supplementary Fig. 22**), pET24a_ACP-KS(Δ 609–640) (residues 1–608 of the tri-domain construct; **Supplementary Fig. 23**) and pET24a_DHD (residues 571–640 of the tri-domain construct; **Supplementary Fig. 24**). The amplification of GbnD5 DH-ACP, GbnD5 DH and GbnD5 ACP from *B. gladioli* gDNA was performed with Phusion DNA polymerase (NEB). The primers used in each instance are as follows: DH-ACP_For(5'-CACCATGACTCATCGCCATGCA-3'), DH-ACP_Rev-(5'-TCATGCAGCCACCGATTTCGCT-3'); DH_For-(5'-CACCATGACTCATCGCCATGCA-3'), DH_Rev-5'-TCAGACGTGCGGCCAGCGGA-3'); ACP_{DH}_For(5'-ACCGTGGCGGCCGGTACGA-3'), ACP_{DH}_Rev(5'-CATGCAGCCACCGATTTCGCT-3'). PCR products were separated on a 1% agarose gel, and bands were excised and purified with a Gene Jet gel extraction kit (Thermo Scientific). The insert was ligated with pET151 (Invitrogen) following the manufacturer's instructions. The resulting vector was used to transform *E. coli* TOP10 cells (Invitrogen), which were plated on LB agar containing ampicillin (100 μ g/mL). Colonies were picked and grown overnight in LB medium. Plasmids were isolated from the culture using a miniprep kit (Thermo), and the inserts were sequenced to verify their integrity. The ACP-KS(C318A)-DHD mutant was constructed using the Q5 site-directed mutagenesis kit (NEB), with the following primers: ACP-KS(C318A)-DHD_For(5'-CGATACGGCCGCCAGCGCGC-3'), ACP-KS(C318A)-DHD_Rev(5'-ACCACCAGTTTCGGGCCG-3').

Protein overproduction and purification. A single colony of *E. coli* BL21(DE3) that had been transformed with the appropriate expression vector was picked and used to inoculate LB medium (10 mL) containing kanamycin (50 μ g/mL) or ampicillin (100 μ g/mL). The resulting culture was incubated overnight at 37 °C and 180 r.p.m., then used to inoculate LB medium (1 L) containing kanamycin (50 μ g/mL) or ampicillin (100 μ g/mL). The resulting culture was incubated at 37 °C and 180 r.p.m. until the optical density at 595 nm reached 0.6, then IPTG (1 mM) was added and growth was continued overnight at 15 °C and 180 r.p.m. The cells were harvested by centrifugation (4,000 \times g for 15 min at 4 °C) and resuspended in buffer (20 mM Tris-HCl, 100 mM NaCl, 20 mM Imidazole, pH 7.4) at 10 mL/L of growth medium, and then lysed using a Constant Systems cell disruptor. The lysate was centrifuged (37,000 \times g for 30 min at 4 °C), and the resulting supernatant was loaded onto a HiTrap Chelating Column (GE Healthcare), which had been pre-loaded with 100 mM NiSO₄ and equilibrated in resuspension buffer (20 mM Tris-HCl, 100 mM NaCl, 20 mM Imidazole, pH 7.4). Proteins were eluted in a stepwise manner using resuspension buffer containing increasing concentrations of imidazole: 50 mM (5 mL), 100 mM (3 mL), 200 mM (3 mL) and 300 mM (3 mL). The presence of the protein of interest in fractions was confirmed by SDS-PAGE, and an additional gel filtration step (Superdex 75/200, GE Healthcare) was used to further purify proteins when necessary. Fractions containing the protein of interest were pooled and concentrated to 130–300 μ M using a VivaSpin centrifugal concentrator at an appropriate MWCO (5 kDa for ACP/DHD domains, 30 kDa for ACP-KS-DHD/DH-ACP/DH constructs). Glycerol was added to the concentrated protein samples to a final concentration of 10% (v/v), and they were snap-frozen in liquid N₂ and stored at -80 °C.

Preparation of U-¹³C-, ¹⁵N-labeled DHD domain. The U-¹³C-, ¹⁵N-labeled DHD domain was obtained from *E. coli* BL21(DE3) cells that were first grown in 1 L of LB medium at 37 °C to an optical density (595 nm) > 1.0. The cells were pelleted by centrifugation at 4,000 \times g for 15 min and washed with PBS. After centrifuging again, the cells were resuspended in 500 mL of M9 medium

supplemented with ¹⁵NH₄Cl and [U-¹³C]glucose. After incubation for 2 h at 37 °C, expression was induced with 1 mM IPTG, and the culture was incubated at 18 °C overnight. U-¹⁵N-labeled DHD was produced analogously using unlabeled glucose instead of [U-¹³C]glucose in 1 L of M9 medium. Subsequent purification steps were conducted as described above, except that the protein was lyophilized for storage at -20 °C.

Acyl transfer assays. GbnD4 ACP-KS-DHD/ACP-KS(C318A)-DHD/ACP-KS(Δ DHD) (~200 μ M) and GbnD5 DH-ACP/BaeL DH-ACP (~130 μ M) proteins were converted to their acetylated and holo forms, respectively, by incubation in 20 mM Tris, 100 mM NaCl, 10% glycerol, 2 μ M Sfp, 400 μ M acetyl-CoA or CoA and 10 mM MgCl₂ at pH 7.4 in a total volume of 50 μ L. Excess acetyl-CoA/CoA-SH was removed by two successive ten-fold dilutions with 20 mM Tris, 100 mM NaCl, followed by concentration to 50 μ L using a VivaSpin centrifugal concentrator with a 30 kDa MWCO. Acyl transfer reactions were performed in 50 μ L total volume of 20 mM Tris, 100 mM NaCl. Acetyl-ACP-KS (100 μ M) was incubated with holo-DH-ACP (50 μ M) at room temperature for 6 h, then analyzed by UHPLC-ESI-Q-TOF-MS.

UHPLC-ESI-Q-TOF-MS analysis of intact proteins. All acyl-transfer assays were analyzed on a Bruker Maxis II ESI-Q-TOF-MS connected to a Dionex 3000 RS UHPLC fitted with an ACE C₄-300 RP column (100 \times 2.1 mm, 5 μ m, 30 °C). The column was eluted with a linear gradient of 5–100% MeCN containing 0.1% formic acid over 30 min. The mass spectrometer was operated in positive ion mode with a scan range of 200–3,000 *m/z*. Source conditions were: end plate offset at -500 V; capillary at -4,500 V; nebulizer gas (N₂) at 1.8 bar; dry gas (N₂) at 9.0 l min⁻¹; dry temperature at 200 °C. Ion transfer conditions were: ion funnel RF at 400 Vpp; multiple RF at 200 Vpp; quadrupole low mass at 300 *m/z*; collision energy at 8.0 eV; collision RF at 2,000 Vpp; transfer time at 110.0 μ s; pre-pulse storage time at 10.0 μ s.

Protein crosslinking. The chloroacrylamido-pantetheine crosslinker was synthesized according to literature procedures^{39,40} and attached to GbnD5 DH-ACP and BaeL DH-ACP using previously described methodology^{22,23}. Typically, loading reactions were carried out using either DH-ACP (130 μ M) or ACP (150 μ M) in 20 mM Tris, 100 mM NaCl, 10 mM MgCl₂ and 10% glycerol (total volume, 50 μ L) at pH 7.4. The loading reaction was initiated by addition of 2 μ M Sfp, 1 μ M CoaA, 1 μ M CoaD, 1 μ M CoaE and 500 μ M crosslinker, and was allowed to proceed for 2 h at 25 °C. Removal of excess crosslinker was achieved by two successive ten-fold dilutions with 20 mM Tris, 100 mM NaCl, followed by concentration using VivaSpin centrifugal concentrators with 30 kDa (DH-ACP) or 5 kDa (ACP) MWCOs. Loading reactions were monitored by UHPLC-ESI-Q-TOF-MS. Crosslinking reactions were performed in a total volume of 25 μ L of 20 mM Tris, 100 mM NaCl. Reactions were initiated by addition of apo-ACP-KS-DHD/ACP-KS(C318A)-DHD/ACP-KS(Δ DHD) (100 μ M) to GbnD5 DH-ACP/BaeL DH-ACP that had been loaded with the crosslinker (200 μ M). The reactions were allowed to proceed for 16 h at 25 °C before analysis by SDS-PAGE.

NMR experiments. U-¹³C-, ¹⁵N-labeled DHD domain was dissolved in 200 μ L of measurement buffer (50 mM potassium phosphate, 150 mM NaCl, 10% D₂O, pH 6.5) to obtain a protein concentration of 450 μ M. Spectra were acquired in a 3-mm tube at 280 K on a Bruker Avance 700 MHz spectrometer equipped with a TCI cryoprobe. TopSpin 3.2 was used to process the spectra, and the data were analyzed with CCPNMR⁴¹. Triple resonance 3D BEST-HNCACB, BEST-HNCA, BEST-HN(CO)CACB, BEST-HNCO, and BEST-HN(CA)CO⁴², were used to assign ¹H, ¹⁵N and ¹³C resonances (deposited at BMRB, accession code 27016). The obtained chemical shifts were compared to predicted random coil chemical shifts and used for neighbor-corrected secondary structure propensity calculations using the ncSSP server. To detect long range ¹H-¹H contacts, a ¹H-¹H NOESY spectrum with a mixing time of 100 ms was acquired. ¹⁵N longitudinal (R₁), ¹⁵N transverse (R₂) and {¹H}-¹⁵N heteronuclear NOE relaxation data were obtained using standard methods⁴³ and fitted using the CCPNMR rate analysis tool⁴¹. Relaxation delays from 0 to 1,000 ms and 316.8 ms were employed for R₁ and R₂ measurements. For relaxation rates, the in CCPNMR implemented

covariance method was chosen to estimate the s.d. The $\{^1\text{H}\}-^{15}\text{N}$ heteronuclear NOEs were obtained for ^1H -saturated and -unsaturated spectra (6 s saturation time; **Supplementary Tables 1–3**). For binding studies, $U\text{-}^{15}\text{N}$ -labeled DHD domain was dissolved in 200 μL of buffer (50 mM potassium phosphate, 150 mM NaCl, 10% D_2O , pH 7.4) to a final concentration of 80 μM . DH-ACP (415 μM) was added in 13 steps to obtain final concentrations of 1.0, 2.0, 4.0, 5.9, 7.8, 9.8, 13.5, 17.2, 24.4, 34.7, 47.6, 62.4, and 76.1 μM . A 2-D ^1H - ^{15}N BEST-TROSY-HSQC spectrum was acquired at each titration step. The data were analyzed using the CCPNMR rate analysis tool⁴¹ and Stadsisk (Triola Statistics) for statistical evaluation. Equation (1) defines the dissociation of the DHD domain (P) from DHD/DH-ACP complex (PL) at a single site on the DH-ACP di-domain (L). This is a valid description because the affinity of each residue in the DHD domain for the DH-ACP di-domain is examined individually. The sum of the concentrations of the bound and unbound states gives the total concentrations of L and P (equation (2))⁴⁴. By combining equations (1) and (2), and including a multiplier constant, A , which takes into account error in the integral of the first data point, equation (3) can be used to determine an approximate value of the dissociation constant (K_d) for each residue of the DHD domain in the DHD/DH-ACP complex. The decay in integral value for each signal as a function of DH-ACP concentration was fitted to equation (3). Regions affected by the interaction were defined as above the threshold, which was calculated as previously described⁴⁵. Briefly the average K_s ($K_s = 1/K_d$) and s.d. were calculated for all residues. Subsequently, residues with K_s values larger than the average plus 3 s.d. were removed and the average and s.d. were recalculated. This process was repeated iteratively until no additional residues with K_s values larger than the average plus 3 s.d. remained in the set. The threshold was set to the average plus the s.d. for the final set of residues. Errors were propagated using Monte Carlo error analysis. The signal decay curves were back-predicted from the best fit parameters, random noise within the estimated range of integral errors was added to the back-predicted points and the data were refitted. Error bars reflect two s.d. over 1,000 iterations of the described procedure. $[L]_t$ represents total ligand concentration and $[P]_t$ represents total protein concentration.

$$K_d = \frac{[P][L]}{[PL]} \quad (1)$$

$$[L]_t = [L] + [PL] \text{ and } [P]_t = [P] + [PL] \quad (2)$$

$$[P] = A \frac{(-K_d - [L]_t + [P]_t) + \sqrt{4K_d[P]_t + (K_d + [L]_t - [P]_t)^2}}{2[P]_t} \quad (3)$$

CD spectroscopic analysis of the DHD domain. CD spectra of the DHD domain (0.2 mg/mL dissolved in 50 mM potassium phosphate, 50 mM sodium chloride, pH 7.4) were recorded between 190 and 260 nm at room temperature on a JASCO J-1500 using a 1-mm path length quartz cuvette. The spectra were analyzed using DichroWeb⁴⁶ and fitted with the CDSSTR method using the reference data set SP 175 (ref. 47).

Carbene footprinting. *Interaction of GbnD5 DH and ACP domains:* A 60- μM solution of DH, ACP or DH-ACP in 10 μL of 20 mM Tris, 150 mM NaCl, pH 7.4 was mixed with an equal volume of a 10-mM solution of aryl diazirine photoreagent in the same buffer. The mixture was left to equilibrate for 5 min at room temperature before 6 μL aliquots were placed in crystal clear vials (Fisher Scientific UK) and snap-frozen in liquid nitrogen. The labeling reaction was initiated by photolysis of the mixture using the third harmonic of a Nd:YLF laser (Spectra Physics, repetition frequency 1,000 Hz, pulse energy 125 μJ) at a wavelength of 347 nm. The frozen samples were irradiated for 16 s. All experiments were performed in triplicate.

Interaction of the GbnD5 DH-ACP di-domain with the GbnD4 DHD domain: A solution of 20 μM DH-ACP in 20 mM Tris, 150 mM NaCl (10 μL total volume), pH 7.4 was mixed with an equal volume of a 40 μM DHD domain solution (or buffer for unbound DH-ACP experiments). A 100-mM solution of photoreagent in the same buffer was added to the mixture to a final concentration of 5 mM. The same labeling protocol described above was used to start the reaction. All experiments were performed in triplicate.

Following irradiation, samples were reduced (DTT, 10 mM in 10 mM ammonium bicarbonate), alkylated (iodoacetamide, 55 mM in 10 mM ammonium bicarbonate) and incubated at 37 °C with trypsin overnight (1:20 protease:protein ratio in 10 mM ammonium bicarbonate) in the presence of Protease MAX (Promega). The digestion was quenched by adjusting the pH to below 6 by adding formic acid to 2% final concentration.

LC-MS Analysis of tryptic digests from carbene footprinting. The analysis of the digests was carried out in load-trapping mode on a Dionex Ultimate 3000 Nano LC system using a C18 trapping column and a Pepmap C18 analytical column (Thermo Scientific, 75 μm , 150 mm, 5 μm particle size, 300 Å pore size). Peptides were eluted using a 30-min linear gradient of solvent B (water/acetonitrile 5/95 with 0.1% formic acid) from 0% to 55%, followed by 5 min at 90% B and a 20-min re-equilibration segment to 100% A (water/acetonitrile 95/5 with 0.1% formic acid) at a flow rate of 0.3 $\mu\text{L}/\text{min}$. The HPLC system was coupled to a Thermo Scientific LTQ FT Ultra mass spectrometer equipped with a nanoelectrospray ionization source. A 1.7 kV voltage was applied to a coated PicoTip emitter (New Objective). The capillary temperature was set at 275 °C, with inner capillary voltage value set on 37 V and tube lens value of 145 V. Spectra were acquired in positive ion mode for a 400–2,000 m/z range at a nominal resolution of 100,000. Identification of peptides was conducted in data-dependent mode. The five most intense ions for each scan were isolated and subjected to CID in the linear ion trap using a nominal energy of 35.0. Signals with +1 charge state were rejected. The data were searched against a custom FASTA database including the DH, DHD and ACP sequences using Bioworks software (Thermo Fisher Scientific). Peptide tolerance was set at 2.0 Dalton, fragment ion tolerance at 0.8, and two missed cleavage sites were allowed. Addition of the photoreagent was included as a variable modification on all residues and carbamoylation of cysteine was set as fixed modification. Targeted CID experiments to locate and quantitate the modification at residue level were performed at a nominal energy of 15.0. The precursors were isolated within a window of 8 Th and the activation time was set at 30 ms.

Analysis of LC-MS data from carbene footprinting. *Peptide level analysis:*

A previously described method was used to quantitate the fraction of each peptide modified²⁷. Briefly, the chromatograms for each singly labeled and unlabeled peptide were extracted within a range of $\pm 0.1 m/z$, and the spectrum for each peak was manually inspected to ensure the sampling of the correct ion only. The peptide fractional modification was calculated using equation (4).

$$P = \frac{A_{\text{labeled}}}{A_{\text{labeled}} + A_{\text{unlabeled}}} \quad (4)$$

Where A_{labeled} and $A_{\text{unlabeled}}$ correspond to the peak area of each labeled and unlabeled peptide, respectively. Differences in the extent of labeling between peptides were considered significant when the P value obtained from a two-tailed unpaired Student's t -test was <0.05 . Peptides that showed significant differences between the different states of the proteins were subjected to MS/MS experiments to locate and quantitate the modification at the residue level.

Residue level analysis: Modification sites on the peptides that showed significant differences at the peptide level were investigated using tandem MS. The fractional modification (f.mod) on the n_i residue was calculated using equation (5).

$$\text{f.mod}(n_i) = \frac{I(n_i^{\text{labeled}})}{I(n_i^{\text{labeled}}) + I(n_i^{\text{unlabeled}})} \quad (5)$$

In which $I(n_i^{\text{labeled}})$ and $I(n_i^{\text{unlabeled}})$ correspond to the measured intensity of the fragment n_i in its labeled and unlabeled versions, respectively. The difference in fractional modification between two consecutive fragments, multiplied by the fractional amount of labeling on the peptide gives the absolute level of modification on the n_i residue using equation (6).

$$\text{abs.mod.} = P[\text{f.mod.}(n_i) - \text{f.mod.}(n_{i-1})] \quad (6)$$

In the case of the impossibility of detection of a n_i ion, its fractional modification was grouped together with its subsequent fragment. The few fractional modification values of residues that showed an anomalous increase or decrease in respect to the fractional modification curve trend were excluded from the data set and were grouped together with the value corresponding to the adjacent residue resulting in only a small loss of resolution. Per-residue modification values from each repeat were averaged and the averages compared between the two states of the protein. Differences in per-residue modification were considered significant when the P value obtained from a two-tailed unpaired Student's t -test was <0.05 .

Homology modeling and docking. Homology modeling of GbnD5 DH (using entry 4LN9 from the PDB as a template) and ACP domains was conducted using the I-TASSER server⁴⁸, and further refinement of the model was achieved using the MolProbity server⁴⁹. Docking of the ACP domain was carried out using the ZDOCK server⁵⁰ with residue constraints on the DH and ACP domains obtained from the carbene footprinting analysis. The residue constraints are detailed in the legend of **Figure 4**.

Life sciences reporting summary. Further information on experimental design and reagents is available in the **Life Sciences Reporting Summary**.

Data availability. NMR data from this study are deposited in the Biological Magnetic Resonance Data Bank with accession code [27016](#). All other data are included in the published article and its supplementary information files, or are available from the authors upon reasonable request.

38. Larkin, M.A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
39. Worthington, A.S. & Burkart, M.D. One-pot chemo-enzymatic synthesis of reporter-modified proteins. *Org. Biomol. Chem.* **4**, 44–46 (2006).
40. Meier, J.L., Mercer, A.C., Rivera, H. Jr. & Burkart, M.D. Synthesis and evaluation of bioorthogonal pantetheine analogues for in vivo protein modification. *J. Am. Chem. Soc.* **128**, 12174–12184 (2006).
41. Vranken, W.F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687–696 (2005).
42. Lescop, E., Schanda, P. & Brutscher, B. A set of BEST triple-resonance experiments for time-optimized protein resonance assignment. *J. Magn. Reson.* **187**, 163–169 (2007).
43. Palmer, A.G., III. NMR characterization of the dynamics of biomacromolecules. *Chem. Rev.* **104**, 3623–3640 (2004).
44. Williamson, M.P. Using chemical shift perturbation to characterise ligand binding. *Prog. Nucl. Magn. Reson. Spectrosc.* **73**, 1–16 (2013).
45. Schumann, F.H. *et al.* Combined chemical shift changes and amino acid specific chemical shift mapping of protein-protein interactions. *J. Biomol. NMR* **39**, 275–289 (2007).
46. Whitmore, L. & Wallace, B.A. DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res.* **32**, W668–W673 (2004).
47. Lees, J.G., Miles, A.J., Wien, F. & Wallace, B.A. A reference database for circular dichroism spectroscopy covering fold and secondary structure space. *Bioinformatics* **22**, 1955–1962 (2006).
48. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat. Methods* **12**, 7–8 (2015).
49. Chen, V.B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
50. Pierce, B.G. *et al.* ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics* **30**, 1771–1773 (2014).