Published in final edited form as:
Nat Chem. 2019 October ; 11(10): 913–923. doi:10.1038/s41557-019-0335-5.

Structural basis for chain release from the enacyloxin polyketide synthase

Simone Kosol[#1], **Angelo Gallo**[#1], **Daniel Griffiths**[#1], **Timothy R. Valentic**[#2], **Joleen Masschelein**[1], **Matthew Jenner**[1,3], **Emmanuel L.C. de los Santos**[3], **Lucio Manzi**[4], **Paulina K. Sydor**[1], **Dean Rea**[5], **Shanshan Zhou**[1], **Vilmos Fülöp**[5], **Neil J. Oldham**[4], **Shiou-Chuan Tsai**[2], **Gregory L. Challis**[1,3,6,*], **Józef R. Lewandowski**[1,*]

[1]Department of Chemistry, University of Warwick, Coventry CV4 7AL, UK

[2]Departments of Molecular Biology and Biochemistry, Chemistry, and Pharmaceutical Sciences, University of California, Irvine, CA 92697, USA

[3]Warwick Integrative Synthetic Biology Centre, University of Warwick, Coventry CV4 7AL, UK

[4]School of Chemistry, University of Nottingham, Nottingham NG7 2RD, UK

[5]School of Life Sciences, University of Warwick, Coventry CV4 7AL, UK

[6]Department of Biochemistry and Molecular Biology, Monash University, Clayton, VIC 3800, Australia

[#] These authors contributed equally to this work.

Summary

Modular polyketide synthases and nonribosomal peptide synthetases are molecular assembly lines consisting of several multienzyme subunits that undergo dynamic self-assembly to form a functional mega-complex. N- and C-terminal docking domains are usually responsible for

[*]Authors to whom the correspondence should be addressed. j.r.lewandowski@warwick.ac.uk and g.l.challis@warwick.ac.uk.

Data availability

The structures of the Bamb_5917 PCP domain and Bamb_5915 are available from the PDB (accession IDs: 5MTI and 6CGO, respectively). NMR assignments for the *apo* and *holo*-Bamb_5917 PCP domain are available from the BMRB (Accession IDs: 34085 and 27304, respectively). Raw NMR and BLI data can be obtained from http://wrap.warwick.ac.uk/123013/. The remaining data supporting the findings of this study are included in the supporting information or are available from the corresponding authors on request. All biological materials are available from the authors on request.

Author contributions

S.K., A.G., D.G. and T.V. contributed equally to this work. J.R.L., G.L.C., S.K., A.G. and D.G. conceived and designed the experiments. S.K., D.G., J.M. and S.Z. designed primers and generated constructs for protein expression, expressed and purified the proteins and performed biochemical assays. S.K., A.G., D.G. and J.R.L. performed and analysed NMR experiments. A.G. calculated NMR structures. P.K.S., D.R., V.F., T.V. and S.T. crystallised Bamb_5915, and T.V. and S.T. solved its structure. D.G., E.S., S.K. and J.R.L. performed the bioinformatics analyses. M.J., L.M. and N.J.O. performed and analysed carbene footprinting. A.G. and J.R.L. performed and analysed MD and docking simulations. J.R.L., G.L.C., S.K., D.G., A.G. and T.V. wrote the paper. All authors discussed the results and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Europe PMC Funders Author Manuscripts

mediating interactions between subunits. Here we show that communication between two nonribosomal peptide synthetase subunits responsible for chain release from the enacyloxin polyketide synthase, which assembles an antibiotic with promising activity against *Acinetobacter baumannii*, is mediated by an intrinsically disordered short linear motif and a β-hairpin docking domain. The structures, interactions and dynamics of these subunits are characterised using several complementary biophysical techniques, providing extensive insights into binding and catalysis. Bioinformatics analyses reveal that short linear motif/β-hairpin docking domain pairs mediate subunit interactions in numerous nonribosomal peptide and hybrid polyketide-nonribosomal peptide synthetases, including those responsible for assembling several important drugs. Short linear motifs and β-hairpin docking domains from heterologous systems are shown to interact productively, highlighting the potential of such interfaces as tools for biosynthetic engineering.

Introduction

Type I modular polyketide synthases (PKSs) and nonribosomal peptide synthetases (NRPSs) are remarkable molecular machines that assemble natural products with a wide range of applications in medicine and agriculture.[1] They typically consist of multiple protein subunits containing one or more catalytic domains that must interact with each other in a tightly defined manner to ensure biosynthetic fidelity. These interactions are frequently mediated by docking domains appended to the N- and/or C-termini of the subunits.[2] One such example is the β-NRPS systemng (βHD) domain, which has been found attached to the N-terminus of NRPS subunits in a handful of hybrid PKS-NRPS systems.[3] βHD domains interact with short linear motifs (SLiMs) appended to the C-terminus of upstream PKS subunits, facilitating the recognition of acyl groups attached to an ACP domain by condensation (C) or heterocyclisation (Cy) domains in the NRPS subunit. The structures of excised βHD domains from the tubulysin and rhabadopetide assembly line as well as a covalently-linked SLiM-βHD domain complex from rhabadopetide assembly line have been determined by NMR spectroscopy[2,4]. In addition, the X-ray crystal structure of a βHD-Cy di-domain from the EpoB NRPS subunit of the epothilone assembly line has been reported.[10]

In the accompanying manuscript,[5] we report that the PKS responsible for the assembly of the acyl chain in enacyloxin IIa, an antibiotic with promising activity against *Acinetobacter baumannii*,[6] employs an unusual dual transacylation mechanism for product release. This involves transfer of the fully-assembled polyketide chain from the acyl carrier protein (ACP) domain of the final PKS module (Bamb_5919) to a peptidyl carrier protein (PCP) domain in Bamb_5917 (Fig. 1a). Subsequent condensation of the polyketide chain with (1*S*, 3*R*, 4*S*)-3, 4-dihydroxycyclohexane carboxylic acid (DHCCA) is catalysed by Bamb_5915, an enzyme with sequence similarity to NRPS C domains (Fig. 1a). Transacylation of the polyketide chain from the Bamb_5919 ACP domain to the Bamb_5917 PCP domain is catalysed by a non-elongating ketosynthase (KS^0) domain appended to the C-terminus of Bamb_5919 (Fig. 1a). This is necessitated by the inability of Bamb_5915 to utilise acyl donors when they are attached to the Bamb_5919 ACP domain, indicating that specific protein-protein interactions facilitate the recognition of substrates bound to the Bamb_5917 PCP domain by Bamb_5915.Here we report extensive structural studies of Bamb_5915 and the Bamb_5917 PCP domain, which illuminate the molecular basis for their interaction. We show that a

SLiM-βHD domain pair plays a key role in their productive association and provide new mechanistic insights into the conformational changes involved in binding and catalysis. Bioinformatics analyses reveal that SLiM-βHD domain pairs, previously identified in only a handful of systems, are widespread in NRPS and hybrid NRPS-PKS assembly lines, and likely facilitate the interaction of carrier protein domains with a range of catalytic domains. The productive interaction of a SLiM and a βHD domain from distantly related taxa illustrates the potential to exploit SLiM-βHD domain interactions for the creation of novel biosynthetic pathways.

Results and Discussion

X-ray crystal structure of Bamb_5915

The 56 kDa Bamb_5915 protein consists of a C-terminal C domain and an N-terminal βHD domain (Fig. 1b & Supplementary Fig. 1-4; PDB ID: 6CGO). The C domain is a V-shaped pseudodimer with each subdomain harbouring the chloramphenicol acyltransferase fold[7] and the conserved active site HHxxxD motif[8] at their central interface (Fig. 1b). A phosphate ion is bound in the active site of the protein. A 30 Å channel runs through the centre of the C domain, allowing access to the active site from two separate faces (Supplementary Fig. 5). The entrance near the binding site for the Bamb_5917 PCP domain (the acyl donor entrance) is located at the anterior of the C domain, in accord with other NRPS C domain structures (Supplementary Figs 6 and 7). A second active site entrance is located at the posterior of the C domain and is defined by a groove between α-helix 6 of the N-terminal subdomain and the β-sheet of the C-terminal subdomain (Supplementary Fig. 5). Although the Bamb_5915 C domain structure is similar to other C domains, two features make it unique. First, the channel entrance adjacent to the canonical acyl acceptor PCP domain binding site (the acceptor side entrance) is obstructed by α-helix 4, and the loops between β-strands 13 and 14 (cyan in Fig. 1b) and β-strand 12 and α-helix 9 (magenta in Fig. 1b; Supplementary Fig. 8 and Movie 1). Thus, either the acyl acceptor is delivered through the secondary posterior entrance, or a conformational change occurs at the acceptor entrance to allow substrate entry into the active site. Second, the “latch region” (V419-S454) in the C-terminal subdomain does not donate a β-strand to the β-sheet in the N-terminal subdomain as in other C domain structures, but instead forms a largely unstructured loop (Supplementary Fig. 9). The “latch” β-strand observed in most other C domain structures has been suggested to disengage from the N-terminal subdomain during either acyl donor binding or product release[9]. The Bamb_5915 crystal structure provides the first observation of a “latch” in a disengaged state.

An N-terminal βHD domain (green in Fig. 1b) with αββαα topology is appended to the C domain, similar to crystal structures of the heterocyclization domain from the EpoB subunit of the epothilone PKS-NRPS[10]. According to a Dali[11] search, the Bamb_5915 βHD domain has high structural homology (Supplementary Fig. 10), but less than 35% sequence similarity to the βHD domains from EpoB[10] and the TubC subunit of the tubulysin PKS-NRPS[12]. In the structures of the EpoB βHD-Cy di-domain and Bamb_5915 both the βHD domains and the linker regions connecting them to the downstream catalytic domains adopt distinct orientations (Supplementary Fig. 11). This suggests that the inter-domain linker is

flexible,[3] allowing βHD domains to access multiple conformational states, which may facilitate PCP recognition.

Solution NMR spectroscopic analysis of the Bamb_5917 PCP domain

The structure of the 10.7 kDa *apo*-PCP domain of Bamb_5917 (i.e. without the substrate-carrying phosphopantetheine (Ppant) prosthetic group) forms a four-helix bundle typical of such carrier proteins (Fig. 1c and Supplementary Fig. 12; PDB ID: 5MTI). The stable conformation of the Bamb_5917 PCP domain resembles the A/H state of the TycC3 PCP domain[16] (Supplementary Fig. 12), also observed in most X-ray and NMR structures of carrier proteins[17]. The disordered termini of the PCP domain are highly mobile on the fast ps-ns timescale as indicated by low heteronuclear nuclear Overhauser effects (NOEs), transverse relaxation rates (R_2), and high longitudinal relaxation rates (R_1) for these regions compared to the structured core (Fig. 1d and Supplementary Fig. 13). The overall rotational diffusion[13] is consistent with a monomeric state, with a compact folded core and disordered termini (Supplementary Materials and Methods). The dynamic, disordered nature of the terminal regions is further confirmed by the low predicted S^2 parameters (Fig. 1d and Supplementary Fig. 14; S^2 of 1 and 0 indicate complete rigidity and unrestricted motion, respectively), typical chemical shift values for a random coil and a lack of long-range NOEs. S^2 predictions suggest that the full-length Bamb_5917 protein exhibits similar dynamic behaviour. A few localised changes in ^{15}N chemical shifts between the *holo*- and *apo*-forms of the Bamb_5917 PCP domain (BMRB 27304 & 34085; Supplementary Fig. 15), observed only near the Ppant attachment site, suggest that these two forms have nearly identical major conformations with similar dynamics on the fast ps-ns time scale (Supplementary Fig. 13). Notably, the C-terminus, which contains a docking domain, shows dynamic behaviour typical of an intrinsically disordered region (IDR).

Interactions between Bamb_5915 and the Bamb_5917 PCP domain

In the NMR titrations of the Bamb_5917 *holo*-PCP domain with both full-length Bamb_5915 and its excised βHD domain the system is in slow to intermediate exchange[14] (Supplementary Fig. 16). Higher than average local apparent association constants, K_a, in the titrations with the excised βHD domain identify a Short Linear Motif (SLiM; S307-R315)[11] in the disordered C-terminus of Bamb_5917 as the major interaction site (Fig. 2a-b). Other regions appreciably affected by binding of the βHD domain are located at the N-termini of α-helices 1 and 2, and loop 1 which connects these two helices (Fig. 2b). While the perturbation of adjacent surface residues (A231 and T230 in α-helix 1) and I225 may indicate a secondary interaction interface with the βHD domain, most of the residues near loop 1 have buried side chains in direct contact with each other (Supplementary Fig. 17) suggesting conformational change upon binding as a more likely interpretation of the observed effects.

In the titrations of the Bamb_5917 PCP domain with full-length Bamb_5915 (Fig. 2c), in addition to the regions identified by the titrations with the excised βHD domain, surface residues near the C-terminus of α-helix 1 (A238 and Q234) and the N-terminus of α-helix 2 (Q256 and A263) also have above average local apparent K_as. These likely report on interactions with the catalytic domain of Bamb_5915. Indicators of conformational change

near loop 1 extend further into the hydrophobic core (e.g. L229 on α-helix 1 and L287 on α-helix 4) compared to the experiment with the isolated βHD domain.

Binding of the SLiM to the βHD domain is further confirmed by the proportional increase in ^{15}N R_2 values for the SLiM and the observation of a minor state in ^{15}N Chemical Exchange Saturation Transfer (CEST)[15](exchange rate of 0.58 ± 0.02 s^{-1} and minor population of ~9%) when Bamb_5915 is present in solution (1:5 and 1:10 molar ratio; Supplementary Figs 18 and 19). Typically, caution needs to be exercised in quantifying local apparent K_as from the rate of disappearance of signals in NMR titrations (Supplementary Figs 20 and 21)[14]. However, because the K_as obtained from signal growth in the titration with the excised βHD domain (Supplementary Fig. 22) are consistent with $K_a > 10^6$ M^{-1} ($K_d < 1$ μM) and the K_a for binding of the Bamb_5917 PCP domain to Bamb_5915 determined using biolayer interferometry is $14x10^5 \pm 4x10^5$ M^{-1} ($K_d = 0.7 \pm 0.2$ μM; Supplementary Fig. 23), it appears that the K_as obtained from peak disappearance for residues in the intrinsically disordered SLiM are dominated by binding events.

We also conducted carbene footprinting[16] to assess changes in the solvent accessibility of different residues as a result of binding of the Bamb_5917 PCP domain to Bamb_5915 (Supplementary Figs 24-26; raw data in Supplementary File 1). Masking indicative of a decrease in solvent accessibility upon binding is observed for peptides from the disordered C-terminus and α-helices 1 and 2 of the PCP domain (Fig. 2d). Unmasking of peptides from α-helices 3 and 4 is also observed (Fig. 2d) suggesting an increase of solvent accessibility for these regions as a result of a conformational change in the PCP domain upon binding.

From the NMR titrations and carbene footprinting, we can identify three main regions of the PCP domain that interact directly with Bamb_5915: the C-terminal SLiM (S307-R315) and α-helices 1 and 2. The unmasking of peptides from α-helices 3 and 4 of the PCP domain is consistent with conformational changes in these regions, rather than direct binding, as indicated by the NMR titrations.

The carbene footprinting experiments aid identification of interaction sites and conformational changes in Bamb_5915. Digested peptides masked as a result of binding include α-helix 1 of the βHD domain, and terminal residues in α-helices 6, 10, 12 and 13 (Fig. 2). The region masked in the βHD domain is adjacent to, and does not overlap with, the SLiM- βHD interface identified by NMR titrations for the isolated βHD domain from the tubulysin system (Supplementary Fig. 27). Interestingly, no statistically significant masking is observed for the β-hairpin, which is the expected binding interface for the SLiM[4,12]. One possible explanation for the lack of statistically significant masking of the β-hairpin could be its “flapping” motion observed in MD simulations, resulting in increased solvent exposure. If binding of the SLiM stabilizes the β-hairpin in a more open conformation, the two effects, binding and conformational change, would partially counteract each other. The masking of peptides from α-helices 10, 12 and 13 identifies a surface patch that could be an interaction site for the globular part of the PCP domain. The masked A167-R175 peptide from α-helix 6 is remote from the expected PCP binding site and is likely linked to a conformational change in Bamb_5915 upon binding. Finally, unmasking of the C-terminal part of the “latch” region

(peptide A432-I460) suggests that Bamb_5915 undergoes a conformational change upon binding the Bamb_5917 PCP domain that leads to solvent exposure of this region.

Further insights into conformational changes of Bamb_5917 PCP upon binding

Binding of Bamb_5915 to the Bamb_5917 PCP domain does not seem to be a simple rigid body docking but rather involves conformational changes in both proteins. A conformational switch has been observed for the *apo* and *holo*-forms of the TycC3 PCP domain,[17] but the relevance of this observation to assembly line function has been challenged[18].

To characterise the conformation of the *holo*-Bamb_5917 PCP domain-Bamb_5915 complex, which is invisible in solution due to size-related severe resonance broadening, we turned to 60-100 kHz magic angle spinning solid-state NMR spectroscopy[19]. The *de novo* assignment of the signals in a sedimented[20,21] 70 kDa complex of the *holo*-Bamb_5917 PCP domain with Bamb_5915 was prevented by limited resolution and sensitivity but a set of tentative self-consistent assignments was obtained by comparing the signals in solution and solid state spectra. We found that several characteristic signals for the bound form of the PCP domain in the sedimented complex appear to be in approximately the same position as in the solution NMR titrations with the excised βHD domain (assignment of these resonances in solution was made using the chemical shifts of the minor bound state in ^{15}N CEST), e.g. R315 and V284 (Supplementary Fig. 28). The residues in the SLiM and most residues in the globular part of the PCP domain, are largely immobilised in the Bamb_5917 PCP:Bamb_5915 complex, as evidenced by the appearance of their signals in the dipolar coupling based correlation spectra (Supplementary Fig. 28). Most of the residues in the PCP domain for which the C^{α} resonances are shifted by more than 1 ppm, indicating conformational change, face the interior of the protein (Fig. 2e). The observed pattern supports the idea that the hydrophobic core of the PCP domain rearranges upon binding to Bamb_5915. This rearrangement involves loop 1, α-helix 1 and α-helix 3. In addition, the C^{α} signal for V284, which has a chemical shift more consistent with a random coil than an α-helix (61.8 ppm in the complex compared to 67.7 ppm in the free form; random coil value for Val is 62.1[22]), suggests that this conformational change is accompanied by partial unwinding of α-helix 4 at its N-terminus.

The conformational changes suggested by the solid-state NMR data and carbene footprinting experiments appear largely consistent with the overall conformational changes found in 0.5 μs accelerated molecular dynamics[23] (aMD) simulation of the *holo*-Bamb_5917 PCP domain. The rearrangement of the hydrophobic core in the simulation involves rotation and translation of α-helix 1 and a large conformational change in loop 1 (Fig. 3c and Supplementary Movie 5). The motion of loop 1 leads to increased solvent accessibility at the C-terminus of α-helix 3 and the N-terminus of α-helix 4, consistent with the unmasking observed in the carbene footprinting.

Further insights into conformational changes of Bamb_5915 upon binding

Molecular dynamics (MD) simulations[20] shed light on the conformational changes of Bamb_5915. One of the most notable conformational changes observed in a standard 100 ns MD simulation of Bamb_5915 is modulation of the degree of association between the N-

and C-terminal subdomains (Supplementary Fig. 29), which has been proposed for other condensation domains to have a potentially significant role in catalysis[9,24,25]. Interestingly, this overall conformational change appears to be correlated with changes to the accessibility of the active site entrances. In particular, an entrance on the acyl acceptor side is created when the degree of association between the N- and C-terminal subdomains increases (Supplementary Fig. 29). This implies that the unusual DHCCA acyl acceptor employed by Bamb_5915 can access the active site through the same route as in C domains that use more conventional aminoacyl-PCP domains as acyl acceptors (Supplementary Fig. 8). To capture slower large-scale motions, we also performed a 1 μs aMD simulation on Bamb_5915 (Fig. 3) [23,26]. Hydrogen bonds are formed between residues from the "latch" region of the C-terminal subdomain (G424 and G423) and residues in the N-terminal subdomain (P97 and V99, respectively) in this simulation (Fig. 3a), resulting in a conformation similar to that observed in other structures of C domains, but with a smaller interface between the N- and C-terminal subdomains. The "donation" of the "latch" to the N-terminal subdomain appears to stabilise the high degree of association between the C- and N-terminal subdomains and consequently the open state of the acyl acceptor entrance (Supplementary Movie 2). The C domain appears to operate a pair of interlocked shutter valves: the size of the acyl acceptor entrance is inversely correlated with the size of the opening on the acyl donor side. The extent of the opening depends on the conformations of the "latch" and the loop between β-strand 12 and α-helix 9 for the acyl acceptor entrance and the "latch" and "floor loop" region[25] (loop between β-strands 11 and 12, in yellow in Figs 3b and 1b) for the acyl donor entrance. The corresponding regions are highly variable between different condensation domains and may contribute to the substrate specificity of the enzymes by influencing the state of the active site entrances.

Assuming that binding of the Bamb_5917 PCP domain stabilises Bamb_5915 in conformations similar to the ones observed in the aMD simulation, we can explain the results of the carbene footprinting experiments for regions with changing solvent accessibility that are remote from the expected binding site. For example, opening of the acyl acceptor channel leads to increased solvent accessibility of the C-terminal portion of the "latch", consistent with the unmasking of that region (Supplementary Fig. 30). Moreover, opening of this channel is accompanied by partial burial of α-helix 6 (orange in Fig. 3b; Supplementary Fig. 31), and masking of this region (Fig. 2d). The aMD simulations also suggest that masking of α-helix 12 can be, at least partially, attributed to conformational change rather than binding (Supplementary Fig. 32). Stabilisation of the "open acceptor channel conformation" of Bamb_5915 by binding of the donor PCP domain, suggests that the acyl acceptor may not be able to access the active site until the acyl donor is bound. However, further experiments will be required to test this hypothesis.

Examination of the structures of Bamb_5915 homologues, including condensation, epimerisation, heterocyclisation and X domains, suggests that the correlation between the degree of association of the C- and N-terminal subdomains and the accessibility of the acyl donor and acceptor active site entrances may be a general feature (Supplementary Fig. 33). In examples with a low degree of association between the N- and C-terminal subdomains the acyl acceptor channel is closed and the acyl donor channel is open, whereas in cases where

the degree of subdomain association is high the converse is observed. In intermediate conformational states both entrances are open.

Interaction model for the Bamb_5917 PCP-Bamb_5915 complex

We used High Ambiguity Driven biomolecular DOCKing (HADDOCK) to produce a series of models of the dynamic Bamb_5917 PCP domain:Bamb_5915 complex[27,28](Supplementary Fig. 34 & 35). We used the X-ray crystal structure of Bamb_5915 and conformations sampled from the aMD simulations of the *holo*-Bamb_5917 PCP domain as the starting point. Conformational changes in the PCP domain are required to position the thiol group of the Ppant arm close to the highly conserved H205 active site residue. One model of the complex employing a conformation that allows the Ppant-bound substrate to be positioned close to H205 is shown in Fig. 3d. The distance between the thiol group of the Ppant arm and the conserved active site H205 residue is ~5 Å. The orientation of the Bamb_5917 PCP domain in this model differs from those observed for the PCP domains in the crystal structures of the GrsA PCP-E and TqaA PCP-C_T di-domains (Supplementary Fig. 36).[29,30] The R420 side chain appears to stabilise the “donated latch” conformation, in a similar way to His and Asn residues in the structure of the TqaA PCP-C_T di-domain (Supplementary Fig. 37).[29]

SLiM-βHD domain interactions are important for chain release

Using (1*S*, 3*R*, 4*S*)-3-amino-4-hydroxycyclohexane-1-carboxylic acid (AHCCA) as an acyl acceptor,[5] we assessed the ability of Bamb_5915 to off-load the acetyl group from a Bamb_5917 PCP domain construct lacking the C-terminal SLiM. Relative to the full-length acetylated Bamb_5917 PCP domain, the amount of N-acetyl-AHCCA produced in this reaction was substantially reduced (Fig. 4a-d). This indicates that the SLiM plays an important role in recruitment of the PCP domain to the C domain. We were unable to conduct the converse experiment in which the ability of full length Bamb_5915 to catalyse chain release is compared to its isolated C domain, because Bamb_5915 constructs lacking the βHD domain were insoluble in *E. coli*. We thus investigated the ability of the isolated βHD domain to inhibit the condensation reaction. A decrease in the quantity of N-acetyl-AHCCA formed was observed as the concentration of the isolated βHD domain was increased (Fig. 4e and Supplementary Fig. 38). Analogous results were obtained when a synthetic SLiM peptide was used in place of the isolated βHD domain (Fig. 4f and Supplementary Fig. 38). Together, these data confirm that the interaction of the SLiM with the βHD domain plays an important role in chain release from the enacyloxin PKS. However, because residual product formation is observed in all of these experiments, it is clear that the globular portion of the PCP domain can engage in productive protein-protein interactions with the C domain. It thus appears that the purpose of the SLiM-βHD domain interaction is to increase the effective concentration of the acyl donor.

SLiM-βHD domain pairs are found in numerous assembly lines

We used the sequences of the Bamb_5915, TubC[12] and EpoB[10] βHD domains to create a profile hidden Markov model (pHMM). The MIBIG[31] database was searched using this pHMM and the hits were used to refine the model (Fig. 5a). To identify the SLiM partners of these βHD domains, we first assumed co-linearity between the order of subunits in assembly

lines and the order of the genes encoding them. We then identified subunits upstream of those with an N-terminal βHD domain, which contain C-terminal sequences that do not conform to pHMMs for known PKS and NRPS domains, and are predicted by IUPRED to be disordered[32]. In these systems the disordered C-terminal regions are rich in charged residues (typically glutamate, but other charge patterns are observed; Fig. 5a), as observed for the SLiMs that interact with the Bamb_5915, TubC and EpoB βHD domains. In a few cases, we manually examined gene clusters to identify those encoding a protein with a C-terminal SLiM that don't conform to the co-linearity assumption. Using this procedure, >80 SLiM-βHD domain pairs were identified in the MIBIG database,[31] with some assembly lines containing multiple examples. Among these were 58 pathways for known metabolites, including several clinically-used antibiotics and anticancer agents (Fig. 5b and Supplementary Table 13). Several subclasses of SLiM-βHD domain pairs were identified, based on the charge pattern of the SLiM and the type of catalytic domain connected to the βHD domain, for example dual epimerisation/condensation (C/E) domains, which are associated with a characteristic SLiM (Supplementary Fig. 39).

When this procedure was applied to the Genbank database[32] >1400 SLiM-βHD domain pairs were identified (Supplementary File 1), including additional pathways for known metabolites not in the MIBiG database (Supplementary Table 14). This shows that such subunit interfaces are not only much more prevalent than previously appreciated, but also mediate communication between a wide range of catalytic domains. SLiMs were identified at the C-termini of oxidase and Cy domains, in addition to ACP and PCP domains (Fig. 5c), and βHD domains were found not only at the N-termini of C and Cy domains, but also C/E domains, thioesterases, thioester reductases (including cases where a β-lactamase-like domain is fused to the N-terminus of the βHD domain), halogenases, methyltransferases and other proteins of unknown function (Fig. 5c). In some cases, a single SLiM appears to recruit multiple catalytic subunits to the same PCP domain, e.g. in the cyanopeptolin NRPS a C-terminal SLiM in McnC likely interacts with N-terminal βHD domains in NRPS and halogenase subunits encoded by *mcnE* and *mcnD*, respectively.

Heterologous SLiM-βHD domain pairs are compatible

The abundance and diversity of SLiM-βHD domain interfaces in biosynthetic assembly lines suggests it may be possible to exploit heterologous interfaces for the production of novel natural product hybrids. As an initial test of feasibility, we examined whether Bamb_5915 could interact productively with a PCP domain from an NRPS subunit (Sven_0512) in *Streptomyces venezuelae* involved in watasemycin assembly (Fig. 6)[34]. The SLiM appendend to this PCP domain is very similar to that in Bamb_5917. The pattern of negatively charged and hydrophobic residues are largely preserved, but it lacks the C-terminal Arg residue (Fig. 6e; Supplementary Fig. 40). In its native context, this SLiM interacts with a βHD domain (~35% sequence identity with the Bamb_5915 βHD domain; Supplementary Fig. 41) appended to a Cy domain. In NMR titrations of the Bamb_5917 PCP domain with the excised βHD domain from Sven_0517, residues in the SLiM were observed to interact strongly (Supplementary Fig. 42-43). Incubation of the acetylated Sven_0512 PCP domain with Bamb_5915 and AHCCA led to the formation of *N*-acetyl-AHCCA (Fig. 6a), demonstrating that SLiMs and βHD domains from heterologous systems

can cross-talk productively. The important role played by the interaction of the heterologous SLiM with the βHD domain is highlighted by the progressive reduction in product formation when increasing concentrations of isolated Sven_0512 SLiM or Bamb_5915 βHD domain are added to the reaction mixtures (Fig. 6c,d).

Conclusions

Previous structural studies have focused on excised βHD domains and the juxtaposition of such domains to downstream Cy domains[2–4]. The structural relationship between the SLiM interaction partners for βHD domains and their upstream carrier protein domains has hitherto not been investigated. Our work shows that the SLiM in Bamb_5917 forms part of a larger IDR appended to the C-terminus of the PCP domain. Together with our recent work on another type of IDR-embedded SLiM, which facilitates subunit interaction in *trans*-AT PKSs by binding to the surface of a dehydratase (DH) domain[35], our results suggest that IDRs may be more widely exploited for inter-subunit communication in biosynthetic assembly lines than previously thought.

A combination of structural, biophysical and computational analyses of the Bamb_5917 PCP-SLiM di-domain and Bamb_5915 enabled us to generate extensive new insights into interactions across SLiM-βHD domain interfaces. In particular, we established that both partners undergo conformational changes upon binding, which may play an important role in C domain catalysis by controlling sequential access of substrates to the active site. Compared to subunits from other systems that communicate via SLiM-βHD domain pairs, Bamb_5917 and Bamb_5915 are relatively small. This, coupled with the fact that the Bamb_5915 C domain utilizes a soluble, rather than carrier protein bound acyl acceptor, and tolerates a range of acyl donors,[5] makes Bamb_5917/Bamb_5915 an attractive model system for future efforts aimed at exploring the specificity and engineer-ability of the SLiM-βHD domain interface.

The pHMM for βHD domains developed in this work reveals that SLiM-βHD domain pairs mediate subunit communication in a far wider range of NRPS and NRPS-PKS systems than previously appreciated. Such systems include the assembly lines for bacitracin, romedepsin, bleomycin and epothilone, all of which are approved antibacterial or anticancer drugs, in addition to cryptophicin, an analogue of which reached phase II clinical trials, but was withdrawn due to side effects. The fact that the Sven_0512 PCP-SLiM di-domain is able to interact productively with the Bamb_5915 βHD-C di-domain is remarkable in light of the divergent taxonomic origin (Gram-positive *Actinobacteria* and Gram-negative *Betaprotobacteria*, respectively) and biosynthetic nature of the watasemycin and enacyloxin assembly lines. This finding highlights the potential for SLiM-βHD domain interfaces to be exploited in biosynthetic engineering approaches to the production of novel natural product derivatives with enhanced therapeutic potential. Our pHMM also revealed that βHD domains are fused to a far wider range of enzymes than canonical NRPS C/Cy domains. Coupled with the ability of SLiMs and βHD domains from heterologous systems to recognize each other, this suggests it may be possible to harness SLiM-βHD domain interactions for the *de novo* creation of biosynthetic pathways to metabolic products hitherto unknown in Nature.

Europe PMC Funders Author Manuscripts

Methods summary

Materials and methods are detailed in the supplementary information.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013; ERC Grant Agreement 639907) supported this research. J.R.L. acknowledges funding from the Royal Society (RG130022), the EPSRC (EP/L025906/1), the BBSRC (BB/L022761/1 and BB/R010218/1), and the Gates Foundation (OPP1160394). The European Commission (Marie Sklodowska-Curie Fellowship; contract no. 656067) and the Research Foundation Flanders funded J.M. G.L.C. acknowledges the BBSRC (BB/L021692/1 and BB/K002341/1) and the Royal Society (Wolfson Research Merit Award WM130033) for funding. The University of Warwick funded P.K.S. through an Institute of Advanced Study fellowship. D.G. and S.Z. were supported by the EPSRC through the Centre for Doctoral Training in Molecular Analytical Science (EP/L015307/1) and the Bridging the Gaps - EPS and AMR initiative (EP/M027503/1), respectively. E.S. is a Research Career Development Fellow in the Warwick Integrative Synthetic Biology Centre supported by the BBSRC and EPSRC (BB/M017982/1). We acknowledge the FP7 WeNMR (261572) and H2020 West-Life (675858) European e-Infrastructure projects for the use of their web portals, which make use of the EGI infrastructure and DIRAC4EGI service with the dedicated support of CESNET-MetaCloud, INFN-PADOVA, NCG-INGRID-PT, RAL-LCG2, TW-NCHC, IFCA-LCG2, SURFsara and NIKHEF, and the additional support of the national GRID Initiatives of Belgium, France, Italy, Germany, the Netherlands, Poland, Portugal, Spain, UK, South Africa, Malaysia, Taiwan and the US Open Science Grid. We thank Andrew Marsh for providing access to the workstation used for the aMD simulation of Bamb_5915. Molecular graphics were generated using UCSF Chimera and Chimera X, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, supported by the NIH (P41-GM103311 and R01-GM129325). We thank Guillaume Bouvignies for assistance with ChemEx.

References

1. Cilia E, Pancsa R, Tompa P, Lenaerts T, Vranken WF. The DynaMine webserver: predicting protein dynamics from sequence. Nucleic Acids Research. 2014; (42)W1:W264, W270.
2. Xu W, Qiao K, Tang Y. Structural analysis of protein–protein interactions in type I polyketide synthases. Crit Rev Biochem Mol Biol. 2013; 48:98–122. [PubMed: 23249187]
3. Weissman KJ, Müller R. Protein-protein interactions in multienzyme megasynthetases. Chembiochem. 2008; 9:826–48. [PubMed: 18357594]
4. Hacker C, et al. Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. Nat Commun. 2018; 9
5. Masschelein J, et al. A dual transacylation mechanism for polyketide synthase chain release in enacyloxin antibiotic biosynthesis. Nat Chem.
6. Mahenthiralingam E, et al. Enacyloxins are products of an unusual hybrid modular polyketide synthase encoded by a cryptic Burkholderia ambifaria Genomic Island. Chem Biol. 2011; 18:665–77. [PubMed: 21609847]
7. Leslie AGW. Refined crystal structure of type III chloramphenicol acetyltransferase at 1·75 Å resolution. J Mol Biol. 1990; 213:167–186. [PubMed: 2187098]
8. De Crécy-Lagard V, Marlière P, Saurin W. Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. C R Acad Sci III. 1995; 318:927–36. [PubMed: 8521076]
9. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen L-O. Structural and Functional Insights into a Peptide Bond-Forming Bidomain from a Nonribosomal Peptide Synthetase. Structure. 2007; 15:781–792. [PubMed: 17637339]
10. Dowling DP, et al. Structural elements of an NRPS cyclization domain and its intermodule docking domain. Proc Natl Acad Sci. 2016; 113:12432–12437. [PubMed: 27791103]

11. Norman NE, Van Roey K, Weatheritt RJ, Toedt G, Uyar B, Altenberg B, Budd A, Diella F, Dinkel H, Gibson TJ. Dali server: conservation mapping in 3D. Mol. Biosyst. 2012; 8(1)
12. Richter CD, Nietlispach D, Broadhurst RW, Weissman KJ. Multienzyme docking in hybrid megasynthetases. Nat Chem Biol. 2008; 4:75–81. [PubMed: 18066054]
13. García de la Torre J, Huertas ML, Carrasco B. HYDRONMR: Prediction of NMR Relaxation of Globular Proteins from Atomic-Level Structures and Hydrodynamic Calculations. J Magn Reson. 2000; 147:138–146. [PubMed: 11042057]
14. Williamson MP. Using chemical shift perturbation to characterise ligand binding. Prog Nucl Magn Reson Spectrosc. 2013; 73:1–16. [PubMed: 23962882]
15. Vallurupalli P, Bouvignies G, Kay LE. Studying “Invisible” Excited Protein States in Slow Exchange with a Major State Conformation. Journal of the American Chemical Society. 2012; 134(19):8148–8161. [PubMed: 22554188]
16. Manzi L, et al. Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. Nat Commun. 2016; 7
17. Koglin A, et al. Conformational switches modulate protein interactions in peptide antibiotic synthetases. Science. 2006; 312:273–6. [PubMed: 16614225]
18. Lohman JR, et al. The crystal structure of BlmI as a model for nonribosomal peptide synthetase peptidyl carrier proteins. Proteins Struct Funct Bioinforma. 2014; 82:1210–1218.
19. Lamley JM, et al. Solid-state NMR of a protein in a precipitated complex with a full-length antibody. J Am Chem Soc. 2014; 136:16800–6. [PubMed: 25381931]
19. Bertini I, et al. Solid-state NMR of proteins sedimented by ultracentrifugation. Proc Natl Acad Sci U S A. 2011; 108:10396–9. [PubMed: 21670262]
21. Mainz A, Jehle S, van Rossum BJ, Oschkinat H, Reif B. Large Protein Complexes with Extreme Rotational Correlation Times Investigated in Solution by Magic-Angle-Spinning NMR Spectroscopy. J Am Chem Soc. 2009; 131:15968–15969. [PubMed: 19839609]
22. Wishart DS. Interpreting protein chemical shift data. Prog Nucl Magn Reson Spectrosc. 2011; 58:62–87. [PubMed: 21241884]
23. Hamelberg D, Mongan J, McCammon JA. Accelerated molecular dynamics: A promising and efficient simulation method for biomolecules. J Chem Phys. 2004; 120:11919–11929. [PubMed: 15268227]
24. Bloudoff K, Rodionov D, Schmeing TM. Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. J Mol Biol. 2013; 425:3137–3150. [PubMed: 23756159]
25. Bloudoff K, Schmeing TM. Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: Discovery, dissection and diversity. Biochim Biophys Acta - Proteins Proteomics. 2017; 1865:1587–1604. [PubMed: 28526268]
26. Bisht NK, et al. Ligand migration and hexacoordination in type 1 non-symbiotic rice hemoglobin. Biochim Biophys Acta. 2011; 1814:1042–53. [PubMed: 20940062]
27. Van Zundert GCP, et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. J Mol Biol. 2016; 428:720–725. [PubMed: 26410586]
28. Wassenaar TA, et al. WeNMR: Structural Biology on the Grid. J Grid Comput. 2012; 10:743–767.
29. Zhang J, et al. Structural basis of nonribosomal peptide macrocyclization in fungi. Nat Chem Biol. 2016; 12:1001–1003. [PubMed: 27748753]
30. Chen WH, Li K, Guntaka NS, Bruner SD. Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases. ACS Chem Biol. 2016; 11:2293–2303. [PubMed: 27294598]
31. Medema MH, et al. Minimum Information about a Biosynthetic Gene cluster. Nat Chem Biol. 2015; 11:625–631. [PubMed: 26284661]
32. Dosztanyi 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. 2005; 21:3433–3434. [PubMed: 15955779]
33. Agarwala R, et al. Database Resources of the National Center for Biotechnology Information. Nucleic Acids Res. 2017; 45:D12–D17. [PubMed: 27899561]

34. Inahashi Y, et al. Watasemycin biosynthesis in Streptomyces venezuelae: thiazoline C-methylation by a type B radical-SAM methylase homologue. Chem Sci. 2017; 8:2823–2831. [PubMed: 28553520]

35. Jenner M, et al. Mechanism of intersubunit ketosynthase–dehydratase interaction in polyketide synthases. Nat Chem Biol. 2018; 14:270–275. [PubMed: 29309054]

Europe PMC Funders Author Manuscripts

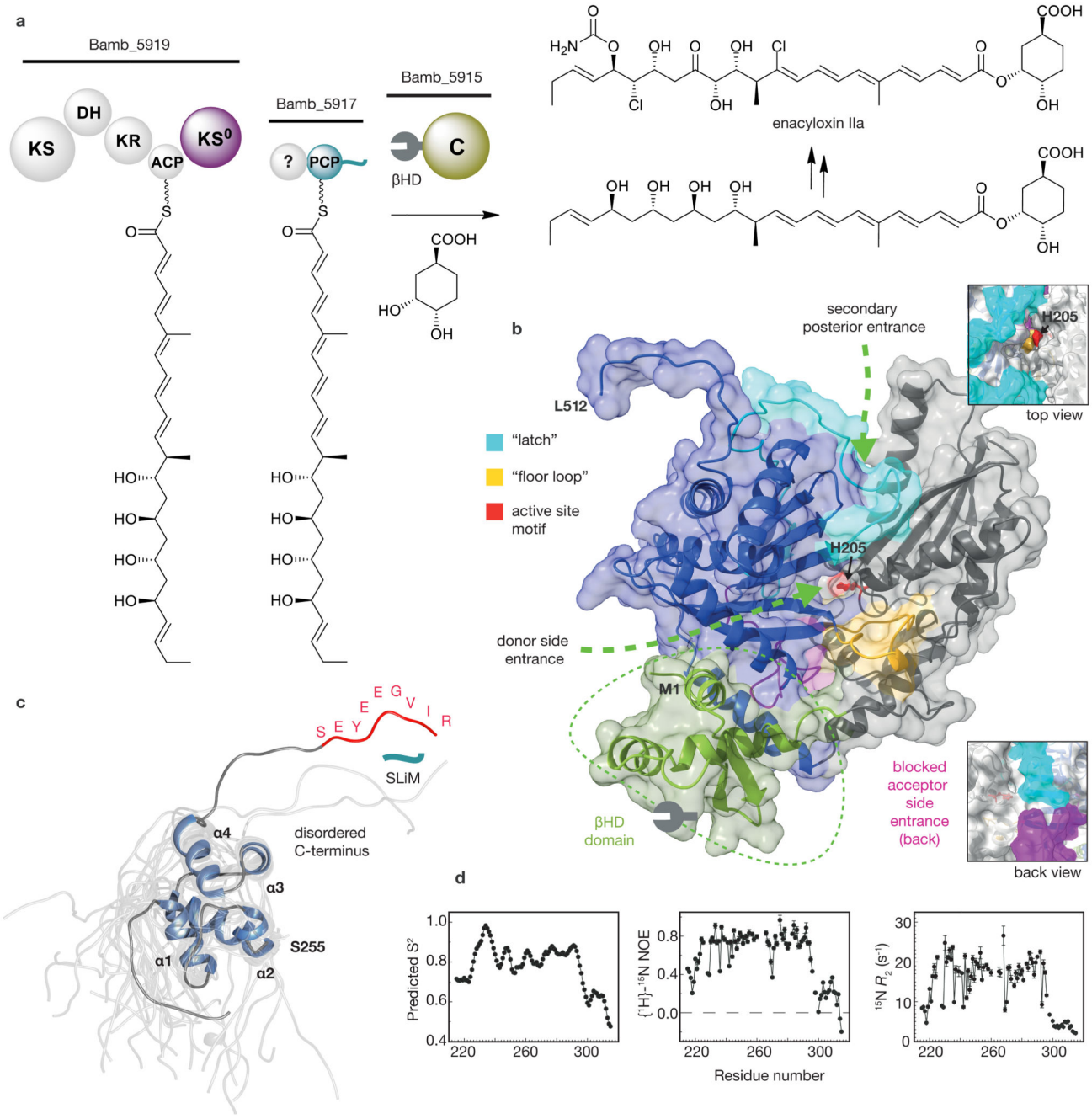

Figure 1. Chain release in enacyloxin biosynthesis

a, Mechanism for chain release from the enacyloxin PKS. Bamb_5917 has 214 amino acids appended to the N-terminus of the PCP domain (marked with ? in Fig. 1a), which appear to be a non-functional remnant of a KS domain that is not involved in chain release (as demonstrated by enzymatic assays[5]). **b**, X-ray crystal structure of Bamb_5915. The βHD domain is coloured green, and the C- and N-terminal subdomains are coloured blue and grey, respectively. The "latch" region is cyan, the "floor loop" is yellow and the active site H205 residue is red. The loops blocking the acceptor entrance to the solvent channel are

magenta and cyan. The inset panels show views of the secondary posterior and blocked acceptor side entrances. **c**, Solution NMR structure of the Bamb_5917 PCP domain. **d**, Dynamine[1] order parameter (S^2) predictions, and $^{15}N–\{^1H\}$ NOE and ^{15}N R_2 relaxation measurements for the *holo*-Bamb_5917 PCP domain.

Europe PMC Funders Author Manuscripts

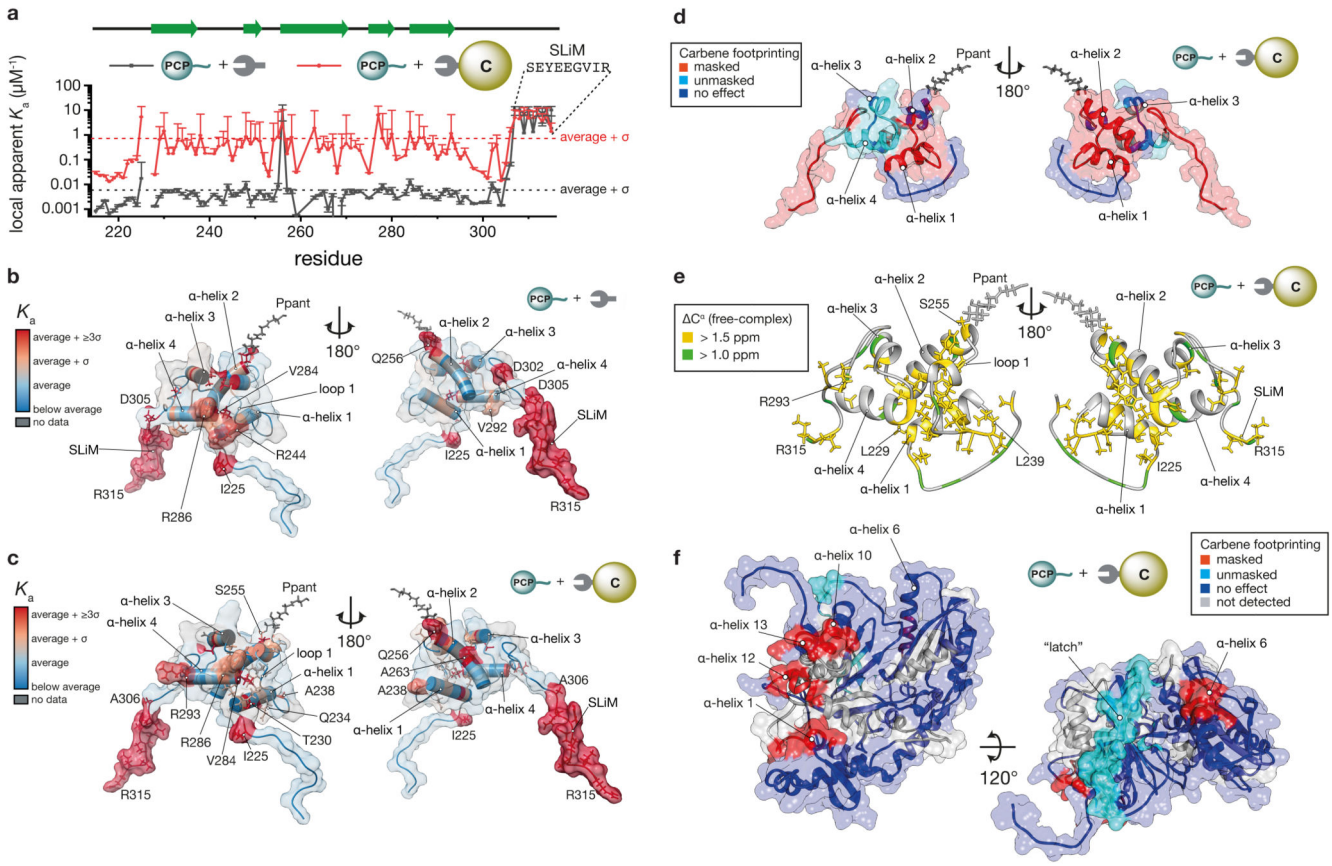

Figure 2. Interactions between the Bamb_5917 PCP domain and Bamb_5915

a, Local apparent K_a values as a function of residue obtained from NMR titrations of the Bamb_5917 PCP domain with Bamb_5915 and the excised βHD domain. The error bars represent two standard deviations over 1000 iterations of Monte Carlo error analysis. **b**, Residues with local apparent K_a values higher than the average in the NMR titration with the excised Bamb_5915 βHD domain. **c**, Residues with local apparent K_a values higher than the average in the NMR titration with full length Bamb_5915. **d**, Changes in solvent accessibility of residues in the Bamb_5917 PCP domain upon binding to Bamb_5915 determined by carbene footprinting mass spectrometry. Red, blue and cyan indicate, respectively, peptides that are masked (i.e. less solvent exposed), unchanged, and unmasked (i.e. more solvent exposed) as a result of complex formation. **e**, $^{13}C^{\alpha}$ chemical shift changes between the Bamb_5917 PCP domain in solution and in a sedimented complex with Bamb_5915, reflecting changes in local conformation in the complex. **f**, Changes in solvent accessibility of residues in Bamb_5915 upon binding to the Bamb_5917 PCP domain determined by carbene footprinting mass spectrometry. The colour key is the same as in **d**. The peptides for which proteolysis products were not detected are shown in grey.

Europe PMC Funders Author Manuscripts

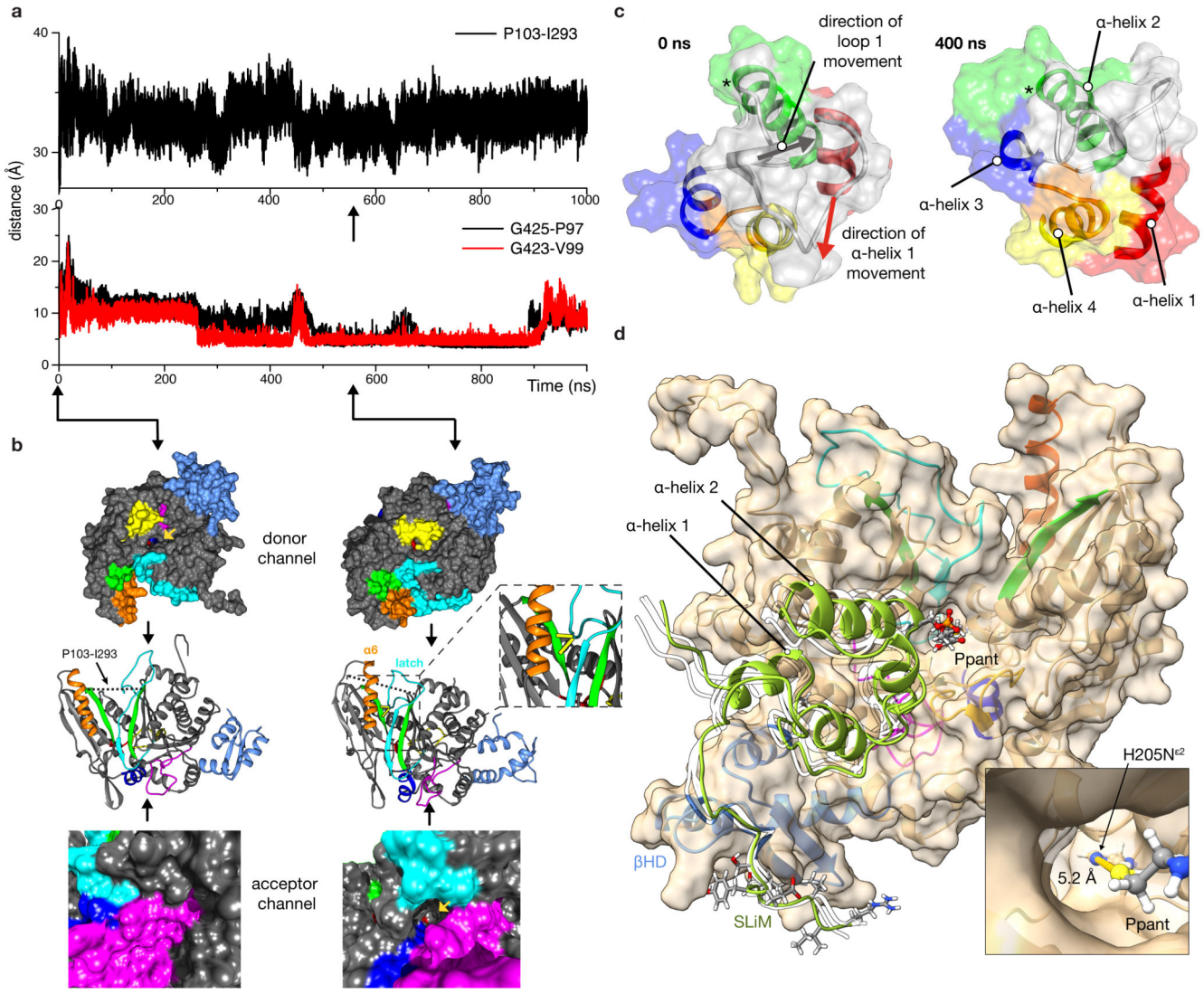

Figure 3. Accelerated Molecular Dynamics (aMD) simulations of Bamb_5915 and the Bamb_5917 PCP domain, and model of the Bamb_5917_PCP domain:Bamb_5915 complex.

a, Distances between centres of mass of residues used to track conformation of Bamb_5915 during 1 μs aMD simulations. P103-I293 reports on the degree of association between C- and N-terminal subdomains. G425-P97 and G423-V99 report on the distances from the N-terminal subdomain β-strand to residues on the C-terminal subdomain latch. **b**, Conformations of Bamb_5915 with closed (left) and open (right) acceptor channels observed in 1 μs aMD simulations. The conformation with the open acceptor channel is stabilised by donation of the latch to the N-terminal subdomain. Residues in the active site motif are red, the βHD domain is light blue, the latch is cyan, β-strands 3 and 9 are green, the floor loop is yellow and α-helix 6 is orange. The yellow lines indicate hydrogen bonds between the latch (cyan) and β-strand 3 (green). **c**, Frames at 0 and 400 ns from 0.5 μs aMD simulations, illustrating the overall conformational changes of the *holo*-Bamb_5917 PCP domain. α-Helices 1, 2, 3, 4 are coloured red, green, blue and yellow, respectively. Red and grey arrows indicate the approximate direction of movement of α-helix 1 and loop 1,

respectively. The Ppant (not shown) is attached to the Ser residue indicted with *. **d**, A model of the *holo*-Bamb_5917 PCP domain:Bamb_5915 complex based on docking the Bamb_5915 X-ray crystal structure and the structure of the *holo*-Bamb_5917 PCP domain observed in a 180 ns frame from the aMD simulations. The distance of 5.2 Å from the Ppant thiol to the active site H205 residue in the model of the complex suggests that conformational changes similar to the ones observed in the aMD simulation of Bamb_5915 are required for the condensation reaction to occur. Additional conformers of the Bamb_5917 PCP domain from the cluster of best solutions from the docking simulation are shown as silhouettes.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

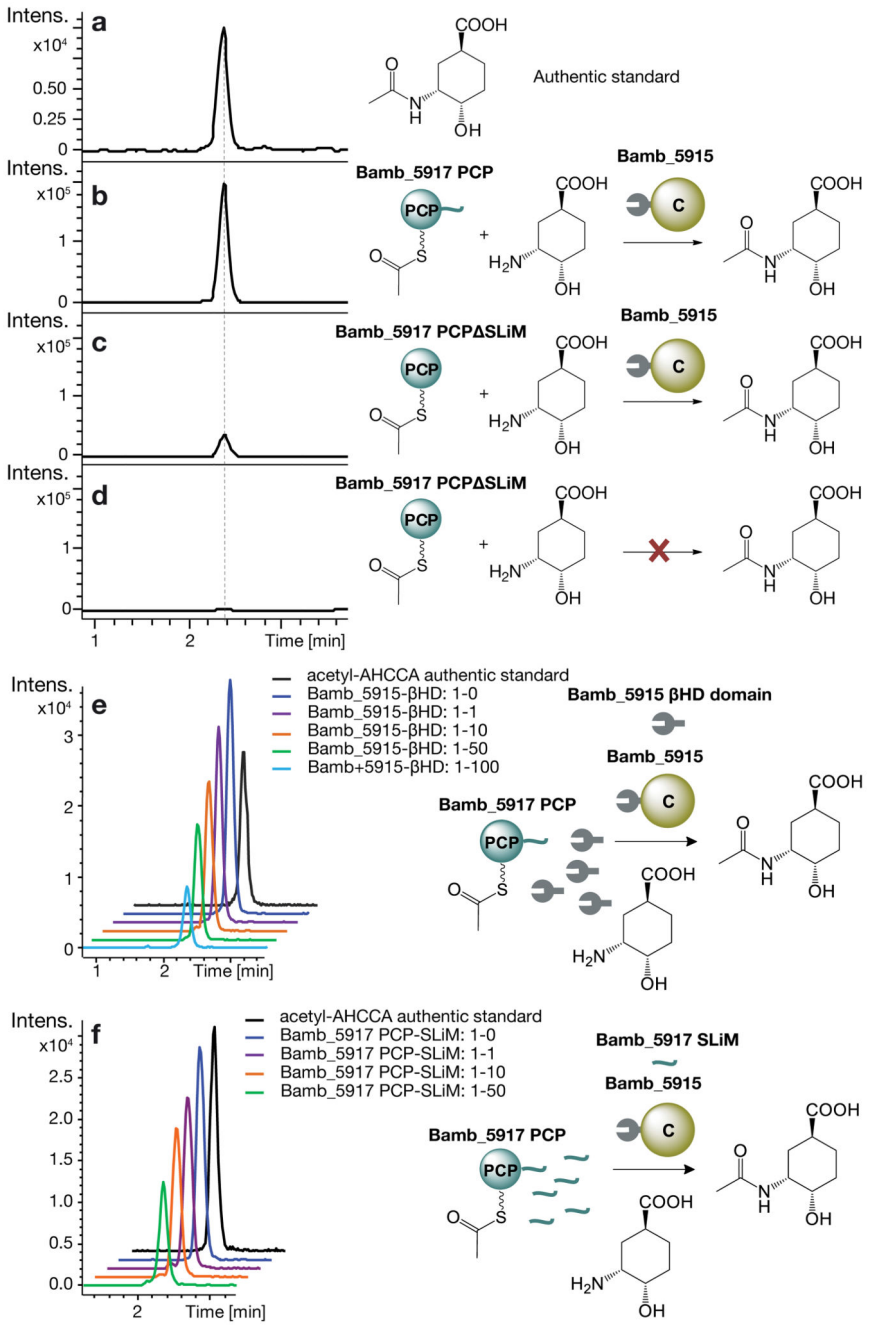

Figure 4. Extracted ion chromatograms from UHPLC-ESI-Q-TOF-MS analyses of chain release reactions

a, chromatogram for the authentic standard of N-acetyl-AHCCA. **b-c**, chromatograms for Bamb_5915 catalysed N-acetylation of AHCCA with the acetylated Bamb_5917 PCP domain (**b**), or the acetylated Bamb_5917 PCP domain lacking the SLiM (Bamb_5917 PCPΔSLiM) (**c**). **d**, chromatogram for a control reaction from which Bamb_5915 was omitted. **e-f**, chromatograms for competition assays for the same experimental set up as in

(**b**) but with increasing amounts of the excised Bamb_5915 βHD domain (**e**) or Bamb_5917 SLiM peptide (**f**). All reactions were performed in triplicate.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

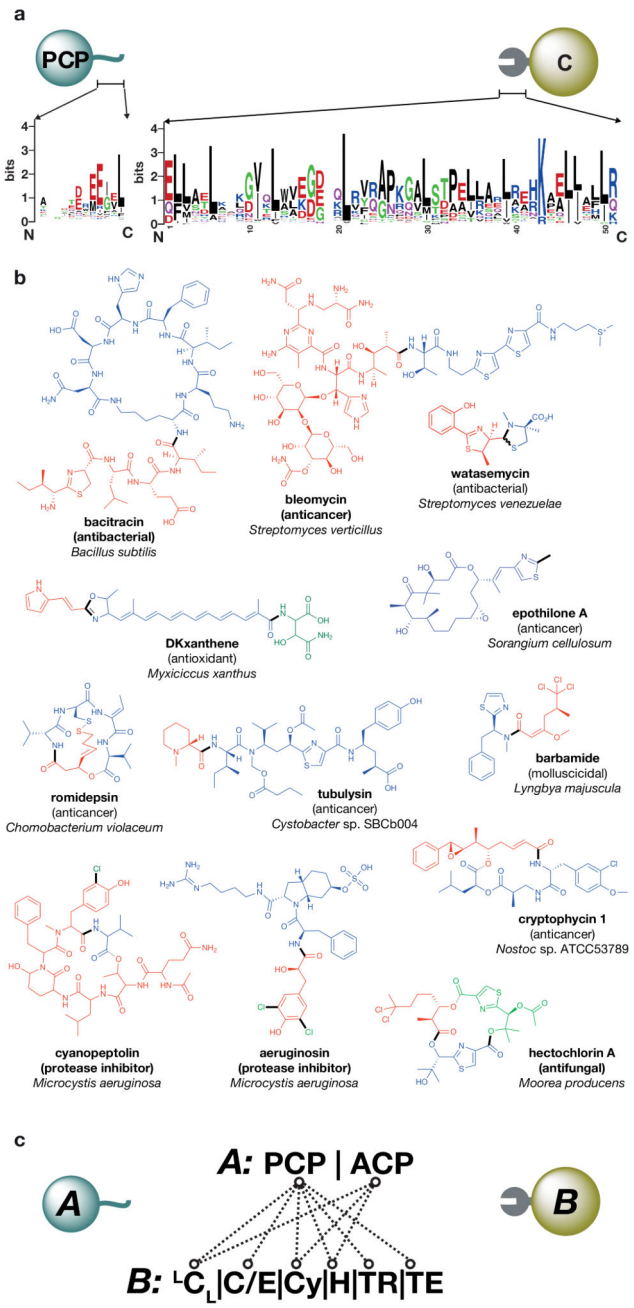

Figure 5. Bioinformatics analyses reveal that SLiM-βHD domain pairs mediate subunit interactions in numerous assembly lines

a, Sequence logos for SLiMs and βHD domains from all the hits in the MIBIG database. **b**, Examples of metabolites produced by assembly lines containing SLiM-βHD domain boundaries. Bonds in black are formed in enzymatic reactions dependent on SLiM-βHD domain communication. **c**, Examples of domain types between which the communication was found to be facilitated by SLiM-βHD domain pairs. Only examples with multiple hits in the database are included. PCP – peptidyl carrier protein, ACP – acyl carrier protein, $^{L}C_{L}$ – condensation domain accepting L-configured donor and acceptor aminoacyl thioesters, C/E

– bifunctional condensation-epimerisation domain, Cy – heterocylisation domain, H-halogenase, TR – thioester reductase, TE – thioesterase.

Europe PMC Funders Author Manuscripts

Europe PMC Funders Author Manuscripts

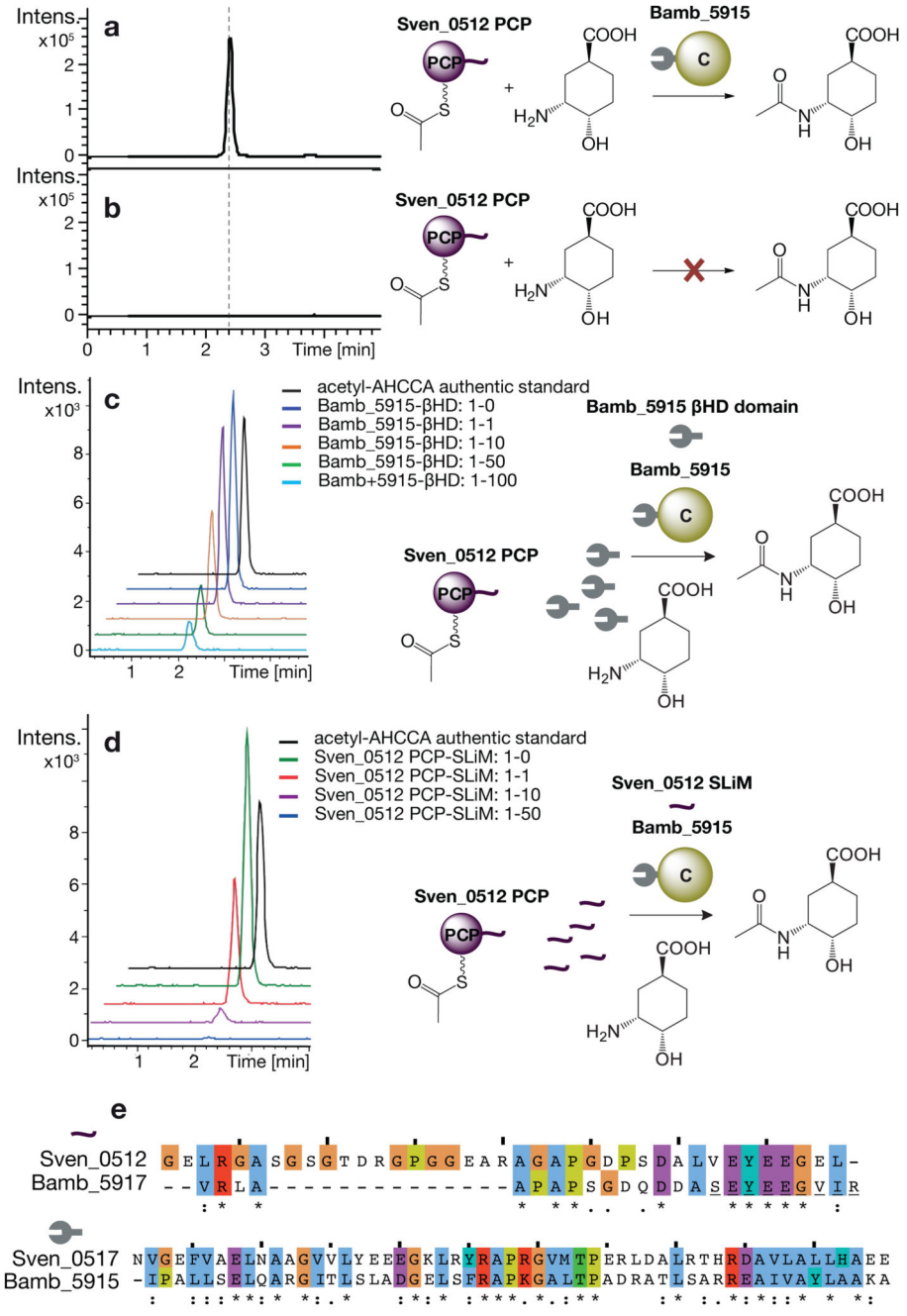

Figure 6. Heterologous SLiMs and βHD domains can interact productively

a, Bamb_5915 C domain-catalysed N-acetylation of AHCCA with the acetylated Sven_0512 PCP domain (40% sequence identity with the Bamb_5917 PCP domain) illustrating that the enacyloxin and watasemycin assembly lines can cross-talk. **b**, Control reaction in the absence of Bamb_5915. **c-d**, Competition assays for same experimental set up as in a but with increasing amounts of the excised Bamb_5915 βHD domain (**c**) or Sven_0512 SLiM peptide (**d**). **e**, Sequence alignment of SLiMs and βHD domains (35% sequence identity)

from the enacyloxin and watasemycin assembly lines. The underlined residues represent the sites in the Bamb_5917 SLiM that interact with the Bamb_5915 βHD domain.

Europe PMC Funders Author Manuscripts