

# Communication Breakdown: Dissecting the COM Interfaces between the Subunits of Nonribosomal Peptide Synthetases

Christopher D. Fage,\* Simone Kosol, Matthew Jenner, Carl Öster, Angelo Gallo, Milda Kaniusaite, Roman Steinbach, Michael Staniforth, Vasilios G. Stavros, Mohamed A. Marahiel, Max J. Cryle, and Józef R. Lewandowski\*



Cite This: *ACS Catal.* 2021, 11, 10802–10813



Read Online

ACCESS |



Metrics & More



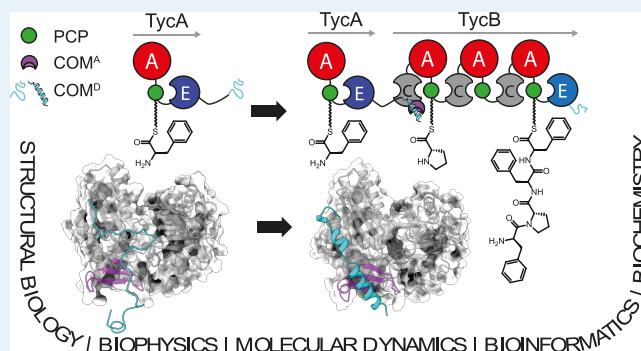
Article Recommendations



Supporting Information

**ABSTRACT:** Nonribosomal peptides are a structurally diverse and bioactive class of natural products constructed by multidomain enzymatic assembly lines known as nonribosomal peptide synthetases (NRPSs). While the core catalytic domains and even entire protein subunits of NRPSs have been structurally elucidated, little biophysical work has been reported on the docking domains that promote interactions—and thus transfer of biosynthetic intermediates—between subunits. In the present study, we closely examine the COM domains that mediate COMmunication between donor epimerization (E) and acceptor condensation (C) domains found at the termini of NRPS subunits. Through a combination of X-ray crystallography, circular dichroism spectroscopy, solution- and solid-state NMR spectroscopy, and molecular dynamics (MD) simulations, we provide direct evidence for an intrinsically disordered donor COM region that folds into a dynamic helical motif upon binding to a suitable acceptor. Furthermore, our NMR titration and carbene footprinting experiments illuminate the residues involved at the COM interaction interface, and our MD simulations demonstrate folding consistent with experimental data. Although our results lend credence to the previously proposed helix-hand mode of interaction, they also underscore the importance of viewing COM interfaces as dynamic ensembles rather than single rigid structures and suggest that engineering experiments should account for the interactions which transiently guide folding in addition to those which stabilize the final complex. Through activity assays and affinity measurements, we further substantiate the role of the donor COM region in binding the acceptor C domain and implicate this short motif as readily transposable for noncognate domain crosstalk. Finally, our bioinformatics analyses show that COM domains are widespread in natural product pathways and function at interfaces beyond the canonical type described above, setting a high priority for thorough characterization of these docking domains. Our findings lay the groundwork for future attempts to rationally engineer NRPS domain–domain interactions with the ultimate goal of generating bioactive molecules.

**KEYWORDS:** biosynthesis, docking domain, intrinsically disordered protein, natural product, nonribosomal peptide synthetase, protein–protein interaction



## INTRODUCTION

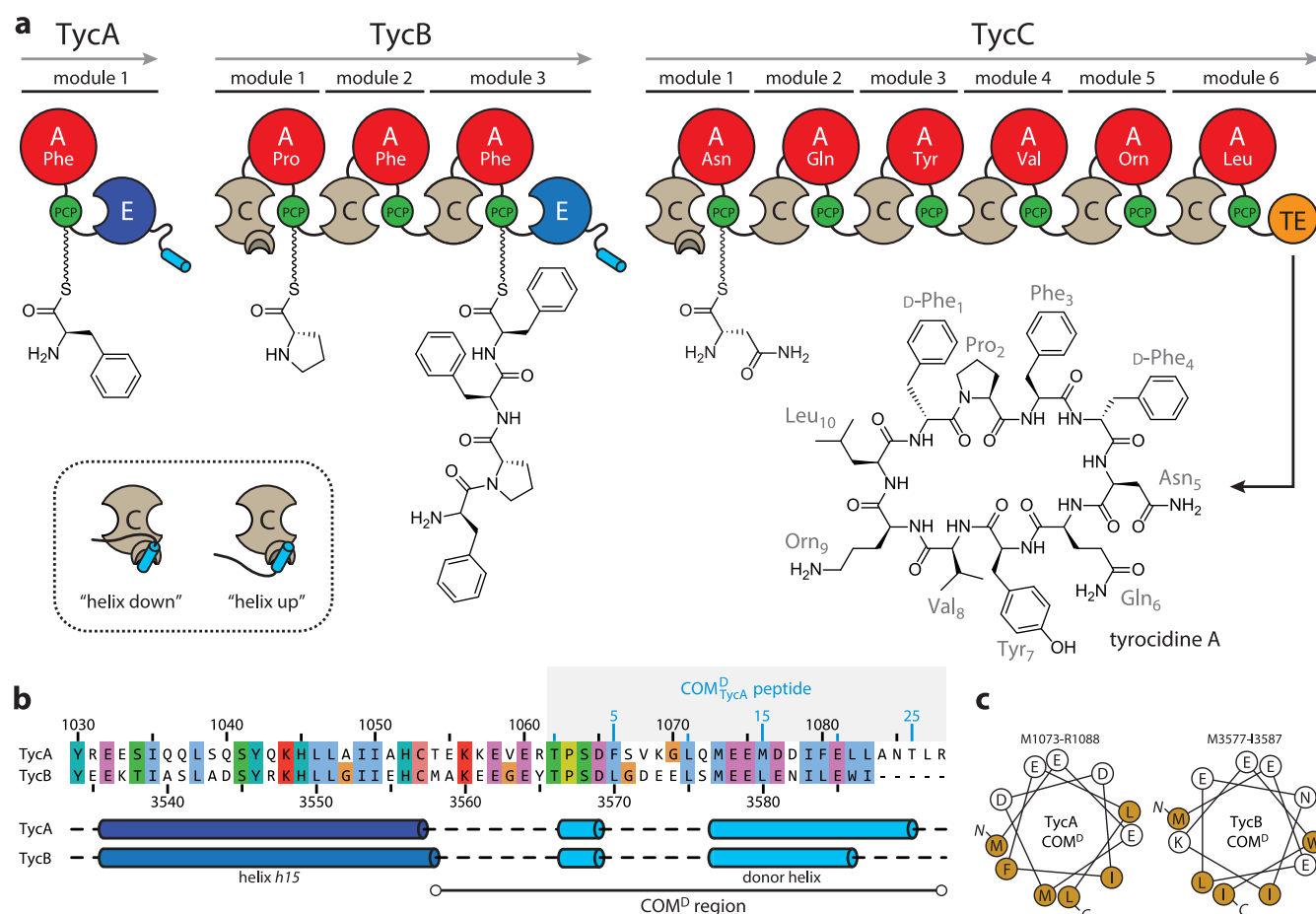
Nonribosomal peptide synthetases (NRPSs) are modular enzyme systems that generate a spectrum of natural products known as nonribosomal peptides (NRPs). The unique structures of NRPs endow them with an array of biological activities, such as antibacterial (e.g., tyrocidine), antifungal (e.g., enniatin), immunosuppressive (e.g., cyclosporine), and metal-chelating (e.g., bacillibactin).<sup>1</sup> In addition to the 20 canonical amino acids, more than 500 monomers can be incorporated into NRPs to enrich their chemical complexity.<sup>1–3</sup>

Stepwise assembly of NRPs is achieved by chain elongation modules consisting of a core set of NRPS domains: an adenylation (A) domain selects and activates a substrate monomer as an acyl adenylate and passes it to the 4'-

phosphopantetheinyl (Ppant) coenzyme of a peptidyl carrier protein (PCP) domain; the PCP domain then shuttles the thioester-linked substrate to the active site of a condensation (C) domain for elongation with the upstream PCP domain-bound intermediate (Figure 1).<sup>1</sup> Linear NRPSs like the tyrocidine synthetase adhere to the collinearity rule; that is, the order of catalytic modules, and thus the composition of the

Received: May 10, 2021

Revised: July 26, 2021



**Figure 1.** COM interactions are necessary for nonribosomal synthesis of tyrocidine A–D.<sup>9</sup> (a) The illustrated amino acyl and peptidyl substrates at the TycA/TycB and TycB/TycC subunit interfaces can be condensed by TycB1(C) and TycC1(C) owing to interactions between the COM<sup>D</sup> (cyan cylinder) and COM<sup>A</sup> (taupe crescent) regions. The inset illustrates two potential orientations for the COM<sup>D</sup> region when bound to the COM<sup>A</sup> region.<sup>10,11</sup> Domain abbreviations: adenylation (A), condensation (C), epimerization (E), peptidyl carrier protein (PCP), thioesterase (TE). Squiggly lines drawn from thioester substrates to PCP domains represent P<sub>ant</sub> groups; concave surfaces on E and C domains show donor and acceptor sites for PCP domains; substrate specificities (all L-configured) are included for A domains; gray arrows indicate subunit-encoding genes. (b) Alignment of C-terminal sequences from TycA and TycB. Secondary structural elements for both sequences, predicted by Phyre2, are shown below, with the putative COM<sup>D</sup> region highlighted.<sup>12</sup> See Figures S1–S2 for more extensive alignments. (c) Helical wheel representation of predicted amphipathic helices in the COM<sup>D</sup> regions of TycA and TycB. Hydrophobic and hydrophilic residues are shown as orange and white circles, respectively.

assembled product, can be deduced from the layout of the biosynthetic gene cluster. Each module may harbor additional domains, such as epimerization (E) or heterocyclization (Cy) domains that are homologous to C domains, to further diversify the chemical scaffold.<sup>1</sup> Finally, the C-terminal module typically ends in a thioesterase (TE) domain that catalyzes chain release via macrocyclization or hydrolysis.

In linear NRPSs, modules are organized on discrete polypeptide subunits (i.e., individual gene products) that must undergo ordered, noncovalent self-assembly to ensure the biosynthetic fidelity of the natural product.<sup>1</sup> Previous studies indicated that subunits containing terminal E and C domains, such as TycA, TycB, and TycC of the tyrocidine synthetase, interact end-to-end via conserved communication-mediating COM domains; that is, the C-terminal donor COM region of TycA (COM<sup>D</sup><sub>TycA</sub>) associates with the N-terminal acceptor COM region of TycB (COM<sup>A</sup><sub>TycB</sub>), and the C-terminal donor COM region of TycB (COM<sup>D</sup><sub>TycB</sub>) with the N-terminal acceptor COM region of TycC (COM<sup>A</sup><sub>TycC</sub>) (Figure 1).<sup>4,5</sup> One study showed that the relatively high degree of similarity

among certain COM sequences permits noncognate modules, such as TycA and SrfA-C (of the surfactin NRPS), to interact and thus yield dipeptide products in vitro.<sup>4</sup> To elicit productive communication between more divergent modules, such as TycB3 (the third module of TycB) and SrfA-C, substitution of COM<sup>D</sup><sub>TycB</sub> with COM<sup>D</sup><sub>TycA</sub> was found to be sufficient.<sup>4</sup> Moreover, a single K9D mutation in the COM<sup>A</sup><sub>TycC</sub> region permitted productive interaction with the otherwise incompatible TycA.<sup>5</sup> However, deletion of just six residues from the C-terminus of TycA disrupted productive binding with TycB1.<sup>4</sup> COM region substitutions have also led to product formation in vivo.<sup>6</sup> Nevertheless, not all COM interface engineering attempts (particularly those targeting COM<sup>A</sup> regions) have yielded the expected results.<sup>7,8</sup> It is worth noting that the engineering efforts outlined above were designed according to sequence homology and without prior knowledge of higher-order structure. Moreover, the currently available models of a COM interface fall short of explaining how mutations of interfacial residues have resulted in functional interactions between noncognate pairs.

While models for *intra*-subunit interactions in NRPSs have been proposed on the basis of known structures of modules and fragments thereof, such details are generally lacking for *inter*-subunit interactions (i.e., those employing docking domains).<sup>10,13,22,14–21</sup> To date, only two types of NRPS docking domains, each involving C-terminal PCP domains and N-terminal C domains, have been structurally elucidated: (1) in the tubulysin and RXP NRPSs, the  $\beta$ -hairpin motif on an N-terminal subdomain binds the charged and intrinsically disordered tail (or “short linear motif”) at the C-terminus of the upstream subunit;<sup>23,24</sup> (2) in the Pax NRPS, an N-terminal accessory helix binds the largely hydrophobic V-shaped groove between two helices at the C-terminus of the upstream subunit.<sup>25</sup> In contrast, *inter*-subunit interactions in the evolutionarily related polyketide synthases (PKSs) have been more thoroughly characterized.<sup>22,26–33</sup> Recently, hybrid PKS-NRPS interfaces between C-terminal carrier protein domains bearing short linear motifs and N-terminal C or Cy domains bearing  $\beta$ -hairpin docking domains were also examined.<sup>21,34</sup> Notably, structural studies on *inter*-subunit interactions in NRPSs, PKSs, and hybrids thereof have mostly been conducted on docking domains that were excised from their parent domains.<sup>22–27,31,32,34</sup>

The crystal structure of the SrfA-C termination module (C-A-PCP-TE domain topology) revealed an artificial COM-like interface between subunits: an  $\alpha$ -helical myc-His<sub>6</sub> tag at the C-terminus of the TE domain, sharing modest sequence homology with COM<sup>D</sup><sub>SrfA-B</sub>, is grasped by a hand-like  $\beta$ -sheet on the N-terminal C domain of a second copy of SrfA-C within the crystal.<sup>10</sup> Whether this represents a physiological or artifactual arrangement has remained inconclusive. Results of another study in which MS/MS fragments were mapped following photo-cross-linking between GrsA and TycB1 coincide with the helix-hand model, although the authors suggested an opposite (i.e., upward) orientation for the helix (see inset of Figure 1a).<sup>11</sup> However, this approach combined subunits from two different NRPS pathways and relied on the incorporation of a bulky unnatural amino acid into the COM<sup>D</sup> region, potentially perturbing the nature of the interaction.<sup>11</sup> To date, no biophysical methods have been employed to examine a native COM interface, and all reported structures of E domains, including those of TycA(E), *apo*-GrsA(PCP-E), and *holo*-GrsA(PCP-E), lack the flexible COM<sup>D</sup> region.<sup>35,36</sup>

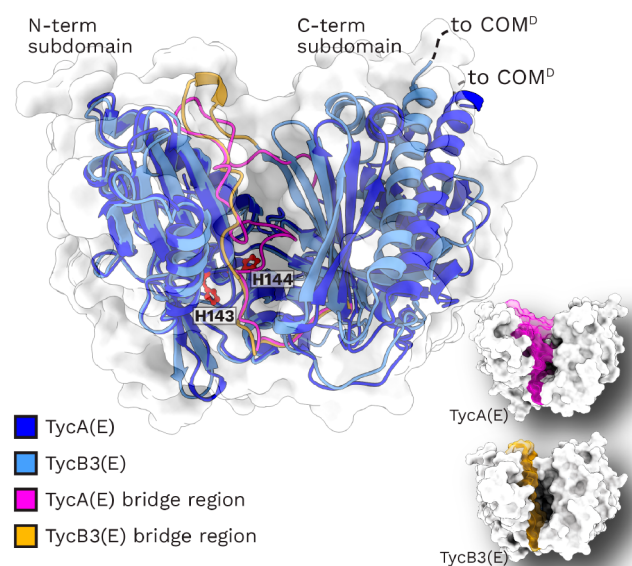
In the present study, we utilized a highly interdisciplinary biophysical and structural approach to capture in molecular detail a COM interface that is essential for the efficient biosynthesis of antibiotics tyrocidine A-D. Furthermore, a global assessment into the prevalence of COM domains in NRPS pathways afforded the discovery of uncharacterized domain architectures for COM interfaces.

We contend that our analyses, based on a set of complementary methods, may be used to characterize a variety of interfaces too flexible to readily crystallize or too finely detailed to capture with electron microscopy. Moreover, our findings add to the current understanding of NRPS subunit interactions and may, in harmony with existing methods like those involving exchange units (i.e., substitution of C domain subdomains), contribute alternative routes to NRPS assembly line diversity, thus facilitating the rational design of novel natural products.<sup>37,38</sup>

## RESULTS AND DISCUSSION

**Crystal Structure of a Chain Elongation Module E Domain.** To gain insight into the architecture of a COM interface, we sought to capture the TycA/TycB and TycB/TycC interfaces crystallographically. Equimolar solutions of TycA(E)/TycB1(C) and TycB3(E)/TycC1(C) were prepared and subjected to sparse matrix crystal screening, with the latter yielding crystals. Phases of a single crystal were solved by molecular replacement with a Phyre2 homology model of TycB3(E) based on the E domain from *holo*-GrsA(PCP-E) (PDB code 5ISX).<sup>12,36</sup> Although crystals grew in the presence of TycC1(C), only a pair of TycB3(E) monomers was found in the asymmetric unit. Notably, the C-terminal 27 and 26 residues of chains A and B (comprising the putative COM<sup>D</sup> region) could not be modeled because of the lack of electron density. Structural statistics are summarized in Table S10.

TycB3(E) is the first E domain from a chain elongation module to be structurally elucidated. As is typical of C-like domains, two chloramphenicol acetyltransferase-like subdomains form a V-shaped pseudodimer, between which the active site of TycB3(E) is located (Figures 2 and S3).<sup>35,36,39</sup> The



**Figure 2.** Superposition of TycB3(E) (PDB code 6TA8, chain B) with the very similar TycA(E) (PDB code 2XHJ; 50% sequence identity)<sup>43,44</sup> gives a relatively high r.m.s.d of 1.71 Å over 372 C $\alpha$  atoms, with inconsistencies in the bridge regions and angles between the N- and C-terminal subdomains. Both E domains are displayed as cartoons, and TycB3(E) is further rendered as a semitransparent white surface. The COM<sup>D</sup> region, which is not resolvable in either structure, is indicated at its point of attachment with a dashed line. Side chains of the conserved H143 and H144 in the HHxxxD motif of TycB3(E) are shown in red ball-and-stick representation. Surface renderings at bottom right highlight the bridge region and intersubdomain cavity of each E domain.

enzyme assumes a conformation nearly identical to that of TycA(E) and GrsA(E) from initiation modules (PDB codes 2XHJ and 5ISX, respectively) (Figures 2 and S4). However, one salient difference can be seen in the bridge region of TycB3(E) (residues  $\sim$  E372–L400), which rests above the Ppant thiol in the structure of *holo*-GrsA(PCP-E): the loop is 8 residues shorter than in TycA(E) and GrsA(E), and does not snake into the active site as in the latter two (Figures 2 and



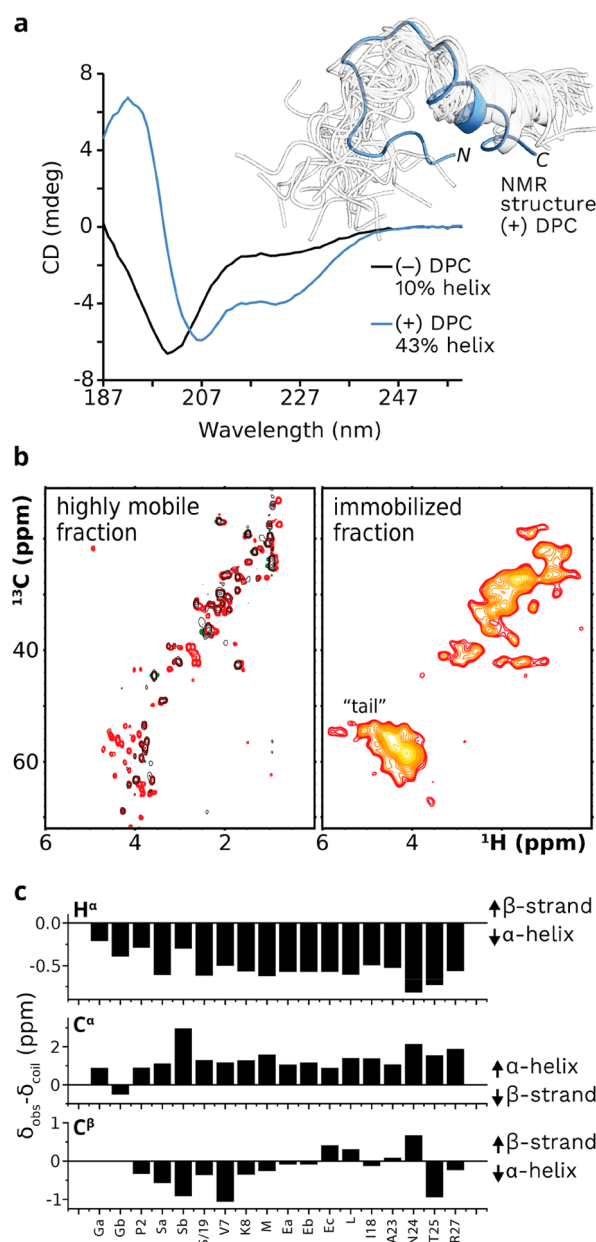
S4).<sup>35,36</sup> This feature is presumably optimized for binding of peptidyl substrates to TycB3(E) and amino acyl substrates to TycA(E) and GrsA(E), in line with published preferences.<sup>40–42</sup> Indeed, except for those lining the bridge region, all residues directly adjacent to the Ppant thiol of *holo*-GrsA(PCP-E) are identical to those of TycB3(E) (Figure S5). Finally, although the donor channel is obstructed in the crystal structure of TycB3(E), the channel opens during MD simulations as a result of a scissoring motion between the N- and C-terminal subdomains (Figure S6 and Movie S1) as proposed previously.<sup>34</sup>

**Solution Structure of COM<sup>D</sup> Peptide.** As with GrsA(E), no electron density is present for the COM<sup>D</sup> region of TycB3(E), despite its inclusion in the expression construct. Moreover, the region was necessarily excluded from TycA(E) to promote crystallization.<sup>45</sup> COM<sup>D</sup> regions are predicted by the Phyre2 server<sup>12</sup> to form an amphipathic  $\alpha$ -helix preceded by disordered loops and a smaller  $\alpha$ -helix (Figures 1b,c and S1). Given these findings, we surmised that in the absence of a downstream binding partner, the COM<sup>D</sup> region assumes a range of flexible conformations relative to its N-terminal E domain. Thus, we turned to circular dichroism and NMR spectroscopy to examine the precise nature of the COM<sup>D</sup> fold.

Synchrotron radiation circular dichroism (CD) spectroscopy experiments were conducted on a solution of a synthetic 27-residue COM<sup>D</sup><sub>TycA</sub> peptide (T1062-R1088 of TycA; Figure 1b) for secondary structure estimation. Our analysis suggested only minor helical content ( $\sim 10\%$ ) for the peptide (Figure 3a). However, we considered whether the predicted amphipathic helix would be unstable without the complementary hydrophobic surface of an acceptor. Measurements were therefore repeated after addition of micelle-forming dodecylphosphocholine (DPC) to simulate the hydrophobic environment provided by COM<sup>A</sup><sub>TycB</sub>. Indeed, addition of DPC above the critical micelle concentration induced a significant shift in the helical content of the peptide ( $\sim 43\%$ ) (Figure 3a).

To visualize which residues of the COM<sup>D</sup><sub>TycA</sub> region adopt a helical conformation, we employed NMR spectroscopy. COM<sup>D</sup><sub>TycA</sub> peptide exhibited typical random coil behavior in aqueous solution, as evidenced by low chemical shift dispersion and a lack of medium- and long-range nuclear Overhauser effects (NOEs) (Figures S7a,b and S8a; Table S13). Measurements performed in the presence of DPC, however, led to the appearance of new peaks with chemical shifts characteristic of  $\alpha$ -helices (Figures S7c,d and S8b; Table S14), consistent with our CD spectroscopy findings. Structure calculations using NOEs and dihedral angles show a well-defined  $\alpha$ -helix composed of residues E14-R27 on the C-terminal half of the peptide (E1075-R1088 of TycA) (inset of Figure 3a; Figures S9–S10; Table S15).

**Solid-State NMR Analyses of COM<sup>D</sup> Peptide with an Acceptor Domain.** To confirm whether the COM<sup>D</sup><sub>TycA</sub> region folds in the presence of TycB1(C), we prepared a sedimented complex of <sup>13</sup>C,<sup>15</sup>N-labeled donor peptide (N-terminal Gly followed by P1063-R1088 of TycA) and unlabeled acceptor domain for analysis by solid-state NMR spectroscopy, which does not suffer from the system size limitations present in solution.<sup>34,46–49</sup> We observed a dominant, highly mobile population of the peptide with fast (ps–ns), large-amplitude motions in scalar coupling-based experiments and a smaller, heterogeneous population of immobilized peptide in dipolar coupling-based experiments (Figure 3b). While the highly mobile (Figures S11–S15)

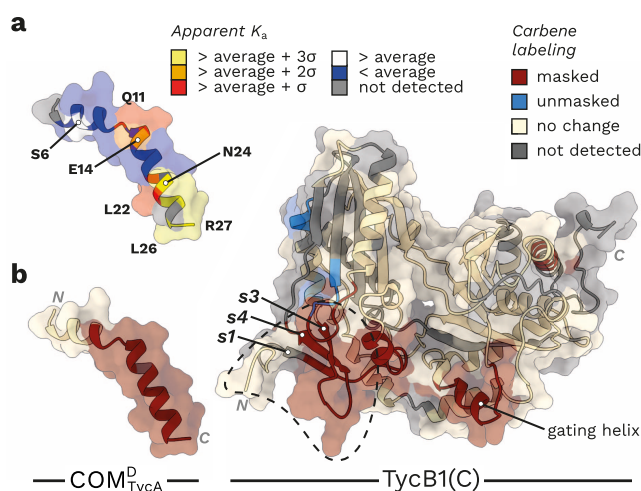


**Figure 3.** Structural analyses of COM<sup>D</sup><sub>TycA</sub> peptide. (a) CD spectra of the peptide, in the presence and absence of DPC. The inset depicts the solution-state NMR structure of the peptide in DPC-containing buffer. (b) Solid-state NMR spectra for the dominant, highly mobile fraction (left) and minor, immobilized fraction (right) of <sup>13</sup>C,<sup>15</sup>N-COM<sup>D</sup><sub>TycA</sub> peptide in a sedimented complex with TycB1(C). Red peaks indicate resonances observed for freshly prepared complex, where both random coil (free peptide) and mostly  $\alpha$ -helical (bound peptide) conformations are registered; black peaks indicate resonances observed after extended sedimentation, where only bound peptide is registered. (c) Secondary chemical shifts for assigned residues in the dominant, dynamic population of the peptide in the sedimented complex.

bound fraction was initially in equilibrium with the intrinsically disordered unbound fraction, prolonged sedimentation under magic angle spinning conditions shifted the equilibrium substantially toward the bound form (compare red and black spectra in Figure 3b).

In the highly mobile population of the peptide, we were able to unambiguously identify resonances corresponding to residues P2 (P1063 of TycA), V7, K8, I18, A23, N24, T25, and R27. Moreover, we identified two distinct sets of peaks for Gly (G1/9) and Ser (S3/6) residues; three sets of  $C^\beta$  and  $C^\gamma$  peaks for Glu (E13/14/20) residues; and single (e.g., overlapping) sets of peaks for Leu (L10/21/22/26), Phe (F5/19), and Met (M12/15) residues. No resonances consistent with residues Q11 and D1/4/16 could be resolved in the highly mobile fraction. Most of the identified residues appear to adopt a helical conformation, according to their secondary chemical shifts (Figure 3c). This suggests that a large fraction of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide folds upon interaction with TycB1(C)—not only residues corresponding to the DPC micelle-induced C-terminal helix but also several residues at the N-terminus (Tables S16–S17).

Because of the low efficiency of the dipolar coupling-based magnetization transfer and limited resolution, 3D experiments with sufficient quality for de novo assignment were not obtainable. However, comparison of the experimental spectra to simulated spectra indicates that most of the observed peaks are consistent with helical and random coil conformations (Figure S16), except for a Gln that assumes an extended conformation (see “tail” in immobilized fraction in Figure 3b). Thus, the overall conformation of bound  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide likely involves two highly flexible helices linked by a loop encompassing Q11 (Q1072 of TycA), as depicted in Figure 4.



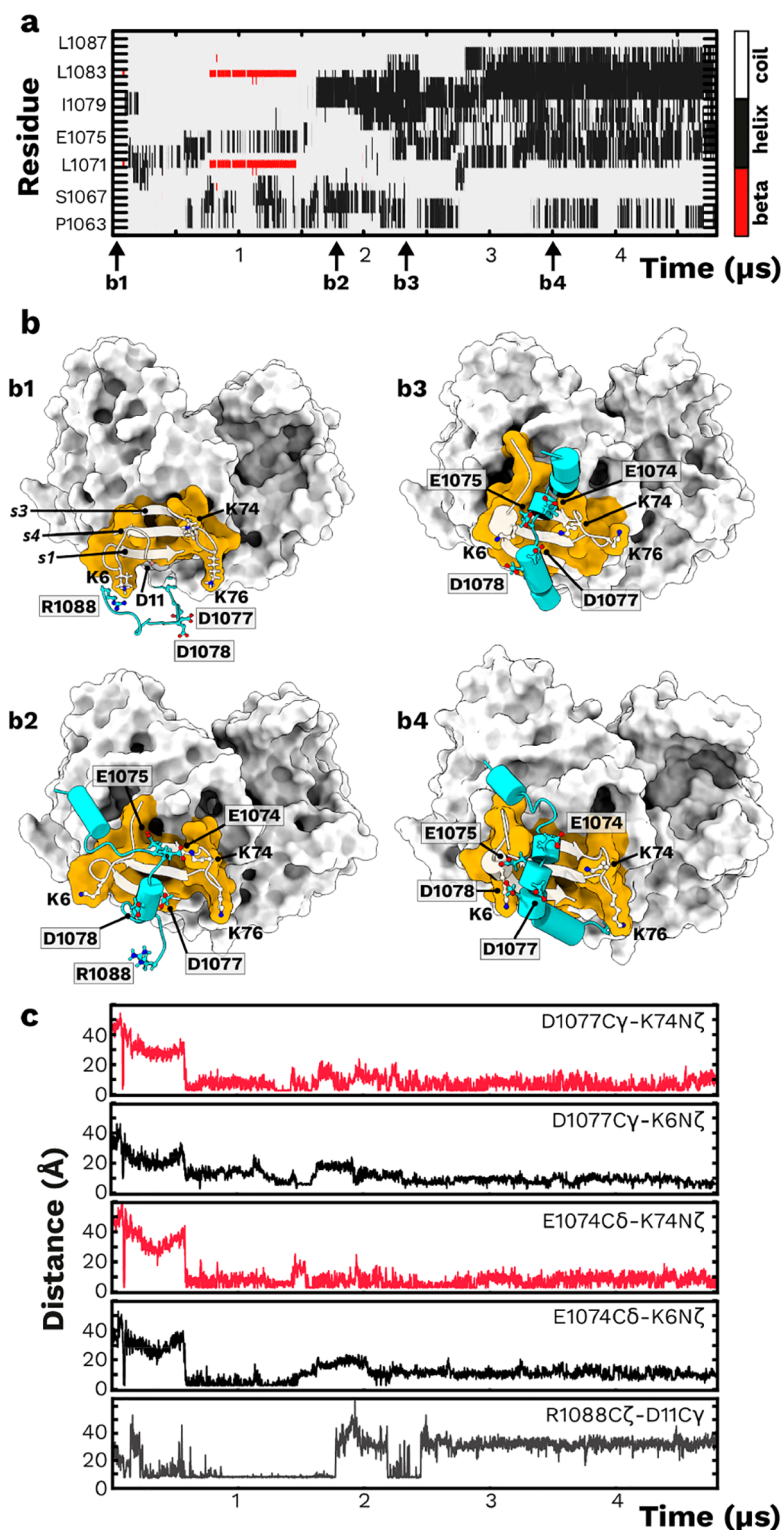
**Figure 4.**  $\text{COM}^{\text{D}}_{\text{TycA}}$ /TycB1(C) interactions. (a) Results of NMR titration of  $^{15}\text{N}$ - $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide with TycB1(C) in solution. Residues are colored according to the magnitude of the apparent association constant,  $K_a^*$  (yellow = largest). (b) Carbene footprinting analysis of the COM interface. Results are mapped onto (left) one conformer of a  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide ensemble consistent with chemical shifts observed in a sedimented complex with TycB1(C) and (right) a homology model of TycB1(C) based on the C domain from SrfA-C.<sup>10,12</sup> Strands s1, s3, and s4 of the  $\beta$ -hand motif (dashed crescent), the acceptor channel-gating helix h14, and N- and C-termini are labeled.

**Solution-State NMR Titrations of  $\text{COM}^{\text{D}}$  Peptide with an Acceptor Domain.** To follow changes in  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide upon addition of its acceptor domain TycB1(C), we conducted NMR titrations in solution. Increasing concentrations of TycB1(C) resulted in broadening and disappearing amide signals of  $^{15}\text{N}$ - $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide, as is characteristic of

slow exchange and formation of a complex too large to be observed by solution-state NMR spectroscopy. Nevertheless, by following the rate of disappearance of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide peaks, we could identify interaction sites with residue-level resolution. After fitting the signal decay as a function of TycB1(C) concentration, it became clear that the C-terminal part of the peptide, particularly L21, A23, N24, L26, and R27 (L1082, A1084, N1085, L1087, and R1088 of TycA), is most affected by the interaction (Figures 4a and S17–18). Overall, the apparent binding strength decreases gradually from the C- to N-terminus, indicating a more transient nature for interactions at the N-terminus. For residues with  $K_a^*$  (i.e., apparent  $K_a$ ) values above the threshold, the  $K_d^*$  (or  $1/K_a^*$ ) values range over 0.1–5  $\mu\text{M}$ , in agreement with isothermal titration calorimetry (ITC) measurements (vide infra). When titrations of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide with TycB1(C) were carried out in the presence of DPC, neither line broadening nor changes in peak intensity were detected (Figure S8c–d). This can be attributed to occupation of the binding site for TycB1(C) by DPC (available in large excess over the acceptor domain) or the inability of TycB1(C) to enter a DPC micelle. This also suggests that the C-terminus of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide drives association with TycB1(C).

**Mapping the COM Interface with Carbene Footprinting.** To further pinpoint which residues contribute to the COM interface, we employed the recently described carbene footprinting methodology.<sup>31,34,50</sup> This structural mass spectrometry technique exploits covalent labeling of solvent-exposed protein residues by a reactive carbene species, formed by in situ photolysis of the corresponding diazirine. Proteolytic cleavage and LC-MS analysis affords peptide-level labeling information. By labeling the protein in the presence and absence of its binding partner, peptides at the solvent-accessible surface (unmasked due to conformational changes upon binding) or solvent-excluded surface (masked due to an interaction interface or conformational changes upon binding) can be identified.

Footprinting experiments were carried out on solutions containing  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide, TycB1(C), or a mixture of the two. Differential labeling showed statistically significant masking of residues G9–M15 and D17–R27 (G1065–M1076 and D1078–R1088 of TycA) at the C-terminus of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide (Figures 4b and S19; Tables S18–S22). Together with the results of the NMR titrations, this indicates that Q11–R27 (Q1072–R1088 of TycA) is the primary fragment that binds TycB1(C), with more transient interactions, if any, for the N-terminal part of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide. For TycB1(C), masking was observed for peptides D11–L28, T68–K74, and Q77–K84 at the N-terminus, corresponding to the  $\beta$ -hand motif and neighboring helix in a homology model (Figures 4b and S20; Tables S23–S27).<sup>10,12</sup> (This homology model is based on the 40% identical C domain from SrfA-C, with which TycA is known to interact. Residues on strands s1, s3, and s4 of the C domains from TycB1 and SrfA-C are identical or conserved in properties; see Figure S2.)<sup>4</sup> Interestingly, residues D124–R128 of TycB1(C), modeled as a loop above strand s3 of the  $\beta$ -hand motif, showed significant unmasking, suggesting a conformational change that increases solvent exposure in this region, perhaps to accommodate binding of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide (further suggested by masking of T118–K123 on the same loop). In addition, we observed masking of D338–I350, which belong to the helix involved in gating the acceptor channel,<sup>34</sup>



**Figure 5.** Folding of  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide in the presence of TycB1(C) during molecular dynamics simulations. Classical MD simulations ( $4.8 \mu\text{s}$ ) were carried out with an ff14SB force field and TIP3P water model. (a) Evolution of secondary structure of the peptide. (b) Illustrated frames b1–b4 from panel a. The catalytic body of a TycB1(C) homology model (based on the C domain from SrfA-C)<sup>10,12</sup> is depicted as a white surface, the  $\beta$ -hand motif as a white cartoon and yellow surface, and  $\text{COM}^{\text{D}}_{\text{TycA}}$  peptide as a cyan cartoon with cylindrical helices and key charged side chains in ball-and-stick representation. (c) Evolution of select distances between charged side chains that play an important role in the folding mechanism (see also Figure S21).

and L420–A428, which are far from the expected binding interface. Finally, peptide D48–T53 showed unmasking.

**Molecular Dynamics Simulations and In Silico Folding of  $\text{COM}^{\text{D}}$  Peptides.** The experimental results discussed above suggest that the  $\text{COM}^{\text{D}}_{\text{TycA}}$  region folds



upon binding TycB1(C). To investigate whether consistent folding could be observed in MD simulations, we performed a series of cMD simulations (up to 5  $\mu$ s) and accelerated molecular dynamics (aMD) simulations (up to 2  $\mu$ s) starting from free COM<sup>D</sup><sub>TycA</sub> peptide (corresponding to the NMR construct, with an N-terminal Gly followed by P1063-R1088 of TycA) or linker+COM<sup>D</sup><sub>TycA</sub> peptide (T1054-R1088 of TycA; i.e., all residues C-terminal to the TycA(E) core), respectively, in random, extended conformations and a homology model of TycB1(C) based on the C domain from SrfA-C (Tables S28–S29).<sup>10,12</sup> While the points of contact gleaned from our carbene footprinting experiments guided these simulations, various initial orientations of the peptides relative to the  $\beta$ -hand motif were explored to minimize bias. Simulations with linker+COM<sup>D</sup><sub>TycA</sub> peptide were performed to confirm whether findings from experiments and simulations involving the shorter COM<sup>D</sup><sub>TycA</sub> peptide apply when the entire linker is considered.

During the cMD simulation, COM<sup>D</sup><sub>TycA</sub> peptide folded upon interaction with TycB1(C) (Figure 5a,b). Within  $\sim 2$   $\mu$ s, residues at the C-terminus assumed a dynamic helical conformation in a “helix down” orientation consistent with that proposed by Tanovic et al. (inset of Figure 1a).<sup>10</sup> Residues at the N-terminus sampled helical conformations more rapidly but fluctuated between helical and random coil conformations throughout the simulation. Folding was initiated via threading of the C-terminal end of the linear peptide through the  $\beta$ -hand motif, with interaction between R27 of COM<sup>D</sup><sub>TycA</sub> peptide (R1088 of TycA) and D11 of TycB1(C) controlling the depth of threading (Figure 5c). Folding of the C-terminal helix involved an interplay between charged residues, particularly the negatively charged E13, E14, D16, and D17 of the COM<sup>D</sup> peptide (E1074, E1075, D1077, and D1078 of TycA) and positively charged K6, K74, and K76 of the  $\beta$ -hand motif (see Figure 5b,c), as well as hydrophobic residues, particularly F19 and L22 on the COM<sup>D</sup> peptide (F1080 and L1083 of TycA) and M12, L79, and V81 on TycB1(C).

To sample greater conformational space at lower computational cost, we turned to aMD simulations.<sup>51</sup> In the presence of TycB1(C), the N- and C-terminal parts of linker+COM<sup>D</sup><sub>TycA</sub> peptide folded into  $\alpha$ -helices connected by a flexible loop (Figure S22). In the majority of the simulations, the C-terminal helix folded and docked in the “helix down” orientation, with only one simulation yielding the “helix up” orientation (Figure S23). The relatively small difference in free energy of binding for these orientations ( $\sim -57$  and  $\sim -38$  kJ/mol, respectively), calculated via the MM/PBSA approach (Tables S30–S31),<sup>52</sup> may explain why a “helix up” orientation was previously observed for the GrsA/TycB1 interface after introduction of a bulky photo-cross-linker into the COM<sup>D</sup> region.<sup>11</sup> Nevertheless, MD simulations (particularly those involving homology models) should be interpreted with caution, and further experiments will be necessary to validate the binding orientation of the donor COM region relative to its acceptor domain.

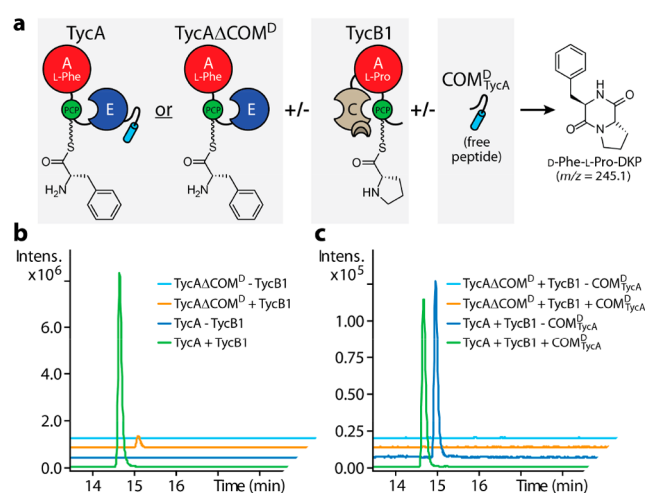
Overall, the folding and interaction patterns we observed during our MD simulations are consistent with our NMR spectroscopy and carbene footprinting results. In addition to accounting for the direct interactions between COM<sup>D</sup><sub>TycA</sub> peptide and TycB1(C), our simulations suggest that masking of the D338-I350 peptide (part of helix *h14* and the preceding loop) of TycB1(C) occurs because of a scissoring motion between the N- and C-terminal subdomains, bringing helices

*h1* and *h14* closer together and restricting the acceptor channel (Figure S24). Thus, a conformational change may be facilitated by COM<sup>D</sup><sub>TycA</sub> peptide binding. Since the  $\beta$ -hand motif binding the COM<sup>D</sup><sub>TycA</sub> region is located near the acceptor site of the C domain but TycA(PCP) acts at the donor site, the flexible linkers connecting TycA(PCP) to TycA(E) and TycA(E) to the COM<sup>D</sup><sub>TycA</sub> region must coordinate to allow for C domain activity on the donor and acceptor PCP domain-bound substrates (Figure S25).

Interestingly, the folding and interaction pattern observed for the COM<sup>D</sup> region resembles that of the well-studied but biologically unrelated pKID/KIX system involved in eukaryotic gene transcription.<sup>53</sup> Upon binding to KIX, the intrinsically disordered peptide pKID folds into two helices: an amphipathic C-terminal helix that assembles when interacting with a hydrophobic patch on KIX (structurally distinct from the  $\beta$ -hand motif) and a mostly polar/charged N-terminal helix that makes minimal contact with KIX.

**Affinity of the COM Interface.** To estimate which region of the donor species contributes most to the *inter*-subunit interaction interface, we prepared individual solutions of the donor module *holo*-TycA, domain TycA(E), and peptide COM<sup>D</sup><sub>TycA</sub> or COM<sup>D</sup><sub>TycB</sub> as well as the acceptor domain TycB1(C), and then carried out ITC measurements. Titration of *holo*-TycA, TycA(E), and COM<sup>D</sup><sub>TycA</sub> peptide with TycB1(C) yielded  $K_d$  values of  $12.8 \pm 4.8$ ,  $2.3 \pm 0.7$ , and  $5.9 \pm 0.6$   $\mu$ M, respectively (Figure S26). These findings are comparable to the micromolar affinities previously obtained for docking domains at the termini of PKS and NRPS subunits,<sup>23,27,28,30</sup> including the COM-facilitated interaction of *apo*-GrsA (63% identical to TycA)<sup>43,44</sup> and TycB1,<sup>54</sup> as well as those estimated herein from NMR titrations (*vide supra*). As anticipated, COM<sup>D</sup><sub>TycB</sub> peptide failed to bind the noncognate acceptor TycB1(C), precluding the association of multiple TycB copies over the correct TycB/TycC pair. The fact that *holo*-TycA, TycA(E), and COM<sup>D</sup><sub>TycA</sub> peptide bind TycB1(C) with similar affinities suggests that the COM<sup>D</sup> region is the most crucial fragment for interaction with the downstream subunit.

**The COM Interface Facilitates Peptide Bond Formation across Subunits.** In order to validate the results of our biophysical experiments, we assayed diketopiperazine (DKP) production by modules TycA (A-PCP-E-COM<sup>D</sup> domain topology) and TycB1 (C-A-PCP domain topology).<sup>11,55</sup> In this assay, L-Phe and L-Pro are selected and activated by the A domains of TycA and TycB1, respectively, and then *trans*-acylated onto adjacent *holo*-PCP domains for condensation by TycB1(C) (after E domain-catalyzed epimerization of L-Phe to D-Phe). The dipeptide intermediate D-Phe-L-Pro then undergoes autocyclization on TycB1 to yield free D-Phe-L-Pro-DKP, which is readily detectable by LC-MS ( $m/z = 245.1$ ) (Figure 6a). A solution of *holo*-TycB1 and *holo*-TycA or *holo*-TycA $\Delta$ COM<sup>D</sup> (27 C-terminal residues deleted) was incubated with substrates L-Phe, L-Pro, ATP, and MgCl<sub>2</sub> in the presence and absence of COM<sup>D</sup><sub>TycA</sub> peptide for 1 h at 30 °C prior to extracting and analyzing by LC-MS. In contrast to a previous report,<sup>4</sup> we found TycA $\Delta$ COM<sup>D</sup> was still able to generate a small but detectable amount of DKP when incubated with TycB1, revealing a nonessential but facilitative role for the COM<sup>D</sup> region in the condensation of amino acyl substrates across subunits (Figure 6b). To investigate whether the donor COM region primes its acceptor C domain for the condensation reaction by inducing a conformational rearrangement upon binding, we included COM<sup>D</sup><sub>TycA</sub> peptide in the



**Figure 6.** D-Phe-L-Pro-diketopiperazine production assays to monitor peptide bond formation across an E-COM<sup>D</sup>/C domain interface in the tyrocidine NRPS. (a) Assay schematic, with domains and Ppat illustrated as in Figure 1. The DKP product was extracted and analyzed by LC-MS. (b,c) Representative extracted ion chromatograms (stacked and staggered). (b) L-Phe-loaded TycA or TycAΔCOM<sup>D</sup> was incubated with L-Pro-loaded TycB1 or buffer. (c) L-Phe-loaded TycA or TycAΔCOM<sup>D</sup> was incubated with L-Pro-loaded TycB1 and a 100-fold molar excess of COM<sup>D</sup><sub>TycA</sub> peptide or buffer. Intensities in panels b,c are not directly comparable because of differences in protein concentrations.

reaction containing TycAΔCOM<sup>D</sup> and TycB1. As addition of free peptide failed to restore DKP production, a sole conformational priming role for the donor COM region can be discounted (Figure 6c). Lastly, we assayed DKP formation after substituting TycA with a charge-swapped R1088E variant. Relative to the wild-type reaction, we observed a decrease in DKP biosynthesis (~35%; Figure S27), supporting a role for R1088 at the interface. This finding is consistent with our MD simulations, wherein the C-terminal Arg aids in threading the COM<sup>D</sup> peptide through the β-hand motif, ultimately leading to a “helix down” orientation. Taken together, these findings reiterate the necessity of COM interfaces for efficient conveyance of intermediates along NRPS assembly lines.

**Modeling COM Interfaces.** We explored whether our findings could explain the COM region compatibility observed in several experimentally tested systems. Previously, a single K9D mutation in the COM<sup>A</sup><sub>TycC</sub> region enabled productive interaction with the noncognate donor TycA.<sup>5</sup> The COM<sup>A</sup> regions of SrfA-C and GrsB, which are highly similar to COM<sup>A</sup><sub>TycB</sub> and contain an Asp at the corresponding position (S2), also productively interacted with their noncognate donor TycA.<sup>5</sup> Indeed, our cMD and aMD simulations indicate that transitory interactions between R1088 and D11 (TycA/TycB) and between E3585 and K9 of (TycB/TycC) govern the depth of C-terminal peptide threading through the β-hand motif, thereby controlling the extent of COM<sup>D</sup> region folding and potentially gatekeeping the tyrocidine NRPS subunit interactions. In each case, these key side chains remained in close proximity until the point of C-terminal donor helix assembly (Figures 5c and S28). Notably, while the model proposed by Tanovic et al. apparently predicts the correct orientation for the C-terminal donor helix in the β-hand motif,<sup>10</sup> it offers no insight into the transient interactions that promote or demote folding.

When we simulated the folding of linker+COM<sup>D</sup><sub>TycA</sub> peptide in the presence of a homology model of the noncognate TycC1(C) (based on the C domain from SrfA-C),<sup>10,12</sup> we found that even though the acceptor’s hydrophobic patch promoted initial folding of the donor’s C-terminal helix, the transitory repulsion between R1088 and K9 of (TycA/TycC) destabilized that helix, prevented its proper insertion into the β-hand motif, and led to unfolding (Figure S29; in another simulation, we observed that folding of the C-terminal helix was stabilized by a hydrophobic patch outside of the β-hand motif). Conversely, in an equivalent simulation with a K9D variant of TycC1(C), a helix with approximately three turns formed and buried itself in the β-hand motif (Figure S30).

Because of sequence divergence (Figures S1–S2), other NRPSs may rely on different residues or alternative forms of gatekeeping. Nevertheless, our model, for example, suggests how a T7D mutation in COM<sup>D</sup><sub>PpsB</sub> of the plipastatin biosynthetic pathway permitted functional interaction with the noncognate PpsE while reducing affinity for its cognate partner PpsC (Figure S31).<sup>7</sup>

Taken together, our model accurately accounts for the results of key engineering efforts on COM interfaces and will thus be of value in the design of future engineering experiments. In particular, interactions which transiently guide folding should be considered with equal import to those which stabilize the final complex.

**Bioinformatics Analyses Reveal New Domain Architectures at COM Interfaces.** For the purpose of estimating COM domain frequency in natural product biosynthetic pathways, we undertook a global bioinformatics assessment. We constructed a profile hidden Markov model (pHMM)<sup>56</sup> from sequences presented by Hahn et al.<sup>5</sup> and searched the MIBiG repository;<sup>57</sup> we then supplemented the pHMM with repository hits and repeated the search. This approach revealed a multitude of canonical *inter*-subunit COM interfaces (i.e., split E-COM<sup>D</sup>/C domains) in addition to *intra*-subunit COM interfaces (i.e., fused E-COM<sup>D</sup>-C domains).<sup>11</sup> We also manually examined and predicted secondary structures for hits below the default cutoff and for interfaces not found by the pHMM. Overall, from 605 NRP entries in the MIBiG repository, we found that 16.9% of assembly lines contain one or more (≤4) split interfaces, 23.0% contain one or more (≤7) fused interfaces, and 6.8% contain both split and fused interfaces (Table S32 and Figures S32–S35). A recent study on reengineering of the teicoplanin NRPS concluded that the existing *inter*-domain linker between E and C domains (corresponding to the fused COM<sup>D</sup> region in our searches) should be retained when transplanting E domains from other modules.<sup>58</sup> This finding suggests an important role for fused COM<sup>D</sup> regions, likely in constraining NRPS subunit architecture. We hypothesize that, like split COM<sup>D</sup> regions, fused COM<sup>D</sup> regions are nonstatic moieties that undergo binding/folding or unbinding/unfolding as needed to support NRPS transitions; however, the covalent constraint between the donor and acceptor entities may limit such dynamics. Future experiments will clarify this point.

Our pHMM-based analyses also identified an apparently undiscovered domain topology for *inter*-subunit COM interfaces in 3.6% of assembly lines: COM<sup>D</sup> regions at the C-termini of PCP domains interact with C domains (i.e., split PCP-COM<sup>D</sup>/C domains) in the machineries responsible for producing the last-resort antibiotics daptomycin, teicoplanin, and vancomycin, for example (Table S32 and Figures S36–



S37). While similar interfaces have been engineered through E domain deletions (e.g., A-PCP-COM<sup>D</sup> domain topology),<sup>5,11</sup> none occurring in nature have been reported until now. Furthermore, we identified previously undescribed COM<sup>D</sup>-like sequences embedded near the N-termini of dual C/E domains<sup>59</sup> that are preceded by split or fused PCP domains in 7.1% of assembly lines (Table S32 and Figures S38–S39). No similarly embedded sequences have been reported in the literature, and whether they function like characterized COM<sup>D</sup> regions remains to be determined. Encouragingly, the Jpred server predicted a helix on these noncanonical COM<sup>D</sup> and COM<sup>D</sup>-like regions, and the SWISS-MODEL server predicted an N-terminal  $\beta$ -hand motif on their associated C or dual C/E domains.<sup>60,61</sup> Curiously, the antiSMASH library associated with the MiBIG repository does not appear to detect these variations on the COM<sup>D</sup> theme.<sup>62</sup>

**Validating a New Domain Architecture at the COM Interface.** To verify a role for the PCP-COM<sup>D</sup> domain in *inter*-subunit communication, we designed a functional assay for peptidyl chain condensation across the subunit interface (Figure 7a). Here, we examined the standalone donor module

subsequently loaded onto the PCP domain of Tcp10. Addition of L-Hpg, ATP, and MgCl<sub>2</sub> permitted A domain-catalyzed loading of L-Hpg onto the *holo*-PCP domain of Tcp11 module 4, priming the system for condensation. After 2 min at 30 °C, intermediates were captured from the NRPS using a methylamine off-loading methodology,<sup>63</sup> and relative quantities were established by LC-MS. Indeed, after short incubation with its acceptor module, the wild-type donor resulted in 5–6 times higher yields of tetrapeptide product than a  $\Delta$ COM<sup>D</sup> donor (34 C-terminal residues deleted) (Figures 7b,c and S41). Overall, these findings showcase the versatility of COM regions in promoting functional domain–domain interactions beyond those of the canonical E and C domain pair.

## CONCLUSIONS

In summary, we elucidated the helix-hand arrangement of an NRPS communication-mediating (COM) interface wherein the  $\beta$ -hand motif of the acceptor subunit grasps the C-terminal  $\alpha$ -helix of the donor subunit. Though intrinsically disordered in isolation, COM<sup>D</sup> regions fold upon binding acceptor C domains in a process conceptually similar to that of the well-characterized but unrelated pKID/KIX system in eukaryotes,<sup>53</sup> as demonstrated through our suite of complementary biophysical, structural biological, and computational experiments. Importantly, effective engineering of COM interfaces may require consideration of interactions which not only stabilize the final complex but also guide folding.

The relevance of the donor COM region in mediating noncovalent interactions with the downstream subunit, and thus efficient peptide bond formation, was further confirmed through activity assays and affinity measurements. Our affinity, NMR spectroscopy, and simulation results suggest that NRPS interactions are transitory at the supermodular level. Our study also adds to the growing body of work highlighting the importance of intrinsically disordered regions in mediating intermolecular interactions in natural product assembly lines.<sup>31,33,34</sup>

Our bioinformatics analyses unveiled the prevalence of both *inter*- and *intra*-subunit COM interfaces throughout biosynthetic assembly lines, including two types of PCP domain-bridging interfaces not previously characterized. It remains to be seen whether NRPS-utilizing organisms have already evolved other iterations of COM or COM-like interfaces to promote natural product diversity.

On the basis of the widespread nature of COM interfaces across natural product pathways, we suggest that the findings reported herein will be of use in future experiments focused on designing noncognate NRPS interactions, in harmony with *intra*-subunit engineering methods such as the recently described exchange unit concept.<sup>37,38</sup> Such interfaces could be exploited to develop novel bioactive molecules with enhanced potency or attenuated side effects.

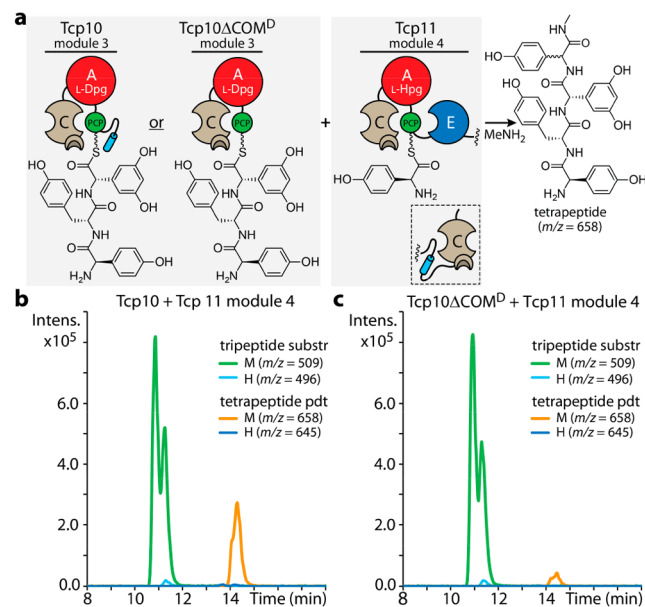
## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acscatal.1c02113>.

#### Movie S1 (MP4)

Experimental details; alignment of sequences containing COM regions; evaluation of known E domains; data, statistics, and analyses from carbene footprinting, ITC, MD simulation, NMR spectroscopy, and X-ray crystal-



**Figure 7.** Tetrapeptide production assays to monitor peptide bond formation across a PCP-COM<sup>D</sup>/C domain interface in the teicoplanin NRPS. (a) Assay schematic, with domains and Ppant illustrated as in Figure 1. Methylamine-offloaded tetrapeptide product was analyzed by LC-MS (C $\alpha$  epimerization was not evaluated). The acceptor's C-terminally fused COM<sup>D</sup>-C domain (inset) does not participate in the reaction. (b, c) Representative extracted ion chromatograms (overlaid). Tripeptide-loaded (b) Tcp10 or (c) Tcp10 $\Delta$ COM<sup>D</sup> was incubated with L-Hpg-loaded Tcp11(C-A-PCP-E-C). A minor fraction of peptide was found to be hydrolyzed (H) rather than methylamidated (M). Abbreviations: 3,5-dihydroxyphenylglycine (Dpg), 4-hydroxyphenylglycine (Hpg), substr (substrate), pdt (product).

Tcp10 (C-A-PCP-COM<sup>D</sup> domain topology) and the N-terminal acceptor module of Tcp11 (C-A-PCP-E-C domain topology), corresponding to modules 3 and 4 (plus the C domain from module 5) of the teicoplanin NRPS (Figure S40). The tripeptide intermediate (D-Hpg-D-Tyr-L-Dpg) was chemically synthesized as a coenzyme A-linked thioester and

lography experiments; bioinformatics analyses; DKP and tetrapeptide production assays; and oligonucleotide primers (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

**Christopher D. Fage** – Department of Chemistry/  
Biochemistry, Philipps-Universität Marburg, 35032 Marburg,  
Germany; Department of Chemistry, University of Warwick,  
Coventry CV4 7AL, U.K.; [orcid.org/0000-0001-5015-644X](https://orcid.org/0000-0001-5015-644X); Email: [c.fage@warwick.ac.uk](mailto:c.fage@warwick.ac.uk)

**Józef R. Lewandowski** – Department of Chemistry, University  
of Warwick, Coventry CV4 7AL, U.K.; [orcid.org/0000-0001-6525-7083](https://orcid.org/0000-0001-6525-7083); Email: [j.r.lewandowski@warwick.ac.uk](mailto:j.r.lewandowski@warwick.ac.uk)

### Authors

**Simone Kosol** – Department of Chemistry, University of  
Warwick, Coventry CV4 7AL, U.K.

**Matthew Jenner** – Department of Chemistry, University of  
Warwick, Coventry CV4 7AL, U.K.; Warwick Integrative  
Synthetic Biology Centre, University of Warwick, Coventry  
CV4 7AL, U.K.

**Carl Öster** – Department of Chemistry, University of  
Warwick, Coventry CV4 7AL, U.K.; Present  
Address: Medicines Discovery Catapult, Alderley Park,  
Cheshire SK10 4TG, U.K.

**Angelo Gallo** – Department of Chemistry, University of  
Warwick, Coventry CV4 7AL, U.K.; [orcid.org/0000-0001-9778-4822](https://orcid.org/0000-0001-9778-4822)

**Milda Kaniusaite** – The Monash Biomedicine Discovery  
Institute, Department of Biochemistry and Molecular Biology,  
Monash University, Clayton, Victoria 3800, Australia;  
EMBL Australia, Monash University, Clayton, Victoria 3800,  
Australia; [orcid.org/0000-0002-1222-893X](https://orcid.org/0000-0002-1222-893X)

**Roman Steinbach** – Department of Chemistry/Biochemistry,  
Philipps-Universität Marburg, 35032 Marburg, Germany

**Michael Staniforth** – Department of Chemistry, University of  
Warwick, Coventry CV4 7AL, U.K.

**Vasilios G. Stavros** – Department of Chemistry, University of  
Warwick, Coventry CV4 7AL, U.K.; [orcid.org/0000-0002-6828-958X](https://orcid.org/0000-0002-6828-958X)

**Mohamed A. Marahiel** – Department of Chemistry/  
Biochemistry, Philipps-Universität Marburg, 35032 Marburg,  
Germany

**Max J. Cryle** – The Monash Biomedicine Discovery Institute,  
Department of Biochemistry and Molecular Biology, Monash  
University, Clayton, Victoria 3800, Australia; EMBL  
Australia, Monash University, Clayton, Victoria 3800,  
Australia; [orcid.org/0000-0002-9739-6157](https://orcid.org/0000-0002-9739-6157)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/acscatal.1c02113>

### Funding

The structure factor amplitudes and atomic coordinates of TycB3(E) were deposited in the RCSB Protein Data Bank under PDB code 6TA8. Trajectories (with frames saved every 1 or 5 ns) from MD simulations performed in this study are available via Zenodo under record 5021365 (<https://doi.org/10.5281/zenodo.5021365>).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

C.D.F. and M.A.M. acknowledge the Deutsche Forschungsgemeinschaft and the LOEWE Center for Synthetic Microbiology (SYNMIKRO), as well as the Engineering and Physical Sciences Research Council (EP/M027503/1) for financial support. 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; 639907). C.D.F., J.R.L., and S.K. also acknowledge funding from the Biotechnology and Biological Sciences Research Council (BB/L022761/1 and BB/R010218/1). M.J. is supported by a BBSRC Future Leader Fellowship (BB/R012121/1). The Bruker Maxis II instrument used in this study was funded by the BBSRC (BB/M017982/1). C.Ö. acknowledges funding from the European Union under a Marie Curie Initial Training Network FP7-PEOPLE-2012-ITN (316630). M.S. thanks the EPSRC for postdoctoral funding (EP/N010825/1). V.G.S. thanks the EPSRC for an equipment grant (EP/N010825/1) and the Royal Society and Leverhulme Trust for a Royal Society Leverhulme Trust Senior Research Fellowship. This work was supported by the National Health and Medical Research Council (APP1140619 to (M.J.C.)) and further supported under the Australian Research Council's Discovery Projects funding scheme (DP190101272 to M.J.C.). This research was conducted by the Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science (CE200100012) and funded by the Australian Government. We thank Beamline ID23-1 at the European Synchrotron Radiation Facility and the MarXtal Facility for support with crystallographic experiments, and Beamline B23 at the Diamond Light Source for support with SRCD experiments (project SM16655). We also thank Julien Tailhades for peptide synthesis.

## REFERENCES

- (1) Süssmuth, R. D.; Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew. Chem., Int. Ed.* **2017**, *56* (14), 3770–3821.
- (2) Caboche, S.; Pupin, M.; Leclere, V.; Fontaine, A.; Jacques, P.; Kuchero, G. NORINE: A Database of Nonribosomal Peptides. *Nucleic Acids Res.* **2007**, *36* (Database), D326–D331.
- (3) Walsh, C. T.; O'Brien, R. V.; Khosla, C. Nonproteinogenic Amino Acid Building Blocks for Nonribosomal Peptide and Hybrid Polyketide Scaffolds. *Angew. Chem., Int. Ed.* **2013**, *52* (28), 7098–7124.
- (4) Hahn, M.; Stachelhaus, T. Selective Interaction between Nonribosomal Peptide Synthetases Is Facilitated by Short Communication-Mediating Domains. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (44), 15585–15590.
- (5) Hahn, M.; Stachelhaus, T. Harnessing the Potential of Communication-Mediating Domains for the Biocombinatorial Synthesis of Nonribosomal Peptides. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (2), 275–280.
- (6) Chiocchini, C.; Linne, U.; Stachelhaus, T. In Vivo Biocombinatorial Synthesis of Lipopeptides by COM Domain-Mediated Reprogramming of the Surfactin Biosynthetic Complex. *Chem. Biol.* **2006**, *13* (8), 899–908.
- (7) Liu, H.; Gao, L.; Han, J.; Ma, Z.; Lu, Z.; Dai, C.; Zhang, C.; Bie, X. Biocombinatorial Synthesis of Novel Lipopeptides by COM Domain-Mediated Reprogramming of the Plipastatin NRPS Complex. *Front. Microbiol.* **2016**, *7*, 1801.
- (8) Cheng, Y.-C.; Ke, W.-J.; Liu, S.-T. Regions Involved in Fengycin Synthetases Enzyme Complex Formation. *J. Microbiol. Immunol. Infect.* **2017**, *50* (6), 755–762.

- (9) Mootz, H. D.; Marahiel, M. A. The Tyrocidine Biosynthesis Operon of *Bacillus Brevis*: Complete Nucleotide Sequence and Biochemical Characterization of Functional Internal Adenylation Domains. *J. Bacteriol.* **1997**, *179* (21), 6843–6850.
- (10) Tanovic, A.; Samel, S. A.; Essen, L.-O.; Marahiel, M. A. Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. *Science* **2008**, *321* (5889), 659–663.
- (11) Dehling, E.; Volkmann, G.; Matern, J. C. J.; Dörner, W.; Alfermann, J.; Diecker, J.; Mootz, H. D. Mapping of the Communication-Mediating Interface in Nonribosomal Peptide Synthetases Using a Genetically Encoded Photocrosslinker Supports an Upside-Down Helix-Hand Motif. *J. Mol. Biol.* **2016**, *428* (21), 4345–4360.
- (12) Kelley, L. A.; Mezulis, S.; Yates, C. M.; Wass, M. N.; Sternberg, M. J. E. The Phyre2 Web Portal for Protein Modeling, Prediction and Analysis. *Nat. Protoc.* **2015**, *10* (6), 845–858.
- (13) Liu, Y.; Zheng, T.; Bruner, S. D. Structural Basis for Phosphopantetheinyl Carrier Domain Interactions in the Terminal Module of Nonribosomal Peptide Synthetases. *Chem. Biol.* **2011**, *18* (11), 1482–1488.
- (14) Sundlov, J. A.; Shi, C.; Wilson, D. J.; Aldrich, C. C.; Gulick, A. M. Structural and Functional Investigation of the Intermolecular Interaction between NRPS Adenylation and Carrier Protein Domains. *Chem. Biol.* **2012**, *19* (2), 188–198.
- (15) Mitchell, C. A.; Shi, C.; Aldrich, C. C.; Gulick, A. M. Structure of PA1221, a Nonribosomal Peptide Synthetase Containing Adenylation and Peptidyl Carrier Protein Domains. *Biochemistry* **2012**, *51* (15), 3252–3263.
- (16) Drake, E. J.; Miller, B. R.; Shi, C.; Tarrasch, J. T.; Sundlov, J. A.; Leigh Allen, C.; Skiniotis, G.; Aldrich, C. C.; Gulick, A. M. Structures of Two Distinct Conformations of Holo-Non-Ribosomal Peptide Synthetases. *Nature* **2016**, *529* (7585), 235–238.
- (17) Reimer, J. M.; Aloise, M. N.; Harrison, P. M.; Martin Schmeing, T. Synthetic Cycle of the Initiation Module of a Formylating Nonribosomal Peptide Synthetase. *Nature* **2016**, *529* (7585), 239–242.
- (18) Marahiel, M. A. A Structural Model for Multimodular NRPS Assembly Lines. *Nat. Prod. Rep.* **2016**, *33* (2), 136–140.
- (19) Tarry, M. J.; Haque, A. S.; Bui, K. H.; Schmeing, T. M. X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. *Structure* **2017**, *25* (5), 783–793.
- (20) Bloudoff, K.; Fage, C. D.; Marahiel, M. A.; Schmeing, T. M. Structural and Mutational Analysis of the Nonribosomal Peptide Synthetase Heterocyclization Domain Provides Insight into Catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114* (1), 95–100.
- (21) Dowling, D. P.; Kung, Y.; Croft, A. K.; Taghizadeh, K.; Kelly, W. L.; Walsh, C. T.; Drennan, C. L. Structural Elements of an NRPS Cyclization Domain and Its Intermodule Docking Domain. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (44), 12432–12437.
- (22) Smith, H. G.; Beech, M. J.; Lewandowski, J. R.; Challis, G. L.; Jenner, M. Docking Domain-Mediated Subunit Interactions in Natural Product Megasynt(et)ases. *J. Ind. Microbiol. Biotechnol.* **2021**, *48* (3–4), kuab018.
- (23) Richter, C. D.; Nietlispach, D.; Broadhurst, R. W.; Weissman, K. J. Multienzyme Docking in Hybrid Megasyntetases. *Nat. Chem. Biol.* **2008**, *4* (1), 75–81.
- (24) Hacker, C.; Cai, X.; Kegler, C.; Zhao, L.; Weickmann, A. K.; Wurm, J. P.; Bode, H. B.; Wöhnert, J. Structure-Based Redesign of Docking Domain Interactions Modulates the Product Spectrum of a Rhabdopeptide-Synthesizing NRPS. *Nat. Commun.* **2018**, *9* (1), 4366.
- (25) Watzel, J.; Hacker, C.; Duchardt-Ferner, E.; Bode, H. B.; Wöhnert, J. A New Docking Domain Type in the Peptide-Antimicrobial-Xenorhabdus Peptide Producing Nonribosomal Peptide Synthetase from *Xenorhabdus Bovienii*. *ACS Chem. Biol.* **2020**, *15* (4), 982–989.
- (26) 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.* **2003**, *10* (8), 723–731.
- (27) Buchholz, T. J.; Geders, T. W.; Bartley, F. E.; Reynolds, K. A.; Smith, J. L.; Sherman, D. H. Structural Basis for Binding Specificity between Subclasses of Modular Polyketide Synthase Docking Domains. *ACS Chem. Biol.* **2009**, *4* (1), 41–52.
- (28) Whicher, J. R.; Smaga, S. S.; Hansen, D. A.; Brown, W. C.; Gerwick, W. H.; Sherman, D. H.; Smith, J. L. Cyanobacterial Polyketide Synthase Docking Domains: A Tool for Engineering Natural Product Biosynthesis. *Chem. Biol.* **2013**, *20* (11), 1340–1351.
- (29) Gay, D. C.; Wagner, D. T.; Meinke, J. L.; Zogzas, C. E.; Gay, G. R.; Keatinge-Clay, A. T. The LINKS Motif Zippers Trans-Acyltransferase Polyketide Synthase Assembly Lines into a Biosynthetic Megacomplex. *J. Struct. Biol.* **2016**, *193* (3), 196–205.
- (30) Zeng, J.; Wagner, D. T.; Zhang, Z.; Moretto, L.; Addison, J. D.; Keatinge-Clay, A. T. Portability and Structure of the Four-Helix Bundle Docking Domains of Trans-Acyltransferase Modular Polyketide Synthases. *ACS Chem. Biol.* **2016**, *11* (9), 2466–2474.
- (31) Jenner, M.; Kosol, S.; Griffiths, D.; Prasongpholchai, P.; Manzi, L.; Barrow, A. S.; Moses, J. E.; Oldham, N. J.; Lewandowski, J. R.; Challis, G. L. Mechanism of Intersubunit Ketosynthase–Dehydratase Interaction in Polyketide Synthases. *Nat. Chem. Biol.* **2018**, *14* (3), 270–275.
- (32) Dorival, J.; Annava, T.; Risser, F.; Collin, S.; Roblin, P.; Jacob, C.; Gruez, A.; Chagot, B.; Weissman, K. J. Characterization of Intersubunit Communication in the Virginiamycin Trans-Acyl Transferase Polyketide Synthase. *J. Am. Chem. Soc.* **2016**, *138* (12), 4155–4167.
- (33) Risser, F.; Collin, S.; Dos Santos-Morais, R.; Gruez, A.; Chagot, B.; Weissman, K. J. Towards Improved Understanding of Intersubunit Interactions in Modular Polyketide Biosynthesis: Docking in the Enacyloxin IIa Polyketide Synthase. *J. Struct. Biol.* **2020**, *212* (1), 107581.
- (34) Kosol, S.; Gallo, A.; Griffiths, D.; Valentic, T. R.; Masschelein, J.; Jenner, M.; de los Santos, E. L. C.; Manzi, L.; Sydor, P. K.; Rea, D.; et al. Structural Basis for Chain Release from the Enacyloxin Polyketide Synthase. *Nat. Chem.* **2019**, *11* (10), 913–923.
- (35) Samel, S. A.; Czodrowski, P.; Essen, L.-O. Structure of the Epimerization Domain of Tyrocidine Synthetase A. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70* (5), 1442–1452.
- (36) Chen, W.-H.; Li, K.; Guntaka, N. S.; Bruner, S. D. Interdomain and Intermodule Organization in Epimerization Domain Containing Nonribosomal Peptide Synthetases. *ACS Chem. Biol.* **2016**, *11* (8), 2293–2303.
- (37) Bozhüyü, K. A. J.; Linck, A.; Tietze, A.; Kranz, J.; Wesche, F.; Nowak, S.; Fleischhacker, F.; Shi, Y.-N.; Grün, P.; Bode, H. B. Modification and de Novo Design of Non-Ribosomal Peptide Synthetases Using Specific Assembly Points within Condensation Domains. *Nat. Chem.* **2019**, *11* (7), 653–661.
- (38) Bozhüyü, K. A. J.; Fleischhacker, F.; Linck, A.; Wesche, F.; Tietze, A.; Niesert, C.-P.; Bode, H. B. De Novo Design and Engineering of Non-Ribosomal Peptide Synthetases. *Nat. Chem.* **2018**, *10* (3), 275–281.
- (39) Leslie, A. G. W.; Moody, P. C. E.; Shaw, W. V. Structure of Chloramphenicol Acetyltransferase at 1.75-Å Resolution. *Proc. Natl. Acad. Sci. U. S. A.* **1988**, *85* (12), 4133–4137.
- (40) Stein, D. B.; Linne, U.; Marahiel, M. A. Utility of Epimerization Domains for the Redesign of Nonribosomal Peptide Synthetases. *FEBS J.* **2005**, *272* (17), 4506–4520.
- (41) Stein, D. B.; Linne, U.; Hahn, M.; Marahiel, M. A. Impact of Epimerization Domains on the Intermodular Transfer of Enzyme-Bound Intermediates in Nonribosomal Peptide Synthesis. *ChemBioChem* **2006**, *7* (11), 1807–1814.
- (42) Luo, L.; Kohli, R. M.; Onishi, M.; Linne, U.; Marahiel, M. A.; Walsh, C. T. Timing of Epimerization and Condensation Reactions in Nonribosomal Peptide Assembly Lines: Kinetic Analysis of Phenylalanine Activating Elongation Modules of Tyrocidine Synthetase B. *Biochemistry* **2002**, *41* (29), 9184–9196.
- (43) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New



Generation of Protein Database Search Programs. *Nucleic Acids Res.* **1997**, *25* (17), 3389–3402.

(44) Agarwala, R.; Barrett, T.; Beck, J.; Benson, D. A.; Bollin, C.; Bolton, E.; Bourexis, D.; Brister, J. R.; Bryant, S. H.; Canese, K.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46* (D1), D8–D13.

(45) Samel, S. A. Arbeiten Zur Strukturellen Charakterisierung von Thioesterase-, Kondensations- Und Epimerisierungsdomänen Nicht-ribosomaler Peptidsynthetasen. Ph.D. Thesis, Philipps-Universität Marburg, 2009.

(46) Lamley, J. M.; Iuga, D.; Öster, C.; Sass, H.-J.; Rogowski, M.; Oss, A.; Past, J.; Reinhold, A.; Grzesiek, S.; Samoson, A.; et al. Solid-State NMR of a Protein in a Precipitated Complex with a Full-Length Antibody. *J. Am. Chem. Soc.* **2014**, *136* (48), 16800–16806.

(47) Mainz, A.; Jehle, S.; van Rossum, B. J.; 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* (44), 15968–15969.

(48) Loquet, A.; Sgourakis, N. G.; Gupta, R.; Giller, K.; Riedel, D.; Goosmann, C.; Griesinger, C.; Kolbe, M.; Baker, D.; Becker, S.; et al. Atomic Model of the Type III Secretion System Needle. *Nature* **2012**, *486* (7402), 276–279.

(49) Gardienet, C.; Schütz, A. K.; Hunkeler, A.; Kunert, B.; Terradot, L.; Böckmann, A.; Meier, B. H. A Sedimented Sample of a 59 KDa Dodecameric Helicase Yields High-Resolution Solid-State NMR Spectra. *Angew. Chem., Int. Ed.* **2012**, *51* (31), 7855–7858.

(50) Manzi, L.; Barrow, A. S.; Scott, D.; Layfield, R.; Wright, T. G.; Moses, J. E.; Oldham, N. J. Carbene Footprinting Accurately Maps Binding Sites in Protein–Ligand and Protein–Protein Interactions. *Nat. Commun.* **2016**, *7*, 13288.

(51) Miao, Y.; Feixas, F.; Eun, C.; McCammon, J. A. Accelerated Molecular Dynamics Simulations of Protein Folding. *J. Comput. Chem.* **2015**, *36* (20), 1536–1549.

(52) Genheden, S.; Ryde, U. The MM/PBSA and MM/GBSA Methods to Estimate Ligand-Binding Affinities. *Expert Opin. Drug Discovery* **2015**, *10* (5), 449–461.

(53) Sugase, K.; Dyson, H. J.; Wright, P. E. Mechanism of Coupled Folding and Binding of an Intrinsically Disordered Protein. *Nature* **2007**, *447* (7147), 1021–1025.

(54) Dehling, E.; Rüschbaum, J.; Diecker, J.; Dörner, W.; Mootz, H. D. Photo-Crosslink Analysis in Nonribosomal Peptide Synthetases Reveals Aberrant Gel Migration of Branched Crosslink Isomers and Spatial Proximity between Non-Neighboring Domains. *Chem. Sci.* **2020**, *11* (33), 8945–8954.

(55) Stachelhaus, T.; Mootz, H. D.; Bergendahl, V.; Marahiel, M. A. Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. *J. Biol. Chem.* **1998**, *273* (35), 22773–22781.

(56) Eddy, S. R. Accelerated Profile HMM Searches. *PLoS Comput. Biol.* **2011**, *7* (10), e1002195.

(57) Medema, M. H.; Kottmann, R.; Yilmaz, P.; Cummings, M.; Biggins, J. B.; Blin, K.; de Bruijn, I.; Chooi, Y. H.; Claesen, J.; Coates, R. C.; et al. Minimum Information about a Biosynthetic Gene Cluster. *Nat. Chem. Biol.* **2015**, *11* (9), 625–631.

(58) Kaniusaite, M.; Goode, R. J. A.; Tailhades, J.; Schittenhelm, R. B.; Cryle, M. J. Exploring Modular Reengineering Strategies to Redesign the Teicoplanin Non-Ribosomal Peptide Synthetase. *Chem. Sci.* **2020**, *11* (35), 9443–9458.

(59) Balibar, C. J.; Vaillancourt, F. H.; Walsh, C. T. Generation of D Amino Acid Residues in Assembly of Arthrofactin by Dual Condensation/Epimerization Domains. *Chem. Biol.* **2005**, *12* (11), 1189–1200.

(60) Drozdetskiy, A.; Cole, C.; Procter, J.; Barton, G. J. JPred4: A Protein Secondary Structure Prediction Server. *Nucleic Acids Res.* **2015**, *43* (W1), W389–W394.

(61) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A. P.; Rempfer, C.; Bordoli, L.; et al. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303.

(62) Blin, K.; Shaw, S.; Steinke, K.; Villebro, R.; Ziemert, N.; Lee, S. Y.; Medema, M. H.; Weber, T. AntiSMASH 5.0: Updates to the Secondary Metabolite Genome Mining Pipeline. *Nucleic Acids Res.* **2019**, *47* (W1), W81–W87.

(63) Kaniusaite, M.; Tailhades, J.; Marschall, E. A.; Goode, R. J. A.; Schittenhelm, R. B.; Cryle, M. J. A Proof-Reading Mechanism for Non-Proteinogenic Amino Acid Incorporation into Glycopeptide Antibiotics. *Chem. Sci.* **2019**, *10* (41), 9466–9482.