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 Kaniusaitė, 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

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).1 In addition to the 20 canonical amino acids, more than 500 monomers can be incorporated into NRPs to enrich their chemical complexity.1−3 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′-phosphopantetheinyll (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).1 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
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. 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. 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\textsuperscript{TycA}) associates with the N-terminal acceptor COM region of TycB (COM\textsuperscript{TycB}), and the C-terminal donor COM region of TycB (COM\textsuperscript{TycB}) with the N-terminal acceptor COM region of TycC (COM\textsuperscript{TycC}) (Figure 1).

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. To elicit productive communication between more divergent modules, such as TycB\textsubscript{3} (the third module of TycB) and SrfA-C, substitution of COM\textsuperscript{D\textsubscript{TycB}} with COM\textsuperscript{D\textsubscript{TycA}} was found to be sufficient. Moreover, a single K9D mutation in the COM\textsuperscript{A\textsubscript{TycC}} region permitted productive interaction with the otherwise incompatible TycA. However, deletion of just six residues from the C-terminus of TycA disrupted productive binding with TycB. COM region substitutions have also led to product formation in vivo. Nevertheless, not all COM interface engineering attempts (particularly those targeting COM\textsuperscript{A} regions) have yielded the expected results. 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).\textsuperscript{10,13,22,14–21} 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 β-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;\textsuperscript{23,24} (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.\textsuperscript{25} In contrast, inter-subunit interactions in the evolutionarily related polyketide synthases (PKSs) have been more thoroughly characterized.\textsuperscript{22,26–33} Recently, hybrid PKS-NRPS interfaces between C-terminal carrier protein domains bearing short linear motifs and N-terminal C or Cy domains bearing β-hairpin docking domains were also examined.\textsuperscript{21,34} 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.\textsuperscript{22–27,31,32,34}

The crystal structure of the SrfA-C termination module (C-A-PCP-TE domain topology) revealed an artificial COM-like interface between subunits: an α-helical myc-His tag at the C-terminus of the TE domain, sharing modest sequence homology with COM\textsubscript{Sdfk.B} is grasped by a hand-like β-sheet on the N-terminal C domain of a second copy of SrfA-C within the crystal.\textsuperscript{10} 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).\textsuperscript{11} However, this approach combined subunits from two different NRPS pathways and relied on the incorporation of a bulky unnatural amino acid into the COM\textsubscript{D} region, potentially perturbing the nature of the interaction.\textsuperscript{11}

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\textsubscript{D} region.\textsuperscript{35,36}

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.\textsuperscript{37,38}

## RESULTS AND DISCUSSION

### Crystal Structure of a Chain Elongation Module E Domain

**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).\textsuperscript{12,36} 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\textsubscript{D} 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).\textsuperscript{35,36,39} The enzyme assumes a conformation nearly identical to that of TycA(E) and GrsA(E) from initiation modules (PDB codes 2XHG and 5ISX, respectively) (Figures 2 and S4). However, one salient difference can be seen in the bridge region of TycB3(E) (residues ~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 snak into the active site as in the latter two (Figures 2 and S4).

![Figure 2](image-url)}
S4). 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. 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.

Solution Structure of COMD Peptide. As with GrsA(E), no electron density is present for the COMD region of TycB3(E), despite its inclusion in the expression construct. Moreover, the region was necessarily excluded from TycA(E) to promote crystallization. COMD regions are predicted by the Phyre2 server to form an amphipathic α-helix preceded by disordered loops and a smaller α-helix (Figures 1b,c and S1). Given these findings, we surmised that in the absence of a downstream binding partner, the COMD 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 COMD fold.

Synchrotron radiation circular dichroism (CD) spectroscopy experiments were conducted on a solution of a synthetic 27-residue COMD region of TycA peptide (T1062-R1088 of TycA; Figure 1b) for secondary structure estimation. Our analysis suggested only minor helical content (~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 COMD. Indeed, addition of DPC above the critical micelle concentration induced a significant shift in the helical content of the peptide (~43%) (Figure 3a).

To visualize which residues of the COMD region adopt a helical conformation, we employed NMR spectroscopy. COMD region of TycA 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 α-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 α-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 COMD Peptide with an Acceptor Domain. To confirm whether the COMD region folds in the presence of TycB1(C), we prepared a sedimented complex of [13C,15N]-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 (34,46−49). 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) 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).

![Figure 3](https://doi.org/10.1021/acscatal.1c02113)
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 Cβ and Cγ 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 S16). This suggests that a large fraction of COMD$_{\text{TycA}}$ 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 COMD$_{\text{TycA}}$ peptide likely involves two highly flexible helices linked by a loop encompassing Q11 (Q1072 of TycA), as depicted in Figure 4.

**Solution-State NMR Titrations of COM$^D_{\text{TycA}}$ Peptide with an Acceptor Domain.** To follow changes in COMD$_{\text{TycA}}$ 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 $^1$N-COM$^D_{\text{TycA}}$ 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 COMD$_{\text{TycA}}$ 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–S18). 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 $K_a^*$ (or $1/K_a^*$) values range over 0.1–5 μM, in agreement with isothermal titration calorimetry (ITC) measurements (vide infra). When titrations of COMD$_{\text{TycA}}$ 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 COMD$_{\text{TycA}}$ 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.$^{31,34,50}$ 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 COMD$_{\text{TycA}}$ 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 COMD$_{\text{TycA}}$ 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 COMD$_{\text{TycA}}$ peptide. For TycB1(C), masking was observed for peptides D11-L28, T68-K74, and Q77-K84 at the N-terminus, corresponding to the β-hand motif and neighboring helix in a homology model (Figures 4b and S20; Tables S23–S27).$^{10,12}$ 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.)$^4$ Interestingly, residues D124-R128 of TycB1(C), modeled as a loop above strand s3 of the β-hand motif, showed significant unmasking, suggesting a conformational change that increases solvent exposure in this region, perhaps to accommodate binding of COMD$_{\text{TycA}}$ 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,$^{13,14}$
and L420-A428, which are far from the expected binding interface. Finally, peptide D48-T53 showed unmasking.

**Figure 5.** Folding of COM<sup>D</sup> <sub>TycA</sub> peptide in the presence of TycB1(C) during molecular dynamics simulations. Classical MD simulations (4.8 μs) 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 β-hand motif as a white cartoon and yellow surface, and COM<sup>D</sup> <sub>TycA</sub> 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).

Molecular Dynamics Simulations and In Silico Folding of COM<sup>D</sup> Peptides. The experimental results discussed above suggest that the COM<sup>D</sup> <sub>TycA</sub> region folds...
binding TycB1(C). To investigate whether consistent folding could be observed in MD simulations, we performed a series of cMD simulations (up to 5 μs) and accelerated molecular dynamics (aMD) simulations (up to 2 μs) starting from free COMD\textsubscript{TycA} peptide (corresponding to the NMR construct, with an N-terminal Gly followed by P1063-R1088 of TycA) or linker+COMD\textsubscript{TycA} 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). While the points of contact gleaned from our carbene footprinting experiments guided these simulations, various initial orientations of the peptides relative to the β-hand motif were explored to minimize bias. Simulations with linker+COMD\textsubscript{TycA} peptide were performed to confirm whether findings from experiments and simulations involving the shorter COMD\textsubscript{TycA} peptide apply when the entire linker is considered.

During the cMD simulation, COMD\textsubscript{TycA} peptide folded upon interaction with TycB1(C) (Figure S5a,b). Within ~2 μs, residues at the C-terminus assumed a dynamic helical orientation in a “helix down” orientation consistent with that proposed by Tanovic et al. (inset of Figure 1a). 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 β-hand motif, with interaction between R27 of COMD\textsubscript{TycA} 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 COMD peptide (E1074, E1075, D1077, and D1078 of TycA) and positively charged K6, K74, and K76 of the β-hand motif (see Figure S5b,c), as well as hydrophobic residues, particularly F19 and L22 on the COMD 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. In the presence of TycB1(C), the N- and C-terminal parts of linker+COMD\textsubscript{TycA} peptide folded into α-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 (~57 and ~38 kJ/mol, respectively), calculated via the MM/GBSA approach (Tables S30–S31), 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 COMD region. 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 COMD\textsubscript{TycA} 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 COMD\textsubscript{TycA} peptide binding. Since the β-hand motif binding the COMD\textsubscript{TycA} 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 COMD\textsubscript{TycA} 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 COMD region resembles that of the well-studied but biologically unrelated pKID/KIX system involved in eukaryotic gene transcription. 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 β-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 COMD helix up\textsubscript{TycA} or COMD\textsubscript{TycB} as well as the acceptor domain TycB1(C), and then carried out ITC measurements. Titration of holo-TycA, TycA(E), and COMD\textsubscript{TycA} Peptide with TycB1-C yielded K\textsubscript{i} values of 12.8 ± 4.8, 2.3 ± 0.7, and 5.9 ± 0.6 μM, respectively (Figure S26). These findings are comparable to the micromolar affinities previously obtained for docking domains at the termini of PKS and NRPS subunits, including the COM-facilitated interaction of apo-GrsA (63% identical to TycA)\textsuperscript{53} and TycB1\textsuperscript{54} as well as those estimated herein from NMR titrations (vide supra). As anticipated, COMD\textsubscript{TycB} 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 COMD\textsubscript{TycA} peptide bind TycB1(C) with similar affinities suggests that the COMD 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 diketopiperazinase (DKP) production by modules TycA (A-PCP-E-COMD domain topology) and TycB1 (C-A-PCP domain topology).\textsuperscript{11,55} In this assay, L-Phe and L-Pro are selected and activated by the A domains of TycA and TycB1, respectively, and then trans-acetylated onto adjacent holo-PCP domains for condensation by TycB1(C) (after E domain-catalyzed epimerization of i-Phe to d-Phe). The dipeptide intermediate d-Phe-i-Pro then undergoes autocyclization on TycB1 to yield free d-Phe-i-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-TycAΔCOMD (27 C-terminal residues deleted) was incubated with substrates L-Phe, i-Pro, ATP, and MgCl\textsubscript{2} in the presence and absence of COMD\textsubscript{TycA} peptide for 1 h at 30 °C prior to extracting and analyzing by LC-MS. In contrast to a previous report,\textsuperscript{11} we found TycAΔCOMD\textsubscript{TycA} was still able to generate a small but detectable amount of DKP when incubated with TycB1, revealing a nonessential but facilitative role for the COMD region in the condensation of aminoacyl 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 COMD\textsubscript{TycA} peptide in the
reaction containing TycAΔCOM\(^D\) 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\(^D\) 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\(^A\)TycC region enabled productive interaction with the noncognate donor TycA.\(^5\) The COM\(^A\)
regions of SrfA-C and GrsB, which are highly similar to
COM\(^A\)TycB and contain an Asp at the corresponding position
(S2), also productively interacted with their noncognate donor
TycA.\(^5\) Indeed, our cMD and aMD simulations indicate that
transitory interactions 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\(^D\) PpsB of the lipstatin
biosynthetic pathway permitted functional interaction with
the noncognate PpsE while reducing affinity for its cognate partner PpsC (Figure S31).\(^7\)

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)\(^{56}\)
from sequences presented by Hahn et al.\(^5\) and searched the
MIBiG repository;\(^7\) 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\(^D\)/C domains) in addition to *intra*-subunit COM interfaces (i.e., fused E-COM\(^D\)-C domains).\(^11\) 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 NRPs 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\(^D\) region in our searches)
should be retained when transplanting E domains from other
modules.\(^58\) This finding suggests an important role for fused
COM\(^D\) regions, likely in constraining NRPS subunit
architecture. We hypothesize that, like split COM\(^D\) regions,
fused COM\(^D\) 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\(^D\) regions at the
C-termini of PCP domains interact with C domains (i.e., split
PCP-COM\(^D\)/C domains) in the machineries responsible for
producing the last-resort antibiotics daptomycin, teicoplanin,
and vancomycin, for example (Table S32 and Figures S36–

When we simulated the folding of linker+COM\(^D\)TycA peptide
in the presence of a homology model of the noncognate
TycC1(C) (based on the C domain from SrfA-C),\(^10,22\) 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). However, in both simulations, we observed that folding
of the C-terminal helix was stabilized by hydrophobic patches
outside of the β-hand motif. In contrast, 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).

The findings presented herein support a role for
R1088 in defining the local conformation to which the
C-terminal donor helix assembles, consistent with our MD
simulations. These results also indicate that folding of the
C-terminal donor helix affects the folding of the COM
helix, as supported by the MD simulations. Moreover, the
simulations suggest that the C-terminal donor helix
assembles in a conformation that is stabilized by the
hydrophobic patch outside of the β-hand motif. It is
possible that this stabilization is achieved through hydrophobic
interactions with the noncognate donor TycA, which
potentially gatekeeping the tyrocidine NRPS subunit interac-
tions. In each case, these key side chains remained in close
proximity until the point of C-terminal donor helix assembly
(Figures S5c and S28). Notably, while the model proposed by
Tanovic et al.\(^10\) apparently predicts the correct orientation
for the C-terminal donor helix in the β-hand motif,\(^10\) it
offers no insight into the transient interactions that promote or
demote folding.
S37). While similar interfaces have been engineered through E domain deletions (e.g., A-PCP-COM\(^5\) domain topology), none occurring in nature have been reported until now. Furthermore, we identified previously undescribed COM\(^5\)-like sequences embedded near the N-termini of dual C/E domains\(^9\) 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\(^5\) regions remains to be determined. Encouragingly, the Jpred server predicted a helix on these noncanonical COM\(^5\) and COM\(^5\)-like regions, and the SWISS-MODEL server predicted an N-terminal \(\beta\)-hand motif on their associated C or dual C/E domains.\(^{60,61}\) Curiously, the antiSMASH library associated with the MiBIG repository does not appear to detect these variations on the COM\(^5\) theme.\(^7\)

**Validating a New Domain Architecture at the COM Interface.** To verify a role for the PCP-COM\(^5\) domain in \textit{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 \(\alpha\)-Hpg, ATP, and MgCl\(_2\) permitted A domain-catalyzed loading of \(\alpha\)-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 methyamine off-loading methodology,\(^63\) 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\(^5\) 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\(^5\) 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,\(^{53}\) 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 NPRS 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.\(^{31,33,34}\)

Our bioinformatics analyses unveiled the prevalence of both \textit{inter-} and \textit{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 \textit{intra}-subunit engineering methods such as the recently described exchange unit concept.\(^{35,37,38}\) 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-
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; Email: 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; Email: 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
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
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
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

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 6T8A. Trajectories (with frames saved every 1 or 5 ns) from MD simulations performed in this study are available via Zenodo under record 502136S (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 MaxSII instrument used in this study was funded by the BBSRC (BB/M017982/1). C.O. 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 MaxXtal 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

Structural and Mutational Analysis of the Nonribosomal Peptide Synthetase Proteins Revealing a Flexible Domains.


