




High-resolution structures of a siderophore-producing cyclization domain from *Yersinia pestis* offer a refined proposal of substrate binding

Received for publication, April 19, 2022, and in revised form, August 27, 2022. Published, Papers in Press, September 5, 2022.

<https://doi.org/10.1016/j.jbc.2022.102454>

Andrew D. Gnann¹, Yuan Xia¹ , Jess Soule¹ , Clara Barthélemy¹, Jayata S. Mawani¹, Sarah Nzikoba Musoke¹, Brian M. Castellano², Edward J. Brignole², Dominique P. Frueh³, and Daniel P. Dowling^{1,*}

From the ¹Department of Chemistry, University of Massachusetts Boston, Boston, Massachusetts, USA; ²Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; ³Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Edited by Ruma Banerjee

Nonribosomal peptide synthetase heterocyclization (Cy) domains generate biologically important oxazoline/thiazoline groups found in natural products, including pharmaceuticals and virulence factors such as some siderophores. Cy domains catalyze consecutive condensation and cyclodehydration reactions, although the mechanism is unknown. To better understand Cy domain catalysis, here we report the crystal structure of the second Cy domain (Cy2) of yersiniabactin synthetase from the causative agent of the plague, *Yersinia pestis*. Our high-resolution structure of Cy2 adopts a conformation that enables exploration of interactions with the extended thiazoline-containing cyclodehydration intermediate and the acceptor carrier protein (CP) to which it is tethered. We also report complementary electrostatic interfaces between Cy2 and its donor CP that mediate donor binding. Finally, we explored domain flexibility through normal mode analysis and identified small-molecule fragment-binding sites that may inform future antibiotic design targeting Cy function. Our results suggest how CP binding may influence global Cy conformations, with consequences for active-site remodeling to facilitate the separate condensation and cyclodehydration steps as well as potential inhibitor development.

Modular, multidomain polyketide (PK) synthases (PKSs), and nonribosomal peptide (NRP) synthetases (NRPSs) produce a variety of PK and NRP natural products (1–4). NRP, PK, and hybrid NRP–PK molecules display bioactivities with clear importance to humans (5–8), and the modularity of NRPS and PKS systems makes them prime candidates for bioengineering (9). In addition, NRP and PK secondary

metabolites are primarily produced in bacteria and, to a lesser extent, fungi, and some are virulence factors of pathogenic bacteria (10–14); therefore, a thorough understanding and control of NRPS and PKS catalytic machinery may afford new options for antibiotic targets (15).

NRPSs and PKSs follow an assembly line logic in which each module is responsible for activating and incorporating a specific building block. Canonical NRPS elongation modules consist of adenylation (A), condensation (C), and peptidyl carrier protein (PCP) domains. A domains activate and load specific substrates (e.g., amino acids) onto PCP phosphopantetheine (Ppant) cofactors as thioesters. PCP domains from contiguous upstream and downstream modules (herein called upstream/donor and downstream/acceptor PCPs) shuttle their thioester-tethered substrates to the intervening C domain, which catalyzes their condensation to yield an elongated and intermediate peptide product attached to the acceptor PCP. The product may then act as a donor intermediate in a subsequent condensation within the next module or be released hydrolytically within the final module. Substrates and intermediates are often modified by tailoring domains, giving rise to various oxidation states (16–19), methylation patterns (20–25), and other alterations (26–29). Of particular interest for the present work, many NRPSs contain a variant of C domains, the heterocyclization (Cy) domain, that converts hydroxyl- or thiol-containing peptides into heterocycles during chain elongation. For more information on the C domain family, see the 2017 review by Bloudoff and Schmeing (30).

Thiazoline and (methyl)oxazoline rings formed by Cy domains are prevalent and important for the bioactivity of many natural products (Fig. S1), a prime example being thiazol(in)e-containing siderophores such as yersiniabactin (Ybt) (31). Ybt is a virulence factor of *Yersinia* bacteria, the causative agents of the plague. Ybt is produced under iron-limiting conditions to scavenge trace metal ions from the extracellular environment, including Fe³⁺ from the host (32–34). Ybt is just one of a number of thi-/oxazol(in)e siderophores required for bacterial growth in iron-restricted environments (3, 12, 35–37), and understanding the molecular mechanisms of Cy domains may support development of a new class of antibiotics against these

* For correspondence: Daniel P. Dowling, daniel.dowling@umb.edu.

Present address for Yuan Xia: GenScript Biotech Corporation, Piscataway, NJ, 08854.

Present address for Jess Soule: University of California Los Angeles, Los Angeles, CA 90095.

Present address for Sarah Nzikoba Musoke: Brigham and Women's Hospital, Boston, MA 02115.

Present address for Brian M. Castellano: Stanford University, Stanford, CA 94305.

Structural and docking insights for the NRPS Cy domain

bacteria (38–40). Sequence diversity between Cy domains of different organisms, typically sharing <30% sequence identity, might permit development of selective inhibitors against Cy domains of pathogenic species, avoiding endogenous Cy domains of the human microbiome (41).

The biosynthesis of Ybt provides an example of a noncanonical NRPS–PKS system that entails the activity of several proteins encoded within a pathogenicity island (42) (Fig. 1). We focus on the second cyclization domain, Cy2, of the NRPS protein HMWP2. HMWP2 contains the starter module of the pathway and has the domain order ArCP–Cy1–A(E)–PCP1–Cy2–PCP2, where ArCP is an aryl CP and A(E) is an interrupted A domain embedding an unconventional 37 kDa epimerase domain (24, 29). HMWP1 contains Cy3 and PCP3 in a hybrid PKS–NRPS context. ArCP is loaded with a salicyl thioester by the standalone A domain YbtE, and the A(E) domain of HMWP2 is responsible for loading L-cysteine onto PCP1 and PCP2 (in *cis*) and PCP3 (in *trans*). The embedded E domain generates a mixture of enantiomers of the first thiazoline installed by HMWP2–Cy1, as would be derived from cyclodehydration of peptides formed with either D- or L-cysteine. Since Ybt only contains one stereoisomer, HMWP2–Cy2 is expected to provide stereospecificity. The product of

HMWP2–Cy2, Ppant-tethered 2-hydroxyphenylthiazolinylthiazolidine (2HPTT), is passed at an interprotein NRPS–PKS junction between HMWP2 and HMWP1. HMWP1 combines PKS and NRPS logics to further extend and release the compound, and YbtU reduces the second heterocycle of Ybt to thiazolidine. Both NRPS and PKS systems require movements of CPs to transfer elongating molecules; however, our understanding of NRPS–PKS interfaces outside docking domain studies is limited (43). Therefore, the rich noncanonical architecture of Ybt synthetase is of great interest for exploring structure–function relationships in hybrid NRPS–PKS systems. As a first step toward modeling the NRPS to PKS interface between HMWP2 and HMWP1, we explore the HMWP2–Cy2 domain, which with its acceptor PCP (HMWP2–PCP2) forms the NRPS half of the NRPS–PKS junction.

NRPS Cy domains are typically responsible for both condensation and cyclodehydration reactions, affording (methyl)ox-/thiazoline modifications (Figs. 1 and S2), and their catalytic mechanisms are only partially understood. Five structures of Cy domains have been reported, involved in biosynthesis of epothilones (44), bacillamide (45), Ybt (46), JBIR-34/35 (47), and pyochelin (48). The first two Cy domain

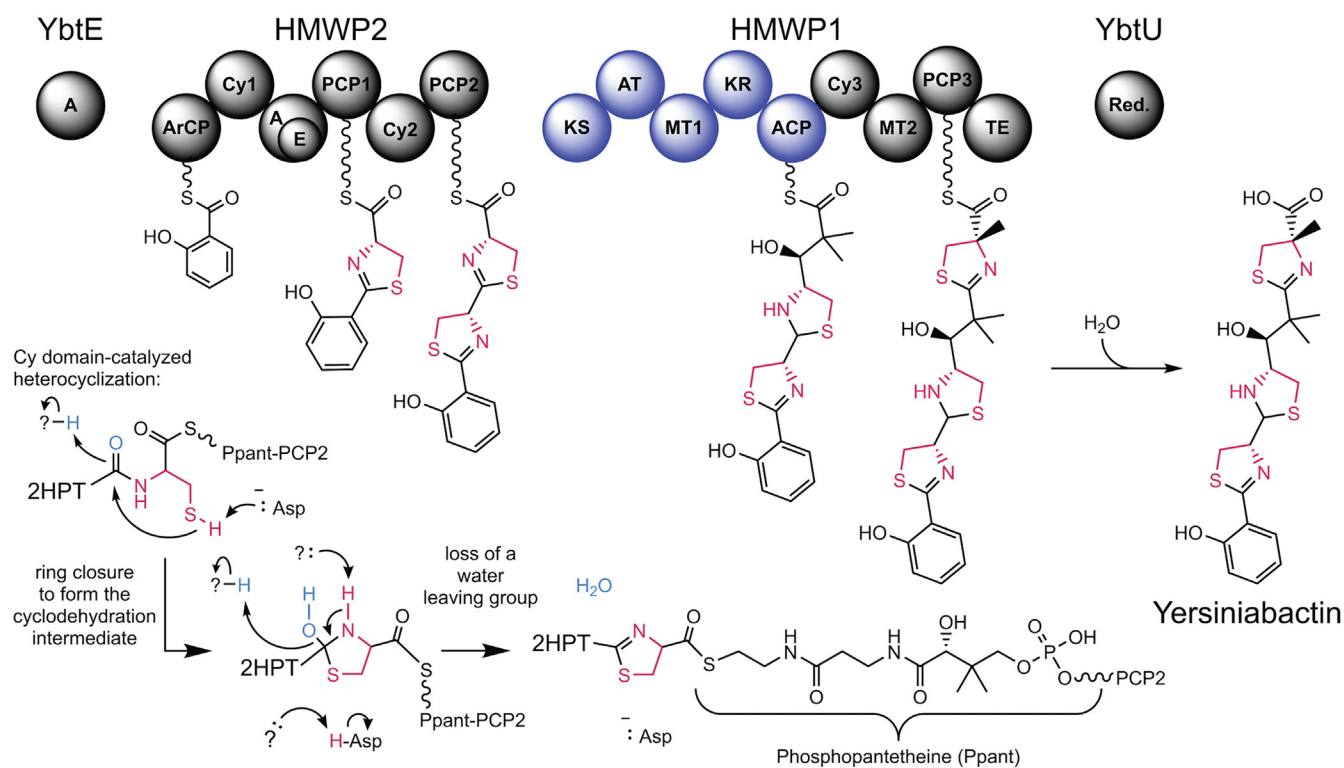


Figure 1. The four protein, hybrid NRPS–PKS, modular assembly line for biosynthesis of yersiniabactin and a mechanistic proposal for HMWP2–Cy2-catalyzed heterocyclization. Each circle represents either an NRPS domain (black) or a PKS domain (blue). Bond-line structures are substrates/products/intermediates carried by their respective carrier proteins. Domain labeling (left to right), and descriptions for PKS domains not detailed in the main text, are A, adenylation; ArCP, aryl carrier protein; Cy, heterocyclization; E, epimerization; PCP, peptidyl carrier proteins; KS, ketosynthase (catalyzes Claisen condensation of acyl donor and malonyl acceptor groups); AT, acyl transferase (catalyzes tethering of malonyl units onto the acyl carrier protein [ACP]); MT, methyltransferase (catalyzes methyl transfer from S-adenosyl methionine to cysteinyl or malonyl C2 positions); KR, ketoreductase (catalyzes NAD(P)H-dependent reduction of the keto group from the acyl donor substrate of the KS); ACP, acyl carrier protein; TE, thioesterase (catalyzes hydrolytic cleavage of the product thioester); red., reductase. (Not shown: YbtS, synthesizes input salicylate from chorismate; YbtD, transfers phosphopantetheine [Ppant] to CPs; YbtT, type II thioesterase putatively hydrolyzing improperly loaded substrates). The Ppant cofactor is represented as a wavy line-S and is drawn out in the bottom right of the figure. Atoms in light blue in the partial mechanism in the bottom left constitute the water-leaving group. Thiazoline heterocycle atoms derived from cysteine are colored magenta throughout the figure. NRPS, nonribosomal peptide synthetase; PKS, polyketide synthase.

structures identified a structural role for aspartates of the characteristic DXXXXD motif, which was in keeping with the roles of corresponding residues in C domains (30, 44, 45). The Cy domain's active site, formed in a tunnel between its N- and C-terminal subdomains, contains a putative catalytic aspartate–threonine dyad (44, 45) as well as a nearby glutamine (44), mutations of which severely affected cyclodehydration. Interestingly, the glutamine is not present in BmdB-Cy2, suggesting complementation of its yet unknown role by some other means. Several additional conserved residues affect product formation but serve unknown roles. Notably, HMWP2-Cy1 has been shown to display global dynamics, and a single mutation leads to a global molecular response (46), highlighting the limitations of traditional structural studies relying on a single static structure and the need to identify conformations critical for function. Finally, there is limited structural information for binding of (amino) acyl/peptidyl-Ppant in a Cy domain or for binding CPs in catalytically competent states, with the recent cryo-EM structures of FmoA3 (47) and PchE (48) providing the first snapshots of CP binding to Cy domains within modular contexts. Overall, more structural snapshots of Cy domains are needed to resolve important gaps in understanding how they function.

Here, we report the 1.94 Å resolution crystal structure of the Cy2 domain of Ybt synthetase (HMWP2-Cy2), which with its acceptor PCP (HMWP2-PCP2) forms the NRPS half of the NRPS–PKS interface. In addition to the role of Ybt in pathogenesis for *Yersinia pestis*, this cyclization domain is of interest because it accepts a donor that already features a heterocycle and because Cy2 is expected to govern the stereochemistry of the end product, Ybt (29). We construct protein–protein docking models suggesting interactions in the tridomain complex HMWP2–PCP1–Cy2–PCP2 contributing to the synthetic directionality of HMWP2. We then take advantage of a less constricted active-site tunnel to report covalent ligand docking experiments of intermediates during

cyclodehydration tethered to PCP2 as oriented in the protein–protein docking result. The results narrow the possible positioning of a cyclodehydration intermediate, and we discuss the impact of the flexibility of HMWP2-Cy2, estimated through normal mode analysis, on positioning the condensation intermediate that precedes cyclization. The Cy conformation we present here provides important clues on communication between Cy domains and their substrates and partner CP domains, suggesting relay mechanisms relying on structural fluctuations that may couple domain engagements with active-site remodeling and global remodeling of the entire Cy domain.

Results

Structure of a Cy domain from siderophore biosynthesis

Two X-ray datasets were obtained for two crystals of HMWP2-Cy2. The first dataset was solved to 1.94 Å resolution by molecular replacement using the structure of the Cy domain from epothilone biosynthesis (44) (EpoB-Cy; 35% sequence identity) as a search model. The second dataset was solved to 2.35 Å resolution using the 1.94 Å resolution HMWP2-Cy2 model (Table S1). These structures demonstrate the pseudodimeric fold characteristic of the condensation domain superfamily, which is a two-lobed arrangement of two chloramphenicol acetyltransferase–like folds (Figs. 2A and S3). The chloramphenicol acetyltransferase–like folds, called the N- and C-terminal subdomains here, each contain a mixed β sheet with intervening α helices. The two subdomains are linked *via* α helices 5 and 6, which resemble a hinge, with the first ~180 residues forming the N-terminal subdomain and the remainder forming the C-terminal subdomain. The C-terminal β sheet is formed by strands 7, 13, 12, 10, 8, and 9 (in order), and the N-terminal β sheet is formed by strands 4, 5, 6, 1, and 11, with strand 11 donated from the C-terminal subdomain. Strand 11 and the preceding loop have been referred to as a latch and are thought to mediate partial dissociation of the

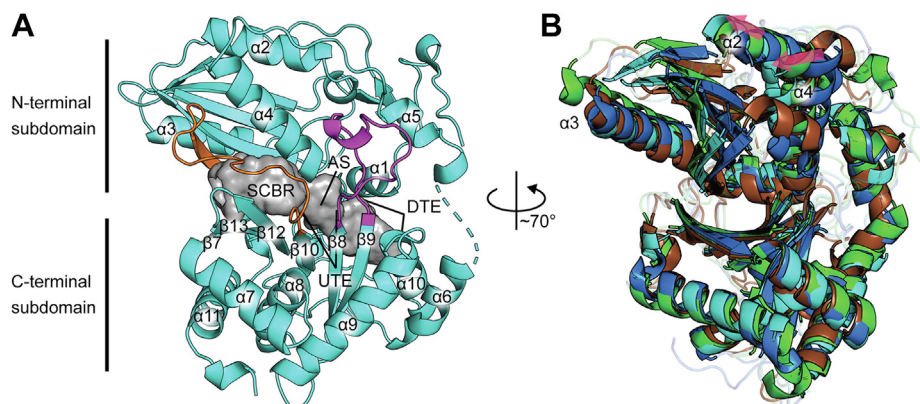


Figure 2. The HMWP2-Cy2 crystal structure exhibits a more open active-site tunnel. *A*, an overview of the 1.94 Å resolution HMWP2-Cy2 crystal structure with the tunnel as a gray surface, the floor loop and latch in light magenta and orange cartoon, respectively, and other features discussed in the text labeled (AS, active site; DTE, downstream tunnel entrance; SCBR, side chain-binding region; UTE, upstream tunnel entrance). *B*, Cy domain C-terminal subdomain Ca alignment demonstrates intersubdomain conformational variation among Cy domains. HMWP2-Cy2 is cyan (PDB ID: 7JTJ), EpoB-Cy is green (PDB ID: 5T7Z), BmdB-Cy2 is blue (PDB ID: 5T3E), and PchE-Cy is brown (PDB ID: 7EN1). Loop regions are transparent. Pink arrows highlight correlated variation between the models.

Structural and docking insights for the NRPS Cy domain

subdomains (Fig. 2A) (49). The two points of crossover between the subdomains are the predominantly loop region between β -strands 8 and 9 in the C-terminal subdomain (floor loop) and the latch.

HMWP2-Cy2 crystallizes in the space group $P4_12_12$ with one molecule in the asymmetric unit. The largest interface area between Cy2 molecules within the crystal buries $\sim 680 \text{ \AA}^2$ (3.6%) of the solvent-accessible surface area, based on analysis using the PISA server (50); therefore, HMWP2-Cy2 is likely a monomer in solution, consistent with the domain's behavior by size-exclusion chromatography (Fig. S4). Electron density maps obtained using the 1.94 \AA resolution data permitted building residues Q1483–L1663 of the N-terminal subdomain and N1669–Q1910 of the C-terminal subdomain (HMWP2 numbering) but did not support building the hinge linker between the N- and C-terminal subdomains. Fortunately, the lower-resolution dataset includes weak electron density for the missing residues except P1666, which remains disordered, allowing for generation of a composite model using coordinates from the higher-resolution dataset with residues from the hinge linker derived from the 2.35 \AA resolution structure. P1666 was built and refined to reduce potential energy in the final composite model and achieve acceptable connectivity. The weak or missing electron density for the connection at the hinge helices is consistent with greater flexibility in this area of HMWP2-Cy2.

Of the prior Cy domain structures, the Cy domain from EpoB (44) most closely resembles HMWP2-Cy2 (RMSD of 1.4 \AA between 338 of 431 $C\alpha$ atoms, aligning against Protein Data Bank ([PDB] ID: 5T7Z). Interestingly, protein alignments improve when considering either the N-terminal subdomain (residues 1483–1665) or the C-terminal subdomain (residues 1666–1910) independently, consistent with movement of these subdomains relative to each other within published NRPS Cy domain structures (Fig. 2B) (44–48). Using a $C\alpha$ alignment of the C-terminal subdomains of HMWP2-Cy2 with other Cy domains depicts a twisting of the N-terminal subdomain β sheet, with concomitant repositioning of helices $\alpha 2$ and $\alpha 4$ on one side of the N-terminal subdomain β sheet and of $\alpha 3$ on the opposite side.

The conformation of HMWP2-Cy2 defines an open and continuous tunnel, with a solvent-accessible surface volume of $\sim 543 \text{ \AA}^3$ (Figs. 2 and S5). As discussed later, the acceptor site is open, whereas the donor-binding site appears closed off, generating a tunnel from the donor substrate side chain-binding region (SCBR) to the acceptor site. Although this tunnel has been observed in EpoB-Cy, BmdB-Cy2, and condensation domains (44, 45, 51, 52), the tunnels of EpoB-Cy and BmdB-Cy2 are more constricted (Fig. S5) with the BmdB-Cy2 tunnel returning a volume of $\sim 382 \text{ \AA}^3$ and EpoB-Cy's appearing as three chambers collectively measuring $\sim 199 \text{ \AA}^3$. In particular, the EpoB-Cy and BmdB-Cy2 structures are constricted in the vicinity of the donor substrate SCBR.

The putative catalytic dyad residues (D1862 and T1830) and helper residue (Q1864) are found in conformations resembling EpoB-Cy (44) and BmdB-Cy2 (45). Intriguingly, the HMWP2-Cy2 active site contains an unexpected metal-binding site against $\beta 1$ and the latch strand. This metal site involves three backbone carbonyl groups (S1520, H1521, and Q1855), a

carboxamide (N1621 following the first aspartate of the DXXXXD motif), and two crystallographic waters, giving rise to a distorted octahedral coordination geometry (Fig. S6). This site refines as a sodium with a B -factor of 28.1 \AA^2 compared with an average of 24.8 \AA^2 for protein atoms with which it interacts. The position after the first aspartate of the DXXXXD motif is most commonly leucine or methionine, with asparagine being the fourth most common (Fig. S7). There also does not appear to be any preference for a polar amino acid at the second position of the DXXXXD motifs within a subset of Cy domain sequences selected from siderophore NRPSs (Fig. S8).

CP interaction surfaces

HMWP2-PCP1 is expected to bind to Cy2 at the N terminus of $\alpha 8$, where the donor Ppant would extend between the C termini of $\beta 8$ and $\beta 10$ in the C-terminal subdomain, near the N termini of helices $\alpha 4$ and $\alpha 8$ (Fig. S9A) (48, 51). For the binding site to become accessible to a donor Ppant in HMWP2-Cy2, $\beta 10$ would need to splay away from $\beta 8$ by 2 to 3 \AA relative to its position in the crystal structure (Fig. S9B). The NMR solution structure of HMWP2-PCP1 (53) was used to investigate PCP1-Cy2 interactions with the conformation displayed by Cy2. To better mimic a tethered Ppant, a HMWP2-PCP1 model bearing a phosphoserine at S1439 (SEP1439) was prepared and docked against HMWP2-Cy2 using HADDOCK2.4 (54–56). Guiding the phosphoserine toward the expected upstream tunnel entrance (which is closed in HMWP2-Cy2) returned clusters of poses (Figs. 3A and S9A) that resemble what is observed in linear gramicidin synthetase, LgrA-PCP1-C2 (51) and the recently reported cryo-EM PchE structure (48) (Fig. S9A). The surface electrostatic potential reveals a charge complementarity footprint that favors this binding mode (Fig. 3B). Interacting polar/charged residue pairs include (in HMWP2 numbering) E1836_{Cy}-R1444_{PCP}, K1717_{Cy}-D1438_{PCP}, and N1801_{Cy}-S1459_{PCP}, and there is hydrophobic packing around L1800_{Cy} and F1462_{PCP} (Fig. 3, C and D). The position of HMWP2-PCP1 permits the six-residue interdomain linker omitted in the docking calculation to reach HMWP2-Cy2 without requiring rearrangements.

At the downstream tunnel entrance, which interacts with the acceptor HMWP2-PCP2, loop 1 (after $\alpha 1$) extends further toward $\alpha 10$ than in EpoB-Cy (44) or BmdB-Cy2 (45), in which it is blocked by the loop following $\alpha 10$ (Fig. S10, A–C). Docking experiments similar to those conducted for HMWP2-PCP1 were conducted with a homology model of HMWP2-PCP2 built on the PCP from *Streptomyces* sp. Acta 2897 ($\sim 34\%$ sequence ID, PDB ID: 4PWV) (16) modified with a phosphoserine at S1977 (SEP1977). Restrained docking in HADDOCK2.4 returned clusters (Fig. 3E) that resemble known PCP_{acceptor} complexes (48, 52, 57, 58). Surface electrostatic potentials again reveal the footprint of charge complementarity that favors this binding mode (Fig. 3F). This interface involves hydrophobic packing of V1507_{Cy} against L1978_{PCP} and V1981_{PCP} and of P1511_{Cy} against L1997_{PCP}, as well as stacking of the phenyl ring of the weakly conserved F1513_{Cy} with both F1971_{PCP} and F2001_{PCP} (Fig. 3, G and H). The other main contribution to HMWP2-PCP2 binding

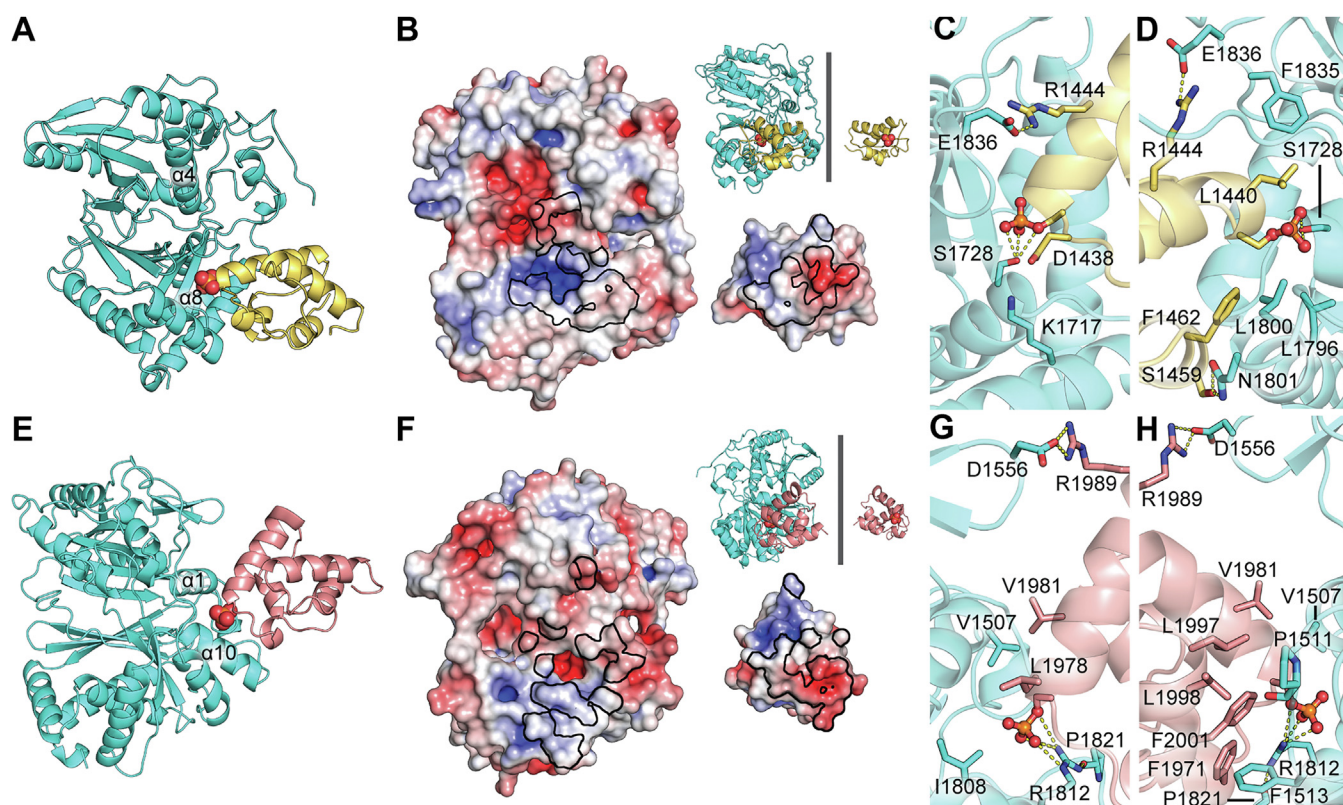


Figure 3. Protein–protein docking poses of HMWP2–PCP1 and HMWP2–PCP2 with HMWP2–Cy2. A, a model of HMWP2–PCP1 (yellow cartoon with the phosphate of phosphoserine in spheres with elemental coloring) is shown bound at the expected upstream PCP-binding site of HMWP2–Cy2 (cyan). B, electrostatic potential maps of the upstream PCP-binding site of HMWP2–Cy2 (left) and the Cy-binding site of HMWP2–PCP1 (right) are related here by a rotation of 180° so that the interaction surfaces of both components are displayed. The orientations of the components can be observed in the inset cartoon representations. Electrostatic potential maps use a color scale from -5 to $+5$ $k_B T/e_e$ (red to blue). The interaction footprint shared by the proteins is outlined in black. C and D, detail of interacting residues (sticks) at the PCP1–Cy2 interface (phosphoserine is displayed in ball and sticks here). E–H, same as A–D but for the downstream HMWP2–PCP2 interaction (PCP2 is in salmon).

according to our model is from electrostatic interaction between a conserved arginine in $\alpha 10$ (R1812) and SEP1977.

Near the downstream interaction surface, two regions of interest differ in Cy domain structures reported to date. First, a conserved arginine in $\alpha 1$ (R1509) forms a salt bridge with the first aspartate of the DXXXXD motif in EpoB–Cy (44) and PchE–Cy1 (48) but is oriented away from this residue in HMWP2–Cy2 (Fig. S10). In BmdB–Cy2 (45), the arginine adopts an intermediate position. Similarly, a conserved glutamine in the loop after the latch (Q1858) forms hydrogen bonds to loop 1 in HMWP2–Cy2 but projects its side chain into the tunnel in EpoB–Cy (and PchE–Cy1), whereas it adopts an intermediate position in BmdB–Cy2 (Figs. S10 and S11). Positioning of R1509 and Q1858 may provide a link between binding of PCP2, remodeling of its binding site, and remodeling of the active-site tunnel for catalysis in the presence of substrates.

Exploration of cyclodehydration intermediate binding in HMWP2–Cy2 highlights potential roles for active-site tunnel residues

We sought to determine whether the active-site tunnel of the HMWP2–Cy2 crystal structure could accommodate its cognate cyclodehydration intermediate, following cyclization but

preceding dehydration. Because nucleophilic attack during cyclization can occur on either face of the condensation product amide, we employed rigid receptor covalent docking to sample binding modes of both *2R* and *2S* diastereomers of the intermediate mimic (*2R/S,4R*)-2,2-hydroxy-(2'-(2"-hydroxyphenyl)-thiazolynyl)-thiazolidinyl Ppant (Ppant-2HPTT-OH) (Fig. S12). The covalent docking approach permits identification of docked poses that are compatible with covalent linkage to the conserved serine of our HMWP2–Cy2–PCP2 complex (SEP1977). Both diastereomers of the hydroxythiazolidine moiety docked with favorable Glide scores in the HMWP2–Cy2 tunnel, returning in total 120 poses, which were then refined. Models of top-ranked poses obtained by this approach are displayed in Figs. S13–S18. These poses can be organized into three categories based on the direction in which the leaving group oxygen of the hydroxythiazolidine moiety is directed: (1) hydroxyl roughly toward the N-terminal end of helix $\alpha 4$, (2) hydroxyl toward the N-terminal subdomain β sheet, or (3) hydroxyl roughly toward the putative catalytic dyad. Depending on the chirality, the other reactive groups of the intermediate, namely the thiazolidine ring N–H and S, interact with alternative regions of Cy2. Tests with both protonation states of the active-site dyad favored the neutral form (presumed after cyclization) for both diastereomers. Figure 4 shows the top pose for the *R* intermediate, which directs the leaving group

Structural and docking insights for the NRPS Cy domain

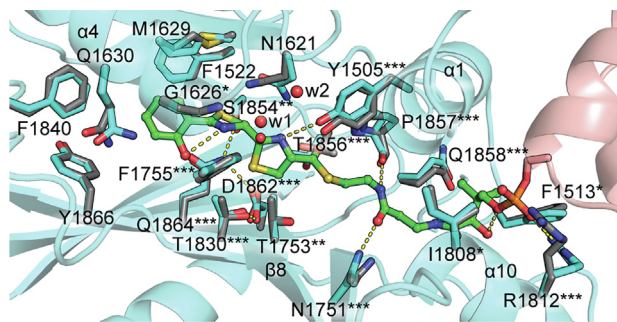


Figure 4. The HMWP2-Cy2 crystal structure acceptor tunnel and active site accommodate a cyclodehydration intermediate. Covalent docking of Ppant-2HPTT-OH in an HMWP2-Cy2-PCP2 complex based on the Cy2 crystal structure and PCP placement from protein-protein docking supports a model in which catalysis proceeds through an *R* hydroxythiazolidine intermediate (green) in HMWP2-Cy2 (refined covalent docking Cy model in cyan and PCP in salmon with transparent cartoon, and crystal structure coordinates for displayed residues in gray). This pose (#10 in Fig. S16) is selected as a top candidate based on its Glide docking score from covalent docking, MM-GBSA-approximated binding energy for the corresponding noncovalent phosphonate-S1977A complex and ligand and receptor strain energies. Two crystallographic water sites (red spheres) are shown in the vicinity of the active site (w1 and w2). HMWP2 residues marked with asterisks agree with Cy domain conservation trends at positions > ca. 2 bits in the sequence logo (Fig. S7) that are somewhat conserved (*property conserved or HMWP2-Cy2 residue in at least ~25% of aligned sequences), moderately conserved (**primarily two options and HMWP2-Cy2 residue at least ~50%), or highly conserved (**overwhelmingly one option). MM-GBSA, molecular mechanics with generalized Born and surface area; PCP, peptidyl carrier protein; Ppant, phosphopantetheine.

oxygen (previously of the carbonyl electrophile) toward the N terminus of helix α 1. Excitingly, the top *R* intermediate pose places the hydroxythiazolidine moiety between the positions expected for the nucleophile and electrophile based on existing structures (51, 57, 59) (Fig. S19), and the placement agrees well with the product-bound structure reported for PchE-Cy1 (48) (Fig. S20).

Although comparable docking metrics were obtained for the top *R* and *S* poses, the *R* hydroxyl-toward- α 4 binding mode warrants further consideration for several reasons. First, in one category of poses, the leaving group oxygen is directed roughly toward the partial positive charge of the N terminus of α 4. This position is more compatible with expectations based on trapped condensation states (51) and with the cyclodehydration model proposed by Bloudoff *et al.* (45), although the leaving group oxygen is directed away from T1830, rather than toward it. This set of poses places the ring sulfur in the middle of the catalytic dyad and the ring nitrogen such that it could form hydrogen bonds with the hydroxyls of Y1505 and T1856. The Ppant arm interacts with N1751, P1857, and Q1858, and its dimethyl group is rotated toward the loops following helices α 1 and α 10. Alternative poses suggest that T1856 could interact with the Ppant-2HPTT-OH thioester. Two crystallographic water molecules found in the active site between helix α 4, the DXXXXD motif loop, and Y1505 were found to be roughly compatible with Ppant-2HPTT-(*R*)OH positioning, although the water molecule nearest helix α 4 might instead resemble positioning of the water-leaving group. Additional observations within the full set of docking results, including descriptions of possible donor side-chain interactions, are

presented in the *Supporting Results* section of the [Supporting information](#) document.

Low-frequency vibrational normal mode analysis for HMWP2-Cy2

Because Cy domain crystal structures without donor CPs display closed conformations that preclude donor substrate binding, we explored the flexibility of HMWP2-Cy2 around its putative cyclodehydration-competent conformation using normal mode analysis. Although our top cyclodehydration intermediate candidate pose is promising, it is also interesting to consider how flexibility of the Cy domain alters relationships between groups within the active site and remodels donor and acceptor CP-binding sites.

The 10 lowest-frequency normal modes of the HMWP2-Cy2 conformation were calculated using a model supplemented with water molecules to fill voids (primarily in the tunnel). The observed modes involve relative motions of the N- and C-terminal subdomains and affect the shape of upstream and downstream entrances to different extents (see movies in the [Supporting information](#) section). This flexibility is reminiscent of variations among crystal structures. Strikingly, rearrangements of both tunnel entrances co-occur across the calculated modes, supporting the idea that remodeling of one entrance upon CP binding may be coupled with global conformational change affecting the other entrance.

Mode 5 is a highly localized motion involving R1817 at the C terminus of α 10 at the downstream opening. R1817 forms a salt bridge with D1679 in α 6, following the hinge region that is disordered in the HMWP2-Cy2 crystal structures. Although mode 5 does not directly involve this hinge, the hinge undergoes motion in all the other modes. Higher-frequency modes 7–10 feature more motions within subdomains, especially between two lobes of the C-terminal subdomain consisting of (1) the floor loop, the strands preceding and following it (β 8/9) and helices α 6, 9, and 10; and (2) strands β 7, 10, 12, and 13 and helices α 7, 8, and 11. We note that mode 7 resembles an amplified combination of the lower-frequency modes in that rotations of N- and C-terminal subdomains result in modulation of the α 3- α 11 distance and displacement of α 10 relative to α 1 (Fig. S21). These motions result in opening of the upstream tunnel entrance by over 3 Å, which would accommodate donor Ppant binding.

We next considered relative motions of residues in the first and second shells of the active site (Fig. 5). Several modes demonstrate modulation of the active-site dimensions in ways that would influence substrate binding. As mentioned previously, many of the modes influence access to the upstream tunnel entrance, measured by displacement of L1754 and S1831. Modes 1, 7, 8, and 10 modulate width along an axis between the floor loop and the latch strand (T1772–T1856). Modes 7 and 9 modulate the distance between the conserved aspartate of the floor loop (D1770) and the active-site aspartate (D1862), as well as the distance between the dyad threonine (T1830) and T1856 in the latch. Modes 6 to 10 appreciably

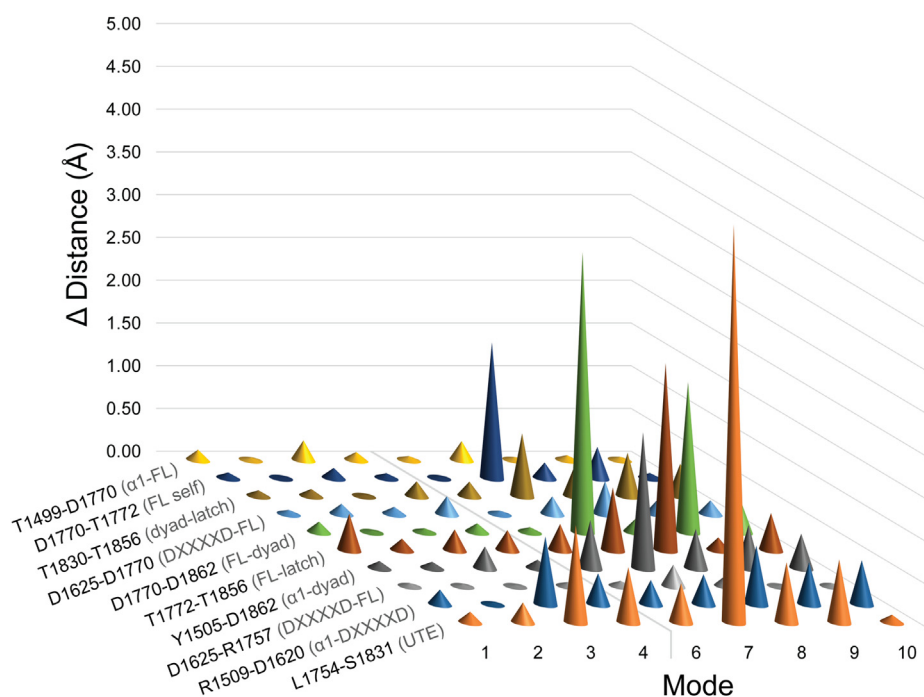


Figure 5. Change in inter-residue distances with the top 10 low-frequency normal vibrational modes of HMWP2-Cy2. A model of HMWP2-Cy2 returns a variety of motions in the normal mode analysis, primarily involving changes of orientation between the N- and C-terminal subdomains. Marked in the rows of this plot are inter-residue distances relevant for considerations of active-site shape and/or the structure of PCP/Ppant-binding regions. See [Figures 4](#), [S9B](#), and [S10C](#) for the structural contexts of these residues. The *light gray line* separating modes 4 and 6 indicates omission of mode 5, which involved only side-chain motion of R1817 and did not influence the measurements presented here. PCP, peptidyl carrier protein; Ppant, phosphopantetheine.

modulate the distance between Y1505, which interacts with the substrate in the acceptor tunnel, and the active-site aspartate.

These modes involve motions associated with the transition between open-for-condensation and closed-for-cyclodehydration conformations, bringing the catalytic dyad toward and away from Y1505. These observations, when considered alongside our top *R* cyclodehydration intermediate pose, suggest that global motion may couple to substrate rearrangements after condensation to obtain a cyclization reaction complex. Notably, the flexibility predicted by this analysis occurs without substantial displacement of the salt bridge D1625–R1757 flanking the upstream entrance, which is consistent with a PCP–C interface from LgrA (PDB ID: 6MFX) ([51](#)) but contrasts the recent PchE–ArCP–Cy1 cryo-EM model, which shows this salt bridge broken in the CP–Cy complex presumed to resemble the open-for-condensation state. The other salt bridge between R1509 and D1620, linking the DXXXXD motif to helix α 1, does undergo slight displacements in modes 3 to 10, which is reminiscent of variation at this position in the Cy crystal structures. The other interesting interaction with α 1, the hydrogen bond between T1499 and D1770 in the floor loop, is relatively consistent across the modes. How these networks of interactions are maintained or broken across the reaction cycle of Cy domains will be interesting, as modulation of their environments by CP binding or chemical changes in the active site may be key for productive trajectories through the conformational landscape.

Potential inhibitory hot spots for HMWP2-Cy2

To identify locations within the HMWP2-Cy2 structure for future inhibitor development, we analyzed the structure with the FTMap server, which returns sites for a set of chemical probes ([Fig. 6](#)) ([60](#)). Several hot spots were identified within the protein tunnel ([Fig. 6, A–C](#)) and may guide designs of competitive inhibitors mimicking the Ppant-tethered intermediates. Notably, a cluster in the SCBR favors ring structures, including phenols, and roughly mimics the position and orientation of the 2-hydroxyphenyl-thiazolonyl moiety in our docking models for the *R* intermediate pose ([Fig. S22](#)). Clusters near the N terminus of helix α 4 or next to the active-site dyad place polar groups in orientations consistent with positions of polar groups in the intermediate docking results. In contrast, the Ppant-binding region does not return clusters exhibiting specific polar contacts, but probes are found engaged in hydrophobic interactions. Excitingly, three hot spots were identified outside the protein tunnel: a minor site located at the expected donor-binding site ([Fig. 6D](#)), and two more populated sites nestled between the floor loop and α 1 ([Fig. 6, E and F](#)). Both sites seem driven by hydrophobic packing, but the donor site also features polar contacts with S1728 at the predicted Ppant phosphate-binding site. Therefore, in addition to competitive inhibitors mimicking intermediates, alternative classes of inhibitors may be developed to bind PCP sites or the floor loop. The floor loop has been shown to undergo fluctuations in surface electrostatic potential across the conformational variation of the C domain family ([46](#)) despite divergent

Structural and docking insights for the NRPS Cy domain

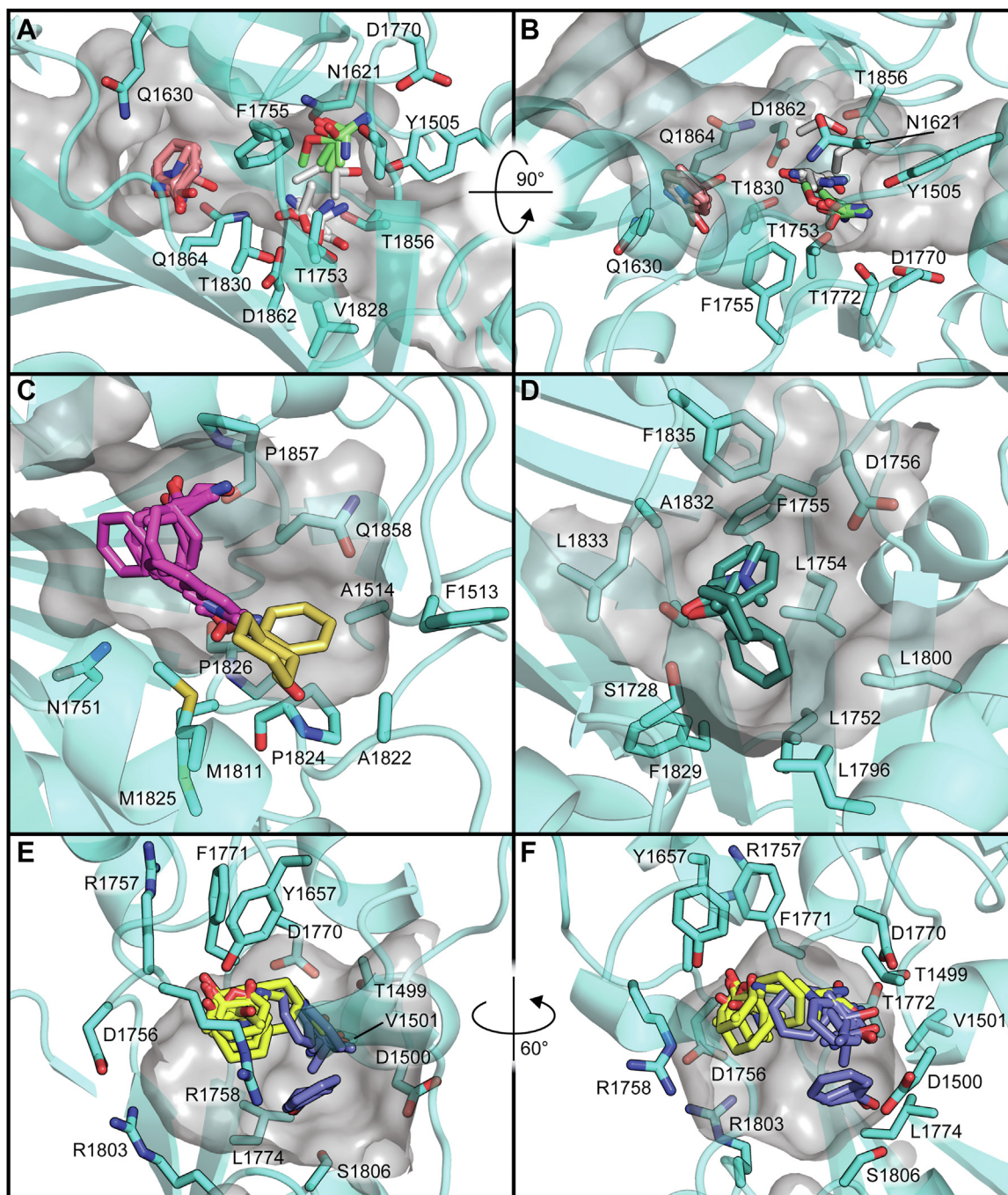


Figure 6. FTmap server results for HMWP2-Cy2. *A* and *B*, two views of small ligand clusters in the HMWP2-Cy2 active site are shown related by a 90° rotation around the *X*-axis. *C*, two clusters of small organic ligands in the downstream tunnel are shown. *D*, binding of a cluster of small organic ligands at the upstream tunnel entrance relies largely on hydrophobic packing, but specific polar contacts could also occur with S1728, which marks the expected phosphopantetheine phosphate-binding site at the end of helix α_8 . *E* and *F*, two clusters of small organic ligands bind in a cavity between the floor loop, hinge, and helices α_1 and α_9 . Although some polar contacts are observed between protein and members of this cluster, docking at this site is probably largely driven by burying hydrophobic surface area.

sequences. Blocking key conformational transitions for a specific Cy domain may permit targeting specific organisms.

Discussion

The crystallographic structures of HMWP2-Cy2 presented here add to the limited array of structural information about Cy domains, and the docking studies and normal mode

analysis performed here suggest how malleability may be related to distinct steps of catalysis. What seems clear is that Cy domains prefer closed upstream entrances in the absence of PCP interactions, so introduction of a PCP may be required to shift the conformational equilibrium toward an open-for-condensation-like state, as observed in cryo-EM studies of PchE (48). In the absence of PCPs, obstruction of the donor site comes in part from a conserved phenylalanine at the C-

terminal end of $\beta 8$ leading into the floor loop (HMWP2 F1755), as seen in crystallographic studies. Mutation of this residue was found to decrease cyclodehydration activity severely (45, 48). Jointly, these observations form the basis for the suggestion that Cy domains exclude bulk solvent from the cyclodehydration active site to facilitate the dehydration step (45). In the presence of a donor PCP, this phenylalanine residue rotates away from the end of helix $\alpha 4$ (48), and it might be that early recognition events between the Cy domain and the PCPs alter the conformational preferences of the phenylalanine and nearby residues. Among these residues, T1753 and T1830 flank the upstream opening and would sense the presence of donor substrates. However, T1830 is also part of the putative catalytic dyad and interacts with the acceptor substrate and D1862. It is likely then that breaking or forming interactions with T1753 and T1830 would influence the global conformational change between open and closed states as well as chemical properties of the catalytic dyad. In some of our docking models, T1753 is nearly able to interact with the sulfur of the acceptor Ppant thioester linkage and is primed to induce such conformational changes. A malleable cyclization scaffold is in agreement with the aforementioned solution NMR studies (46), which revealed global dynamics in the cyclization domain under study, and with cryo-EM data of NRPS modules (47, 48) where cyclization domains display lower resolutions than other domains. The solution NMR studies also indicate that the conserved and putative catalytic aspartate plays a key role in global dynamics that mediate allosteric communication of CP binding (46), which works with our model to suggest how PCP binding, global conformation, and the stage of catalysis may be related.

In comparison to C domains, there are additional catalytic demands for the cyclization reaction (Fig. 7): the geometry of the condensation intermediate must be controlled to overcome ring strain when forming the hydroxythiazolidine intermediate and to allow efficient proton transfer for dehydration. All the Cy domains (44–46, 48) share features of the downstream tunnel that could contribute to binding the Ppant moiety in the reported manner, and these interactions are distinct from those in C–PCP models (51, 52, 58). Structural alignments with existent structures of C–PCP_{acceptor} complexes harboring Ppant molecules demonstrate that Ppant conformations in those models are inconsistent with accommodation of a cyclodehydration intermediate in the Cy active site. This adds some credence to the possibility of separate enzyme conformations for the condensation and cyclodehydration intermediates. Collectively, the Cy domain structures to date, structural comparison with prior PCP_{donor}–C and C–PCP_{acceptor} complexes, and molecular docking work presented here lead us to propose that conserved residues of the downstream tunnels in Cy domains, thought to be involved in Ppant binding, may dynamically contribute to the energy landscape of an open-to-closed transition involving translation of the substrate between condensation- and cyclodehydration-competent conformations. Translation along that path could

be facilitated by rearrangement of T1753 and the Q1858 and N1751 carboxamides and by the ability of Y1505 to act as either hydrogen bond donor or acceptor to the Ppant-linking thioester or the newly formed hydroxythiazolidine's nitrogen. In addition, we show how the distinct structure of Cy domains at the downstream tunnel provides a basin of attraction that could support rearrangement of the Ppant dimethyl moiety during the open-to-closed transition. Perhaps these features of Cy domains could be targeted in inhibitor design, especially considering the identification of ligand-binding hot spots in the downstream tunnel.

In the open-for-condensation state, the putative catalytic dyad would be shifted away from the opposing side of the active site (formed in part by Y1505 in helix $\alpha 1$). This would open additional tunnel volume for binding the cysteine thiol near the dyad in the event of initial condensation by the amine nucleophile. Our studies support this positioning and imply an amine-first condensation step. Envisioning extending the C–S bond formed in the cyclization reaction from its length in our docked cyclodehydration intermediate model to a pre-cyclization state reveals the potential for change in this bond-forming coordinate to be accompanied by a large-scale conformational change between open and closed states approximated in our low-frequency normal mode analysis. The conformational change we describe brings the putative catalytic dyad closer to the center of the active site, potentially positioning it to act as a base while compressing the C–S axis for bond formation (Fig. 7, step 2). Binding of the cyclodehydration intermediate in the manner we report would permit sequestration of the water-leaving group near the N terminus of $\alpha 4$ (which holds water in the crystal structure) and deprotonation of the ring nitrogen by the mostly conserved Y1505.

The results presented here contribute toward deciphering the complex behavior of Cy domains and present possible sites for inhibitor development targeting Cy domain activity in a crucial enzyme of a problem pathogen.

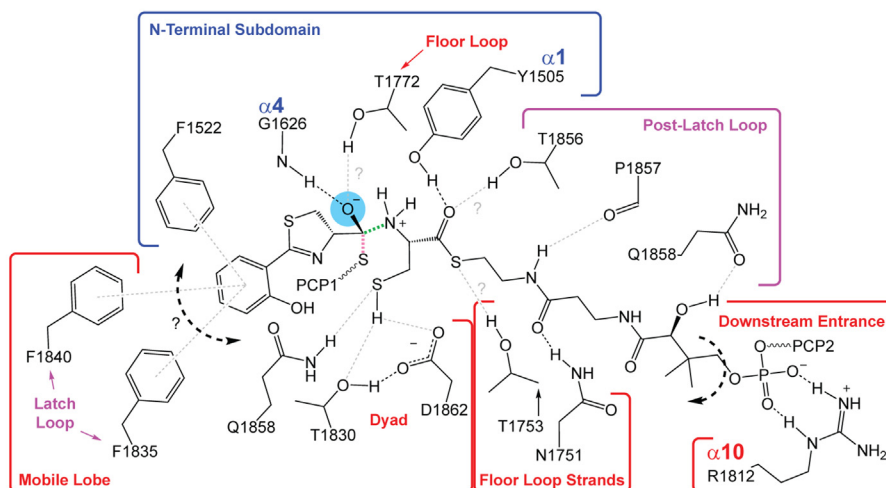
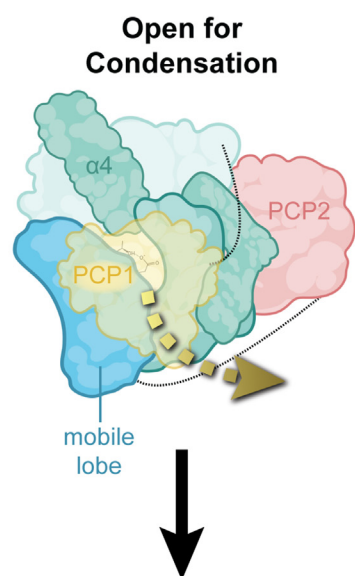
Experimental procedures

Cloning and expression of HMWP2-Cy2

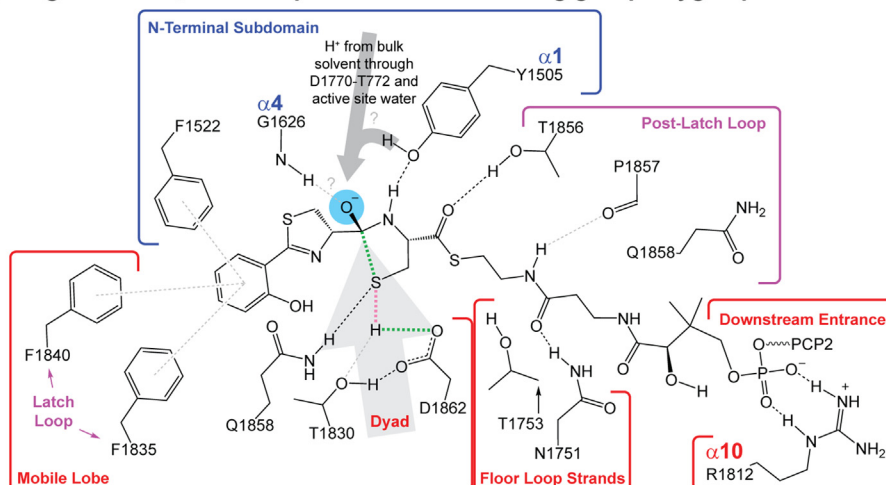
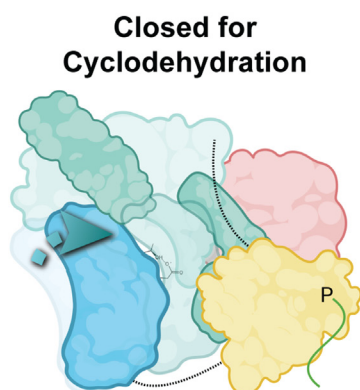
The encoding sequence for HMWP2-Cy2 was amplified from an expression vector containing the entire HMWP2 gene, kindly provided by the C.T. Walsh Laboratory (61). The sequence encoding P1480–Q1910 was amplified using Pfu Ultra II (Agilent) with primers purchased from IDT (forward primer: 5'-GGA AAA CCC A/TA TGC CGG TCG AAC AAC-3'; reverse primer: 5'-CGG/GAT CCT ATT GCC AGG CGC TTT CAT CC-3'; NdeI and BamHI binding sites are underlined, and / marks the cut site). Reactions were purified by electrophoresis and extracted using a Qiagen Gel Extraction kit. Subsequently, both pET28a (Novagen) and the recovered HMWP2-Cy2 DNA insert were double digested with NdeI and BamHI (NEB). The digested pET28a was further reacted with shrimp alkaline phosphatase (NEB) overnight at 16 °C. Digestion reactions were purified using a Qiagen Spin Clear

Structural and docking insights for the NRPS Cy domain

1) Condensation



2) Ring formation, sulfur deprotonation and leaving group oxygen protonation



3) Leaving group protonation 2 and loss of water

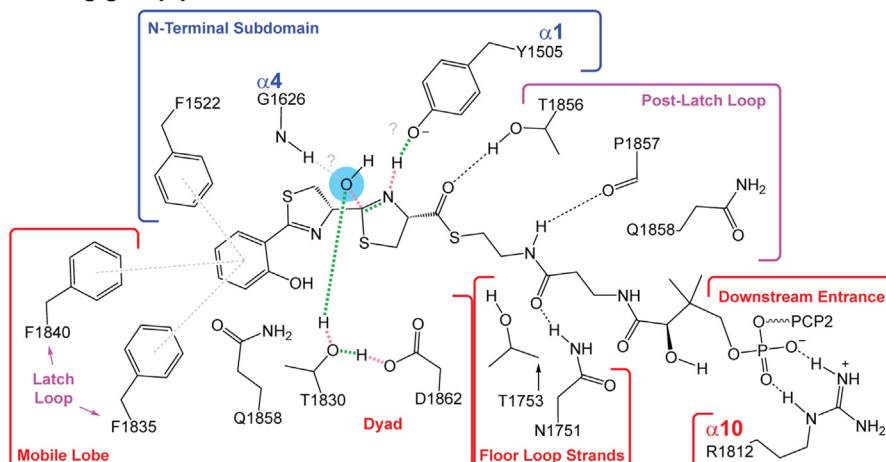


Figure 7. Proposals for a global Cy domain conformational change associated with catalysis and three main reaction steps. On the left is a schematic representation of the conformational changes involved in transitioning from the open-for-condensation state to the closed-for-cyclodehydration state (created using BioRender). The PCP_{donor} (yellow) dissociates from its position in the open-for-condensation state (yellow arrow, Cy domain in green/blue) at least enough to remove the pantetheine moiety from the upstream tunnel entrance. Helix α4 is labeled for reference, and other darker green regions in the open-for-condensation state are the floor loop and hinge regions. To obtain the closed-for-cyclodehydration state, one lobe of the C-terminal subdomain

and Concentration kit, and resultant DNA was used for ligation reactions with T4 DNA Ligase (NEB). Ligated product was transformed directly into chemically competent NEB5 α cells (NEB) and plated on Miller lysogeny broth/agar solid medium supplemented with 50 μ g/ml kanamycin. Transformant DNA was isolated and sequenced (Genewiz) to confirm the presence of the intact HMWP2-Cy2 DNA sequence.

The sequenced HMWP2-Cy2 construct was transformed into BL21(DE3) cells (Stratagene) for protein expression and purification. Eight 1 l lysogeny broth–Miller cultures supplemented with 50 μ g/ml kanamycin were inoculated at 16 °C with 5 ml of an overnight culture and allowed to grow for approximately 72 h. Cells were harvested at 5000 rpm for 15 min in a JLA 9.1000 Beckman rotor at 4 °C. Cell pellets were immediately resuspended in approximately 40 ml of 50 mM Tris (pH 8.0), 100 mM NaCl, 5 mM β -mercaptoethanol, and 5 mM imidazole. The cell suspension was then sonicated using a Branson Sonifier at 50% amplitude and 1 s pulse/2 s off for a total of 2 min for six cycles. The cell lysate was then centrifuged for 1 h at 15,000 rpm in a JA 25.50 rotor at 4 °C. The resulting supernatant, including the soluble HMWP2-Cy2 protein, was further purified at 4 °C.

The HMWP2-Cy2 supernatant was loaded at 1 ml/min onto a 5 ml HisTrap column (GE Healthcare) pre-equilibrated with lysis buffer. Protein was eluted using a 100 ml linear gradient from 0 to 300 mM imidazole. The purest protein-containing fractions as analyzed by SDS-PAGE were collected and dialyzed into 50 mM Tris (pH 8.0) with 5 mM DTT. Dialyzed sample was then loaded onto a 10 ml Mono Q column (GE Healthcare) pre-equilibrated in the same buffer, and protein was eluted over a linear gradient from 0 to 500 mM NaCl. Protein fractions were pooled, concentrated, and run on a 16/600 S200 column (GE Healthcare) pre-equilibrated in 50 mM Tris (pH 8.0), 100 mM NaCl, and 5 mM DTT. Protein-containing fractions were collected and concentrated to 10 mg/ml as determined by the Bradford method with standard bovine serum albumin. Protein was aliquoted, flash frozen in liquid nitrogen, and stored at –80 °C. Analytical gel filtration was performed with an ENrich SEC 650 10 \times 300 column (Bio-Rad), equilibrated and eluted with 50 mM Tris pH 8.0, 100 mM NaCl, and 1 mM DTT at a flow rate of 1 ml/min. Protein standards were from Sigma–Aldrich (ovalbumin, bovine serum albumin, and lysozyme) or MP Biomedicals

(bovine γ -globulins) and run in 50 mM Tris (pH 6.8) and 150 mM NaCl at 1 ml/min.

Structure determination

Crystallization of HMWP2-Cy2

Conditions for crystallizing HMWP2-Cy2 were identified from sparse matrix screening using a Phoenix micropipetting robot (Art Robbins). Prepared trays were stored and monitored at 4 °C within a Rock Imager 1000 (Formulatrix). Drops of 450 nl of 10 mg/ml protein in size-exclusion chromatography buffer and 150 nl of MCSG-1 H10 (Anatrace; 0.1 M HEPES [pH 7.5] and 25% [w/v] PEG 3350) yielded singular crystals of diffraction quality. Crystals were cryoprotected with crystallization solution containing 15% (v/v) glycerol, mounted on a nylon loop, and flash cooled in liquid nitrogen.

Data collection, phasing, and refinement

Diffraction data for the 1.94 Å resolution HMWP2-Cy2 structure were collected at the MIT crystallization facility on a rotating copper anode X-ray generator (Micromax 007) using 1.0° oscillations. Data for the 2.35 Å resolution HMWP2-Cy2 structure were collected at a wavelength of 0.9791 Å at 100 K at the Advanced Photon Source, on beamline 24-ID-E, using 0.5° oscillations. Data were indexed to space group $P4_12_12$ using HKL2000 (62), and data statistics are presented in Table S1.

Phases for the 1.94 Å resolution dataset were solved by molecular replacement in PHASER (63–65) using the structure of EpoB-Cy (PDB ID: 5T7Z) (44), minus residues M1–S42 of the NRPS docking domain, as a search probe. The model was converted to the expected sequence of HMWP2-Cy2 with truncated side chains using CHAINSAW (66) within the CCP4 suite (67), which yielded a PHASER solution with overall LLG and TFZ scores of 128 and 13.2, respectively. Initial electron density maps were interpretable and enabled direct building of the structure. Iterative rounds of model building were carried out in Coot software (68, 69), followed by refinement in PHENIX (64, 65, 70). Initial rounds of refinement included simulated annealing, energy minimization, and *B*-factor refinements. As the model refinement converged, only energy minimization and *B*-factor refinements were performed. After building the protein chain, a continuous region of electron density was observed near R1703 of the V-shaped opening

(blue, mobile lobe), containing T1830 and D1862, rotates through 5 to 10° around its point of contact with the floor loop strands (teal arrow). In addition, a gating phenylalanine (HMWP2-Cy2 F1755) in the floor loop swings toward the N terminus of helix α 4 (not depicted). On the right are proposals for active-site organization during three main reaction steps. Green and pink dashed lines indicate bonds formed or broken in each step, respectively. Light gray dashed lines indicate interactions suggested by our model but for which there is less confidence given variation in our docking models. The first step, condensation, occurs in the open-for-condensation state of the Cy domain. Gray question marks in step 1 point out potential interactions between residues around the active site and the Ppant thioester or oxyanion. In the state depicted in step 1, it appears possible that the donor side chain could form a range of interactions with hydrophobic residues around the SCBR (symbolized by the bold, black, dashed arrow, and question mark on the left). Deprotonation of the zwitterionic intermediate may proceed via the thiolate of the departing donor Ppant (not shown). Following displacement of the donor Ppant and PCP after condensation, an acceptor Ppant conformational change characterized by rotation of the dimethyl group by \sim 180° may be concomitant with the open-to-closed transition (bold, black, and dashed arrow on the right). The second step, in which the condensation product collapses into a cyclization reaction complex, may accompany motion of the C-terminal β strands harboring the dyad (large, light-gray arrow) during the transition from the open-for-condensation state to the closed-for-cyclodehydration state. Motion of the dyad could contribute to compression of the nucleophile–electrophile distance and proton abstraction from the nucleophile. Formation of the cyclodehydration active site during the open-to-closed transition would be influenced by the states of the gating F1755 (omitted for clarity) and floor loop. The final step, including loss of a water, is envisioned as happening entirely in the closed-for-cyclodehydration state. Gray question marks in steps 2 and 3 point out remaining ambiguities. PCP, peptidyl carrier protein; Ppant, phosphopantetheine; SCBR; substrate side chain–binding region.

Structural and docking insights for the NRPS Cy domain

between $\beta 7$ and $\beta 11$, just beyond the donor substrate SCBR. Attempts to refine a chain of water molecules in this density returned positive difference electron density that suggested the presence of a polymer. Therefore, a fragment of a PEG polymer from the crystallization solution and cryoprotectant was modeled and refined with reasonable *B*-factor values and electron density. In addition to this truncated PEG molecule, waters, and a sodium ion were placed and refined into electron density. Final rounds of refinement included occupancy refinement for alternate side-chain confirmations, in addition to translation/libration/screw refinement using optimal group definitions determined within PHENIX. The final model was verified using composite omit maps and Ramachandran angle analysis using MolProbity (71), and no residues were found in disallowed regions. The final model contained all except 24 residues from the N terminus, which included the hexahistidine tag, and residues 1663 to 1668 of the loop connecting the N- and C-terminal halves of the Cy domain structure. Refinement statistics are presented in Table S1.

The 2.35 Å resolution HMWP2-Cy2 structure was solved by molecular replacement using the 1.94 Å resolution structure as a search model in PHASER through PHENIX (63–65). Rounds of model building and refinement were performed identically as aforementioned until the structure refinement converged. The final model was verified using composite omit maps and Ramachandran angle analysis using MolProbity (71) and contained all except 24 residues from the N terminus, which included the hexahistidine tag, and P1666.

Model preparation for molecular docking

Cyclization domain

A complete HMWP2-Cy2 model was generated by combining residues 1483 to 1662 and 1669 to 1910 from the 1.94 Å resolution structure with residues 1663 to 1668 of the 2.35 Å resolution structure (keeping a nonredundant set of ordered water molecules from both structures). The model was prepared with neutral -CHO and -NH₂ termini using standard methods in the BioLuminate Protein Preparation Wizard (72–74): missing atoms were built in, charges were predicted using Epik (Schrödinger) (74, 76), expected bond orders were verified, and hydrogen-bonding networks were optimized. Hydrogen coordinates were refined using the OPLS3e forcefield in the program Impact within the BioLuminate Protein Preparation Wizard (72–74, 77–80). All nonwater ligands were removed from Cy domain models before preparation and simulations. Crystallographic water molecules were removed from the HMWP2-Cy2 models generated for molecular docking after preparation steps (see the [Cyclodehydration intermediates](#) subsection for further information on HMWP2-Cy2-PCP2 models for covalent docking). The HMWP2-Cy2 model used for protein–protein docking was modified by altering the rotameric state of R1812 so that its guanidinium group was toward a location similar to the guanidinium group found interacting with the pantetheine phosphate group in most C–PCP complexes (51, 52, 57). The modified R1812 rotamer was installed in a model

with crystallographic waters, which was then subjected to hydrogen-bond network optimization and hydrogen coordinate refinement before water molecules were removed and R1812 only was subjected to Prime (81–83) energy minimization using the automatic method (up to two iterations of 65 minimization steps using conjugate gradient descent for large gradients and truncated Newton otherwise). HMWP2-Cy2 models for protein–protein docking ultimately included a set containing possible combinations of crystallographic rotameric states of F1513 and R1812 and the C–PCP-mimicking R1812 rotamer.

CP₅

The first model in the deposited NMR bundle for apo HMWP2-PCP1, PDB ID: 5U3H (53), was used for docking experiments. HMWP2 residues 1404 to 1476—comprising the folded core of the apo/holo conformation of PCP1—were used. This model excludes much of the flanking and flexible linker regions. In addition, the conserved serine was phosphorylated to approximate electrostatic contributions from the Ppant phosphate of the holo PCP. Since no structure has been reported for HMWP2-PCP2, a homology model of HMWP2 residues 1943 to 2015 was generated in BioLuminate (74, 84) using as template a cytochrome p450-bound holo PCP crystal structure, PDB ID: 4PWV chain B (16), which shares 34% sequence identity with HMWP2-PCP2. HMWP2-PCP2 was also phosphorylated at its conserved serine residue. PCP models were prepared with neutral -CHO and -NH₂ termini in BioLuminate (73–76, 78) similarly to the preparation of the Cy domain models. An orthorhombic TIP3P water environment with 10 Å buffer regions was built around the HMWP2-PCP2 model using the Desmond System Builder tool in BioLuminate (74, 85, 86). The resulting system also contained 150 mM NaCl plus nine neutralizing Na⁺ ions. This system was relaxed using the default Desmond molecular dynamics (MD) relaxation protocol for the constant pressure (NPT) ensemble (see [Supporting information](#) section), then simulated at 300 K, and 1.01325 bar for 10 ns recording coordinates at 5 ps intervals (85, 86). A representative frame was selected with lowest C α RMSD to the average model of the last 1000 frames of the trajectory (*i.e.*, a set of frames with roughly constant C α RMSD to the input model). This model was then minimized for 100 ps using the Desmond MD energy minimization protocol. The protein model was extracted from the energy-minimized system and reprepared in the same fashion as the HMWP2-PCP1 model.

Cyclodehydration intermediates

BioLuminate was used to model potential phosphopantetheinylated HMWP2-Cy2 cyclodehydration intermediates of *R* and *S* chiralities at the hydroxyl-bearing carbon center of the hydroxythiazolidine ((2*R/S,4R*)-2,2-hydroxy-(2'-(2"-hydroxyphenyl)-thiazolinyl)-thiazolidinyl Ppant) (74). Note that the thiazolines of the 2-hydroxyphenyl-thiazolinyl side chains of these intermediates contain the chiral carbon corresponding to the *D* product of the *L/D* epimerization carried out by the

upstream embedded E domain of HMWP2 (29). The cyclodehydration intermediate models were prepared as singly charged anions (at the phosphate) using LigPrep in BioLuminate (74, 87). These molecules will be referred to as Ppant-2HPTT(*R/S*)-OH or *R* or *S* (cyclodehydration) intermediates.

Molecular docking

Protein–protein docking

The prepared HMWP2–PCP1 model was docked at the predicted opening of the upstream HMWP2–Cy2 active-site tunnel (between strands β 8 and β 10 of the prepared model with all water molecules removed and the modified R1812 rotamer) using default settings in the HADDOCK2.4 web server (54–56, 88) with an additional brief MD simulation in explicit solvent. Active residues providing restraints were defined as the PCP phosphoserine (HMWP2 residue SEP1439) and Cy residue S1728, near the N terminus of helix α 8, where the Ppant phosphate is predicted to be in PCP_{donor}–Cy complexes (based on the PCP1_{donor}–C2 complex of LgrA; PDB ID: 6MFX) (51). The HMWP2–PCP2 homology model was docked similarly at the opening of the downstream active-site tunnel of HMWP2–Cy2 (between helices α 1 and α 10), using Cy models with each possible combination of F1513 and R1812 crystallographic rotamers and the C–PCP-mimicking R1812 rotamer. For the results presented here, active residues were Cy F1513 and R1812 and PCP F1971, SEP1977, F2001, and the Cy model contained the F1513 rotamer B, directing the phenyl toward the PCP_{acceptor} binding site, and the C–PCP-mimicking R1812 rotamer. The additional active PCP2 residues F1971 and F2001 were included to promote favorable π -stacking interactions and/or hydrophobic packing with F1513 of the Cy domain, which adopts two rotameric states in the HMWP2–Cy2 crystal structure (one flipped out toward the PCP2-binding site and one away) and is present in over 40% of selected Cy domain sequences aligned in Fig. S7. Runs were also conducted with only R1812 and SEP1977 defined as active, but these and the other rotameric combinations of F1513 and R1812 did not perform as well, in some cases returning poses that resemble PDB ID: 4ZXI or PDB ID: 6N8E but that have considerably less favorable HADDOCK scores, lower buried surface areas, and/or positive Z scores. These results are therefore not presented. In selecting docked poses, scores, energy terms, and cluster sizes output by HADDOCK were considered, and the models were visually inspected for poses that were consistent with the HMWP2 modular architecture and reasonably placed the conserved PCP serine residues (SEP1439 or SEP1977) near the upstream and downstream active-site tunnel entrances.

Cyclodehydration intermediates

Because reasonable HMWP2–Cy2–PCP2 didomain complex models were obtained by protein–protein docking in HADDOCK2.4, the best-ranked of them was chosen as the basis for an approximate complex of the HMWP2–Cy2 crystal structure and the docked PCP2, to provide a Ppant phosphate

site. The complex was formed by superimposing the docked HMWP2–Cy2–PCP2 complex on the HMWP2–Cy2 crystal structure using C α alignment of Cy helices α 1 and α 10 (the secondary structure features forming most of the downstream tunnel entrance) followed by removal of the Cy domain from docking as well as redundant and conflicting interfacial water molecules and crystallographic water molecules in the tunnel that would block substrate access. The rotameric state of R1812 of the HMWP2–Cy2 structure was altered so its guanidinium moiety is positioned similarly to guanidinium moieties of conserved arginine residues observed in C–PCP_{acceptor} complexes. The resulting complex was then preprocessed and subjected to hydrogen-bond network optimization and refinement of PCP2 and the interface waters and residues within 5 Å using the automatic method in Prime (81–83). From this model, a second was generated that contained an aspartic acid rather than aspartate at the putative catalytic dyad. Both models (*i.e.*, anionic and neutral dyads) were then subjected to hydrogen coordinate refinement followed by all-atom energy minimization in Impact with a 0.3 Å heavy atom displacement cutoff preventing the models from deviating too much from their input coordinates.

Phosphoryl groups were then removed from the conserved PCP serine residues in preparation for covalent docking using Schrödinger's Phosphonate_Addition.cdock reaction definition file available at <https://www.schrodinger.com/products/covdock> (also see the Supporting information section). This reaction definition file, although designed for phosphonate reactions, is suitable for displacement of a phosphate oxygen by a serine side-chain oxygen. The covalent docking tool in BioLuminate (74) uses Glide docking (89–91) in a rigid receptor approximation and returns only poses that are compatible with the covalent linkage the user defines (92). Default covalent docking settings were used (maximum of 200 initial poses, keeping only those with Glide scores lower than +2.5, outputting the top 1000 with 30 poses per ligand reaction site, which is one site in this case). The docking grid's inner box, within which the ligand's center must be positioned, was 10 × 10 × 10 Å³ and the outer box had dimensions of 34 × 34 × 34 Å³. The center of the grid was defined as the centroid of tunnel-lining residues so that it was in the center of the tunnel's diameter and positioned roughly halfway between R1812 and F1522, which lie at the ends of the docking region. Docking results were manually inspected to identify categories of hydroxythiazolidine orientation. Each category of pose for the *R* and *S* intermediates was subjected to optimization/refinement in the following scheme inspired by the refinement steps of Glide/Prime Induced Fit Docking protocols in BioLuminate (74, 81–83, 89–94): (1) refinement of protein residues within 5 Å of the docked ligand using 10 iterations of 100 steps each with Prime's automatic conjugate gradient/truncated Newton method, (2) refinement of the ligand and residues within 5 Å of it using 10 iterations of 100 steps each with Prime's automatic method, (3) Monte Carlo protein side-chain optimization for residues within 5 Å of the ligand (Predict Side Chains tool running Prime in BioLuminate), and (4) refinement of the ligand and residues within 5 Å of it using

Structural and docking insights for the NRPS Cy domain

10 iterations of 100 steps each with Prime's automatic method. These refined models were sorted by Prime energy and evaluated using a combination of that value and the docking score. As a final point of comparison between poses, the phosphate-serine linkages of top covalent docking models were cleaved by deletion of the bridging oxygen (S1977 O γ), replacing the broken bonds with bonds to hydrogens on the separated groups to leave a phosphonate and alanine, and molecular mechanics with generalized Born and surface area minimization calculations were performed on the resulting noncovalent complexes using Prime with 1 kcal/mol·Å² restraints on flexible atoms, which were defined as those of the ligand and residues within 10 Å of it (81–83).

Low-frequency vibrational normal mode analysis

The 1.94 Å resolution structure of HMWP2-Cy2 (PDB ID: 7JTJ) was retrieved from the PDB, and the missing loop and side chains were filled using Prime in BioLuminate (74, 81–83). Using the BioLuminate Protein Preparation Wizard (72–80), hydrogen-bond networks were optimized, hydrogen coordinates were refined to minimize potential energy, and all atom coordinates were refined to decrease potential energy with a cutoff before heavy atom RMSD reached 0.3 Å. The PEG fragment and loosely bound surface water molecules were removed. The Desmond (85, 86) solvate pocket tool was used to add water molecules to the active-site tunnel, where the PEG fragment was removed. This model was placed in a solvent box with at least 10 Å separating the protein and the sides of the box, neutralizing with sodium ions and including 150 mM NaCl. The crystallographic sodium ion was replaced with a water molecule in this process. The default Desmond MD (85, 86) relaxation protocol was run, and the output from the last stage with solute restraints (stage 5) was used as input to a series of 100 ps MD runs in the NPT ensemble at 277 K and 1.01325 bar, gradually releasing solute restraints as follows: 10 kcal/mol·Å², 5 kcal/mol·Å², and then 1 kcal/mol·Å². These were followed with a series of 100 ps MD warming steps in 5 K increments from 280 to 300 K, each in the NPT ensemble at 1.01325 bar. An additional 5.5 ns of NPT simulation at 300 K were performed recording frames every 1 ps. A representative frame was selected from the final 2200 frames as the frame with the lowest RMSD to the average structure over this range. Surface water and ions were removed by selecting molecules more than 5 Å from the protein, then extending that selection to water or ions within 5 Å, leaving only the water most closely associated with the protein and the active-site tunnel water undisturbed. An initial energy minimization in MacroModel (Schrödinger) (95) using the Polak–Ribiere Conjugate Gradient method with a cutoff of 0.0418 kJ/mol was performed and followed by normal mode calculation in BioLuminate (74), returning 80 frames for each of the top 10 modes. This allowed for identification of loosely bound water molecules on the surface of the input model that contributed modes with negative eigenvalues. These water molecules were removed from the model used as input to the initial minimization, and the resulting model was subjected to Polak–Ribiere

Conjugate Gradient energy minimization as before. Normal modes reported here were calculated for the minimized model without loosely bound surface water molecules.

FTMap ligand fragment–binding site identification

The composite model built using the 1.94 Å and 2.35 Å resolution HMWP2-Cy2 structures was used as input to the FTMap server (60). The entire protein was defined as the search space, and default settings were used for binding site identification.

Figure preparation

Chemical figures were generated in ChemDraw (PerkinElmer), protein figures were generated using PyMOL (Schrödinger), and electrostatics figures were produced using the APBS PyMOL plugin with charges from the OPLS3e force field available in BioLuminate and van der Waals radii (74, 78, 96). Calculations of approximate tunnel volumes were performed using the CASTp 3.0 web server available at the time of publication at <http://sts.bioe.uic.edu/castp/calculation.html> (97). Sequence alignments were generated using Clustal Omega (98). Representative Cy domain sequences were retrieved from UniProtKB by SANSparallel (99, 100), searching with seed Cy domain sequences from anguibactin, bacillamide, bacitracin, bleomycin, epothilone, JBIR-34/45, mycobactin, myxothiazole, pyochelin, vibriobactin, and Ybt biosynthesis. Sequence redundancy was cut at 80% in JalView (101) before alignment. WebLogo3 (102) was used to generate the sequence logo.

Data availability

The 1.94 Å and 2.35 Å resolution crystal structures for HMWP2-Cy2 have been deposited to the PDB (<http://www.rcsb.org/>) and assigned the accession numbers 7JTJ and 7JUA, respectively.

All files relating to the computational studies reported here are available in the GitHub repository at <https://github.com/adgnann/HMWP2-Cy2>. The file README.md provides a description of the repository contents.

Supporting information—This article contains supporting information including supplemental results and discussion as well as crystallographic data collection and refinement statistics (Table S1) and additional figures (Figs. S1–S22) (44, 45, 48, 51, 52, 57–59, 99–105). Videos of the low-frequency vibrational normal modes are also available online ([Supporting movies](#)).

Acknowledgments—We kindly acknowledge Catherine L. Drennan for support in starting this project, the MIT X-ray facility and Staff Scientist Robert Grant for continued usage of the facility, and Subrata Mishra, Kenny Marincin, and Aswani Kancherla for enlightening discussions. We also acknowledge use of the high-performance computing facilities managed by the Research Computing Department at the University of Massachusetts Boston. This work is based upon research conducted at the Northeastern Collaborative Access Team beamlines, which are funded by the National Institute of General Medical Sciences,

National Institutes of Health (grant no.: P30 GM124165). The Eiger 16M detector on 24-ID-E beamline is funded by a National Institutes of Health-ORIP HEI grant (grant no.: S10OD021527). This research used resources of the Advanced Photon Source, a US Department of Energy Office of Science User Facility operated for the Department of Energy Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357.

Author contributions—A. D. G., Y. X., E. J. B., and D. P. D. conceptualization; A. D. G., Y. X., J. S., C. B., J. S. M., S. N. M., B. M. C., E. J. B., and D. P. D. methodology; A. D. G., Y. X., J. S., C. B., J. S. M., S. N. M., B. M. C., E. J. B., and D. P. D. investigation; A. D. G. and D. P. D. writing—original draft; D. P. F. and D. P. D. writing—review & editing; D. P. D. resources.

Funding and additional information—Support for this research was provided to D. P. D. by the National Institute of General Medical Sciences at the National Institutes of Health (grant no.: 1R15GM123425-01). A. D. G. was supported by a Sanofi Genzyme doctoral research fellowship. B. M. C. was supported by an HHMI EXROP and EXROP Capstone Scholarship. D. P. F. acknowledges support from the National Institute of General Medical Sciences (grant no.: R01 GM104257). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: A, adenylation domain; ArCP, aryl carrier protein; C, condensation domain; CP, carrier protein; Cy, heterocyclization domain; 2HPTT, 2-hydroxyphenylthiazolylthiazolidine; MD, molecular dynamics; NRP, nonribosomal peptide; NRPS, nonribosomal peptide synthetase; PCP, peptidyl carrier protein; PDB, Protein Data Bank; PK, polyketide; PKS, polyketide synthase; Ppant, phosphopantetheine; SCBR, substrate side chain-binding region; Ybt, yersiniabactin.

References

- Wang, H., Fewer, D. P., Holm, L., Rouhiainen, L., and Sivonen, K. (2014) Atlas of nonribosomal peptide and polyketide biosynthetic pathways reveals common occurrence of nonmodular enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 9259–9264
- Crits-Christoph, A., Diamond, S., Butterfield, C. N., Thomas, B. C., and Banfield, J. F. (2018) Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**, 440–444
- Hider, R. C., and Kong, X. L. (2010) Chemistry and biology of siderophores. *Nat. Prod. Rep.* **27**, 637–657
- Zhang, J. J., Tang, X. Y., Huan, T., Ross, A. C., and Moore, B. S. (2020) Pass-back chain extension expands multimodular assembly line biosynthesis. *Nat. Chem. Biol.* **16**, 42–49
- Cañedo, L. M., de la Fuente, J. A., Gesto, C., Ferreira, M. J., Jiménez, C., and Riguera, R. (1999) Agrochelin, a new cytotoxic alkaloid from the marine bacteria *Agrobacterium* sp. *Tetrahedron Lett.* **40**, 6841–6844
- Kobayashi, S., Nakai, H., Ikenishi, Y., Sun, W. Y., Ozaki, M., Hayase, Y., et al. (1998) Micacocidin A, B and C, novel antimycoplasma agents from *Pseudomonas* sp. II. Structure elucidation. *J. Antibiot. (Tokyo)* **51**, 328–332
- Bollag, D. M., McQueney, P. A., Zhu, J., Hensens, O., Koupal, L., Liesch, J., et al. (1995) Epothilones, a new class of microtubule-stabilizing agents with a taxol-like mechanism of action. *Cancer Res.* **55**, 2325–2333
- Johnson, B. A., Anker, H., and Meloney, F. L. (1945) Bacitracin: a new antibiotic produced by a member of the *B. subtilis* group. *Science* **102**, 376–377
- Beck, C., Garzón, J. F. G., and Weber, T. (2020) Recent advances in re-engineering modular PKS and NRPS assembly lines. *Biotechnol. Bioproc. Eng.* **25**, 886–894
- Carniel, E. (2001) The *Yersinia* high-pathogenicity island: an iron-uptake island. *Microbes Infect.* **3**, 561–569
- Quadri, L. E. N., Keating, T. A., Patel, H. M., and Walsh, C. T. (1999) Assembly of the *Pseudomonas aeruginosa* nonribosomal peptide siderophore pyochelin: *in vitro* reconstitution of aryl-4, 2-bisthiazoline synthetase activity from PchD, PchE, and PchF. *Biochemistry* **38**, 14941–14954
- Griffiths, G. L., Sigel, S. P., Payne, S. M., and Neilands, J. B. (1984) Vibriobactin, a siderophore from *Vibrio cholerae*. *J. Biol. Chem.* **259**, 383–385
- Quadri, L. E. N., Sello, J., Keating, T. A., Weinreb, P. H., and Walsh, C. T. (1998) Identification of a *Mycobacterium tuberculosis* gene cluster encoding the biosynthetic enzymes for assembly of the virulence-conferring siderophore mycobactin. *Chem. Biol.* **5**, 631–645
- Pollack, J. R., and Neilands, J. B. (1970) Enterobactin, an iron transport compound from *Salmonella typhimurium*. *Biochem. Biophys. Res. Commun.* **38**, 989–992
- Foley, T. L., and Simeonov, A. (2012) Targeting iron assimilation to develop new antibacterials. *Expert Opin. Drug Discov.* **7**, 831–847
- Haslinger, K., Brieke, C., Uhlmann, S., Sieverling, L., Süsmuth, R. D., and Cryle, M. J. (2014) The structure of a transient complex of a non-ribosomal peptide synthetase and a cytochrome P450 monooxygenase. *Angew. Chem. Int. Ed. Engl.* **53**, 8518–8522
- Du, L., Chen, M., Sánchez, C., and Shen, B. (2000) An oxidation domain in the BlmIII non-ribosomal peptide synthetase probably catalyzing thiazole formation in the biosynthesis of the anti-tumor drug bleomycin in *Streptomyces verticillus* ATCC15003. *FEMS Microbiol. Lett.* **189**, 171–175
- Schneider, T. L., Shen, B., and Walsh, C. T. (2003) Oxidase domains in epothilone and bleomycin biosynthesis: thiazoline to thiazole oxidation during chain elongation. *Biochemistry* **42**, 9722–9730
- Schneider, T. L., and Walsh, C. T. (2004) Portability of oxidase domains in nonribosomal peptide synthetase modules. *Biochemistry* **43**, 15946–15955
- Ansari, M. Z., Sharma, J., Gokhale, R. S., and Mohanty, D. (2008) In silico analysis of methyltransferase domains involved in biosynthesis of secondary metabolites. *BMC Bioinform.* **9**, 454
- Miller, D. A., Walsh, C. T., and Luo, L. (2001) C-methyltransferase and cyclization domain activity at the intraprotein PK/NRP switch point of yersiniabactin synthetase. *J. Am. Chem. Soc.* **123**, 8434–8435
- Mori, S., Garneau-Tsodikova, S., and Tsodikov, O. V. (2020) Unimodular methylation by adenylation-thiolation domains containing an embedded methyltransferase. *J. Mol. Biol.* **432**, 5802–5808
- Patel, H. M., and Walsh, C. T. (2001) *In vitro* reconstitution of the *Pseudomonas aeruginosa* nonribosomal peptide synthesis of pyochelin: characterization of backbone tailoring thiazoline reductase and N-methyltransferase activities. *Biochemistry* **40**, 9023–9031
- Ronnebaum, T. A., McFarlane, J. S., Prinsinzano, T. E., Booker, S. J., and Lamb, A. L. (2019) Stuffed methyltransferase catalyzes the penultimate step of pyochelin biosynthesis. *Biochemistry* **58**, 665–678
- Xu, F., Butler, R., May, K., Rexhepaj, M., Yu, D., Zi, J., et al. (2019) Modified substrate specificity of a methyltransferase domain by protein insertion into an adenylation domain of the bassianolide synthetase. *J. Biol. Eng.* **13**, 65
- Howard-Jones, A. R., Kruger, R. G., Lu, W., Tao, J., Leimkuhler, C., Kahne, D., et al. (2007) Kinetic analysis of teicoplanin glycosyltransferases and acyltransferase reveal ordered tailoring of aglycone scaffold to reconstitute mature teicoplanin. *J. Am. Chem. Soc.* **129**, 10082–10083
- Cadel-Six, S., Dauga, C., Castets, A. M., Rippka, R., Bouchier, C., Tandeau de Marsac, N., et al. (2008) Halogenase genes in nonribosomal peptide synthetase gene clusters of *Microcystis* (cyanobacteria): sporadic distribution and evolution. *Mol. Biol. Evol.* **25**, 2031–2041
- Balibar, C. J., Vaillancourt, F. H., and Walsh, C. T. (2005) Generation of D amino acid residues in assembly of arthrofactin by dual condensation/epimerization domains. *Chem. Biol.* **12**, 1189–1200

Structural and docking insights for the NRPS Cy domain

29. Patel, H. M., Tao, J., and Walsh, C. T. (2003) Epimerization of an L-cysteinylyl to a D-cysteinylyl residue during thiazoline ring formation in siderophore chain elongation by pyochelin synthetase from *Pseudomonas aeruginosa*. *Biochemistry* **42**, 10514–10527
30. Bloudoff, K., and Schmeing, T. M. (2017) Structural and functional aspects of the nonribosomal peptide synthetase condensation domain superfamily: discovery, dissection and diversity. *Biochim. Biophys. Acta Proteins Proteom.* **1865**, 1587–1604
31. Drechsel, H., Stephan, H., Lotz, R., Haag, H., Zähner, H., Hantke, K., *et al.* (1995) Structure elucidation of yersiniabactin, a siderophore from highly virulent *Yersinia* strains. *Liebigs Annalen* **1995**, 1727–1733
32. Perry, R. D., Balbo, P. B., Jones, H. A., Fetherston, J. D., and DeMoll, E. (1999) Yersiniabactin from *Yersinia pestis*: biochemical characterization of the siderophore and its role in iron transport and regulation. *Microbiology* **145**, 1181–1190
33. Miller, M. C., Parkin, S., Fetherston, J. D., Perry, R. D., and DeMoll, E. (2006) Crystal structure of ferric-yersiniabactin, a virulence factor of *Yersinia pestis*. *J. Inorg. Biochem.* **100**, 1495–1500
34. Perry, R. D., and Fetherston, J. D. (2011) Yersiniabactin iron uptake: mechanisms and role in *Yersinia pestis* pathogenesis. *Microbes Infect.* **13**, 808–817
35. Heesemann, J., Hantke, K., Vocke, T., Saken, E., Rakin, A., Stojiljkovic, I., *et al.* (1993) Virulence of *Yersinia enterocolitica* is closely associated with siderophore production, expression of an iron-repressible outer membrane polypeptide of 65,000 Da and pesticin sensitivity. *Mol. Microbiol.* **8**, 397–408
36. De Voss, J. J., Rutter, K., Schroeder, B. G., and Barry, C. E., 3rd. (1999) Iron acquisition and metabolism by mycobacteria. *J. Bacteriol.* **181**, 4443–4451
37. Actis, L. A., Fish, W., Crosa, J. H., Kellerman, K., Ellenberger, S. R., Hauser, F. M., *et al.* (1986) Characterization of anguibactin, a novel siderophore from *Vibrio anguillarum* 775(pJM1). *J. Bacteriol.* **167**, 57–65
38. Brumbaugh, A. R., Smith, S. N., Subashchandrabose, S., Himpfl, S. D., Hazen, T. H., Rasko, D. A., *et al.* (2015) Blocking yersiniabactin import attenuates extraintestinal pathogenic *Escherichia coli* in cystitis and pyelonephritis and represents a novel target to prevent urinary tract infection. *Infect. Immun.* **83**, 1443–1450
39. Tao, P., Mahalingam, M., Kirtley, M. L., van Lier, C. J., Sha, J., Yeager, L. A., *et al.* (2013) Mutated and bacteriophage T4 nanoparticle arrayed F1-V immunogens from *Yersinia pestis* as next generation plague vaccines. *PLoS Pathog.* **9**, e1003495
40. Klebba, P. E., Newton, S. M. C., Six, D. A., Kumar, A., Yang, T., Nairn, B. L., *et al.* (2021) Iron acquisition systems of gram-negative bacterial pathogens define TonB-dependent pathways to novel antibiotics. *Chem. Rev.* **121**, 5193–5239
41. Ellermann, M., and Arthur, J. C. (2017) Siderophore-mediated iron acquisition and modulation of host-bacterial interactions. *Free Radic. Biol. Med.* **105**, 68–78
42. Carniel, E., Guilvout, I., and Prentice, M. (1996) Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. *J. Bacteriol.* **178**, 6743–6751
43. Smith, H. G., Beech, M. J., Lewandowski, J. R., Challis, G. L., and Jenner, M. (2021) Docking domain-mediated subunit interactions in natural product megasynth(et)ases. *J. Ind. Microbiol. Biotechnol.* **48**, kuab018
44. Dowling, D. P., Kung, Y., Croft, A. K., Taghizadeh, K., Kelly, W. L., Walsh, C. T., *et al.* (2016) Structural elements of an NRPS cyclization domain and its intermodule docking domain. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 12432–12437
45. Bloudoff, K., Fage, C. D., Marahiel, M. A., and Schmeing, T. M. (2017) Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 95–100
46. Mishra, S. H., Kancharla, A. K., Marincin, K. A., Bouvignies, G., Nerli, S., Sgourakis, N., *et al.* (2022) Global protein dynamics as communication sensors in peptide synthetase domains. *Sci. Adv.* **8**, eabn6549
47. Katsuyama, Y., Sone, K., Harada, A., Kawai, S., Urano, N., Adachi, N., *et al.* (2021) Structural and functional analyses of the tridomain-nonribosomal peptide synthetase FmoA3 for 4-methylxazoline ring formation. *Angew. Chem. Int. Ed. Engl.* **60**, 14554–14562
48. Wang, J., Li, D., Chen, L., Cao, W., Kong, L., Zhang, W., *et al.* (2022) Catalytic trajectory of a dimeric nonribosomal peptide synthetase subunit with an inserted epimerase domain. *Nat. Commun.* **13**, 592
49. Kosol, S., Gallo, A., Griffiths, D., Valentic, T. R., Masschelein, J., Jenner, M., *et al.* (2019) Structural basis for chain release from the enacyloxin polyketide synthase. *Nat. Chem.* **11**, 913–923
50. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797
51. Reimer, J. M., Eivaskhani, M., Harb, I., Guarné, A., Weigt, M., and Schmeing, T. M. (2019) Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. *Science* **366**, eaaw4388
52. Drake, E. J., Miller, B. R., Shi, C., Tarrasch, J. T., Sundlov, J. A., Allen, C. L., *et al.* (2016) Structures of two distinct conformations of holo-nonribosomal peptide synthetases. *Nature* **529**, 235–238
53. Harden, B. J., and Frueh, D. P. (2017) Molecular cross-talk between nonribosomal peptide synthetase carrier proteins and unstructured linker regions. *Chembiochem* **18**, 629–632
54. Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003) Haddock: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737
55. de Vries, S. J., van Dijk, A. D. J., Krzeminski, M., van Dijk, M., Thureau, A., Hsu, V., *et al.* (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* **69**, 726–733
56. van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastriitis, P. L., Karaca, E., *et al.* (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725
57. Izoré, T., Candace Ho, Y. T., Kaczmarek, J. A., Gavrilidou, A., Chow, K. H., Steer, D. L., *et al.* (2021) Structures of a non-ribosomal peptide synthetase condensation domain suggest the basis of substrate selectivity. *Nat. Commun.* **12**, 2511
58. Kreidler, D. F., Gemmel, E. M., Schaffer, J. E., Wenczewicz, T. A., and Gulick, A. M. (2019) The structural basis of N-acyl-alpha-amino-beta-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat. Commun.* **10**, 3432
59. Bloudoff, K., Alonzo, D. A., and Schmeing, T. M. (2016) Chemical probes allow structural insight into the condensation reaction of non-ribosomal peptide synthetases. *Cell Chem. Biol.* **23**, 331–339
60. Kozakov, D., Grove, L. E., Hall, D. R., Bohnuud, T., Mottarella, S. E., Luo, L., *et al.* (2015) The FTMap family of web servers for determining and characterizing ligand-binding hot spots of proteins. *Nat. Protoc.* **10**, 733–755
61. Keating, T. A., Miller, D. A., and Walsh, C. T. (2000) Expression, purification, and characterization of HMWP2, a 229 kDa, six domain protein subunit of Yersiniabactin synthetase. *Biochemistry* **39**, 4729–4739
62. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Met. Enzymol.* **276**, 307–326
63. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674
64. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., *et al.* (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221
65. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in phenix. *Acta Crystallogr. D Struct. Biol.* **75**, 861–877
66. Stein, N. (2008) CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643
67. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, 235–242

68. Emsley, P., and Cowtan, K. (2004) *Coot*: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132
69. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of *coot*. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501
70. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367
71. Chen, V. B., Arendall, W. B., 3rd, Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., *et al.* (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21
72. Schrödinger Release 2020-4: *Protein Preparation Wizard*. (2020). Epik, Schrödinger, LLC, New York, NY
73. Sastry, G. M., Adzhigirey, M., Day, T., Annabhimoju, R., and Sherman, W. (2013) Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J. Comput. Aided Mol. Des.* **27**, 221–234
74. Schrödinger Release 2020-4. (2020). BioLuminate, Schrödinger, LLC, New York, NY
75. Greenwood, J. R., Calkins, D., Sullivan, A. P., and Shelley, J. C. (2010) Towards the comprehensive, rapid, and accurate prediction of the favorable tautomeric states of drug-like molecules in aqueous solution. *J. Comput. Aided Mol. Des.* **24**, 591–604
76. Shelley, J. C., Cholleti, A., Frye, L. L., Greenwood, J. R., Timlin, M. R., and Uchimaya, M. (2007) Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. *J. Comput. Aided Mol. Des.* **21**, 681–691
77. Jorgensen, W. L., and Tirado-Rives, J. (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.* **110**, 1657–1666
78. Harder, E., Damm, W., Maple, J., Wu, C., Reboul, M., Xiang, J. Y., *et al.* (2016) OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. *J. Chem. Theor. Comput.* **12**, 281–296
79. Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118**, 11225–11236
80. Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., and Sherman, W. (2010) Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. *J. Chem. Theor. Comput.* **6**, 1509–1519
81. Schrödinger Release 2020-4: *Prime*. (2020). Schrödinger, LLC, New York, NY
82. Jacobson, M. P., Friesner, R. A., Xiang, Z., and Honig, B. (2002) On the role of the crystal environment in determining protein side-chain conformations. *J. Mol. Biol.* **320**, 597–608
83. Jacobson, M. P., Pincus, D. L., Rapp, C. S., Day, T. J. F., Honig, B., Shaw, D. E., *et al.* (2004) A hierarchical approach to all-atom protein loop prediction. *Proteins* **55**, 351–367
84. Zhu, K., Day, T., Warshaviak, D., Murrett, C., Friesner, R., and Pearlman, D. (2014) Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction. *Proteins* **82**, 1646–1655
85. Schrödinger Release 2020-4: *Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY*. (2020). Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY
86. Bowers, K. J., Chow, E., Xu, H., Dror, R. O., Eastwood, M. P., Gregersen, B. A., *et al.* (2006) Scalable algorithms for molecular dynamics simulations on commodity clusters. In *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing*, IEEE Computer Society, Los Alamitos, CA: 84
87. Schrödinger Release 2020-4. (2020). LigPrep, Schrödinger, LLC, New York, NY
88. Honorato, R. V., Koukos, P. I., Jimenez-Garcia, B., Tsaregorodtsev, A., Verlati, M., Giachetti, A., *et al.* (2021) Structural biology in the clouds: the WeNMR-EOSC ecosystem. *Front. Mol. Biosci.* **8**, 729513
89. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *J. Med. Chem.* **47**, 1739–1749
90. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.* **47**, 1750–1759
91. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., *et al.* (2006) Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **49**, 6177–6196
92. Zhu, K., Borrelli, K. W., Greenwood, J. R., Day, T., Abel, R., Farid, R. S., *et al.* (2014) Docking covalent inhibitors: a parameter free approach to pose prediction and scoring. *J. Chem. Inf. Model.* **54**, 1932–1940
93. Sherman, W., Day, T., Jacobson, M. P., Friesner, R. A., and Farid, R. (2006) Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem.* **49**, 534–553
94. Schrödinger Release 2020-4: *Glide*. (2020). Schrödinger, LLC, New York, NY
95. Schrödinger Release 2020-4. (2020). MacroModel, Schrödinger, LLC, New York, NY
96. Jurrus, E., Engel, D., Star, K., Monson, K., Brandi, J., Felberg, L. E., *et al.* (2018) Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128
97. Tian, W., Chen, C., Lei, X., Zhao, J., and Liang, J. (2018) CASTp 3.0: computed atlas of surface topography of proteins. *Nucl. Acids Res.* **46**, W363–W367
98. Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucl. Acids Res.* **47**, W636–W641
99. Koskinen, J. P., and Holm, L. (2012) SANS: high-throughput retrieval of protein sequences allowing 50% mismatches. *Bioinformatics* **28**, i438–i443
100. Somervuo, P., and Holm, L. (2015) SANSparallel: interactive homology search against uniprot. *Nucl. Acids Res.* **43**, W24–29
101. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191
102. Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004) WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190
103. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins* **23**, 566–579
104. Stevers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539
105. Williams, C. J., Headd, J. J., Moriarty, N. W., Prisant, M. G., Videau, L. L., Deis, L. N., *et al.* (2018) MolProbity: more and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315