Title: Structural elements of an NRPS cyclization domain and its intermodule docking domain

Authors:
Daniel P. Dowling[1,2,3], Yan Kung[4], Anna K. Croft[5], Koli Taghizadeh[6], Wendy L. Kelly[7], Christopher T. Walsh[8], and Catherine L. Drennan[1,2,6,9,$]

Affiliations:
[1]Howard Hughes Medical Institute, Departments of Chemistry[2] and Biology[9], Massachusetts Institute of Technology, Cambridge, Massachusetts 02139; cdrennan@mit.edu
[3]Current Address, Department of Chemistry, University of Massachusetts, Boston, MA 02125
[4]Current Address, Department of Chemistry, Bryn Mawr College, Bryn Mawr, PA 19010
[5]Department of Chemical and Environmental Engineering, University of Nottingham, University Park, Nottingham NG7 2RD
[6]Center for Environmental and Health Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139
[7]Current Address, School of Chemistry and Biochemistry and the Parker H. Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, Georgia 30332
[8]Stanford University Chemistry, Engineering, and Medicine for Human Health, Stanford University, Stanford, California, 94305
[$]Correspondence to: cdrennan@mit.edu

Running title: Structure of an NRPS cyclization domain

Keywords:
Crystal structure, molecular dynamics, epothilone, natural product, thiazoline, thiazole, PKS

ABSTRACT

Epothilones are natural products with anticancer activity that are biosynthesized by polyketide synthase (PKS)-nonribosomal peptide synthetase (NRPS) enzymes EpoA-F. A cyclization domain of EpoB (Cy) assembles the thiazole functionality from an acetyl group and L-cysteine via condensation, cyclization and dehydration. The PKS carrier protein of EpoA contributes the acetyl moiety, guided by a docking domain, whereas an NRPS EpoB carrier protein contributes L-cysteine. To visualize the structure of a cyclization domain with an accompanying docking domain, we have solved a 2.03 Å resolution structure of this bidomain EpoB unit, comprising residues M1-Q497 (62 kDa) of the 160 kDa EpoB protein. We find that the N-terminal docking domain is connected to the V-shaped Cy domain by a twenty-residue linker, but otherwise makes no contacts to Cy. Molecular dynamic simulations and additional crystal structures reveal a high degree of flexiblity for this docking domain, emphasizing the modular nature of the components of PKS-NRSP hybrid systems. These structures further reveal two 20-Å-long channels that run from distant sites on the Cy domain to the active site at the core of the enzyme, allowing two carrier proteins to dock with Cy and deliver their substrates simultaneously. Through mutagenesis and activity assays, catalytic residues N335 and D449 have been identified. Surprisingly these residues do not map to the location of the conserved HHxxxDG motif in the structurally homologous NRPS condensation (C) domain. Thus, although both C and Cy domains have the same basic fold, their active sites appear distinct.

Significance: Here we investigate the structural basis for cyclization activity in hybrid polyketide synthase-nonribosomal peptide synthetases. This first structure of a cyclization (Cy)

domain reveals an unexpected location for the enzyme active site, providing a fresh perspective on past mutational studies. Our structures also depict two twenty-Å-long channels that create routes for the two tethered substrates to simultaneously reach the buried active site, affording substrate condensation and cyclization. Along with the Cy domain, these structures contain a covalently-attached docking domain, providing insight into how protein modules work together to achieve uni-directionality in the biosynthesis of natural products.

Introduction

Epothilones are hybrid polyketide/nonribosomal peptide natural products that are indicated for treatment of metastatic or locally advanced breast cancers that are taxane-resistant (1, 2). They contain a large macrocycle with a thiazole-containing side chain, which is important for stabilizing microtubules and impairing cell division (Fig. 1*A*) (3–5). Azole heterocycles, such as thiazole, are commonly found in many natural products, and in different oxidation states (i.e. azolines and azolidines). They are more resistant to hydrolysis than a peptide bond, and their incorporation can increase a compound's affinity for a target biomolecule (6). Despite their importance, there is still much to learn about the structure and mechanism of the enzymes involved in their biosynthesis (7, 8). Here we explore the structural basis of activity of the catalytic domain that generates the 2-methylthiazoline precursor of epothilone natural products.

The epothilone biosynthetic gene cluster from *Sorangium cellulosum* is encoded by EpoA-K (Fig. 1*A*). EpoA, a polyketide synthase (PKS), and EpoB, a nonribosomal peptide synthetase (NRPS), are responsible for making the 2-methylthiazole functionality (9–12). 2-methylthiazole derives from the condensation of an acetyl group with L-cysteine (Fig. 1*B*). Before condensation, the acetyl moiety and L-cysteine are covalently attached to the carrier T domains of EpoA (acetyl-*S*-EpoA-T) and EpoB (cysteinyl-*S*-EpoB-T), respectively, via phosphopantetheine (Ppant) linkers (Fig. 1*A*, Fig. S1). Instead of a traditional condensation or C domain, EpoB has a cyclization (Cy) domain that performs both the amide bond forming condensation reaction and the cyclization/dehydration to form the five membered ring structure of 2-methylthiazoline (11). Two mechanisms have been proposed that differ in whether the amide bond forms first or second to the cyclization reaction (13, 14) (Fig. 1*B*). Subsequent oxidation by the flavin-dependent oxidase domain (EpoB-Ox) results in 2-methylthiazole (9, 11,

12). PKS–NRPS docking domains (15), referred to as EpoAdd and EpoBdd for the upstream donor and the downstream acceptor proteins, respectively, serve to localize the T domains to the appropriate intermodule junction, facilitating what has often been referred to as an assembly line biosynthetic process.

There is much that is unknown about Cy domains; the molecular basis for the differentiation of C and Cy domain activity is not established and the key catalytic residues have not been identified. Keating *et al.* predicted that the C and Cy domains would adopt similar structures (16), and early sequence alignments identified a DxxxxDxxS Cy domain sequence that replaces the C domain HHxxxDG catalytic motif (Fig. S2). Although conservation of the DxxxxDxxS sequence within known Cy domains suggests its importance, mutational analyses have been inconclusive as to which residues are critical for catalysis (14, 17, 18).

To provide insight into the cyclization activity of EpoB, we determined the X-ray structures of an EpoB construct that contains residues M1-Q497 (~62 kDa) of the full length EpoB enzyme from *S. cellulosem*. This 62 kDa unit, which we will call EpoBcy, was previously shown (18) to be an active Cy domain, capable of interacting *in trans* with constructs encoding the A, Ox, and T-domains of EpoB, and the T domain of EpoA, to synthesize 2-methylthiazole. This construct also contains the natural N-terminal docking domain, providing the first glimpse of this PKS to NRPS docking domain within a larger protein. These structural results, along with accompanying mutagenesis data, provide insight into the molecular basis of cyclization activity and have important implications regarding PKS/NRPS interprotein interaction.

Results

Structure of an NRPS Cy and docking domain protein complex

We determined two structures of EpoBcy in two different space groups. A 2.6-Å resolution structure of EpoBcy was solved, with two molecules in the asymmetric unit, in space group *R*32, by multiple isomorphous replacement techniques using data from five different heavy atom derivatives (Tables S1 and S2). A 2.03-Å resolution structure was solved in space group $P2_1$, using *R*32-EpoBcy as a molecular replacement search model, with one molecule per asymmetric unit (Table S2). The overall protein fold of the Cy domain (D76 – Q497) of EpoB is "V"-shaped, with the N- and C-terminal segments each comprising approximately one half of the "V" (Fig. 2*A*). The N- and C-terminal segments (D76 – K247 and S248 – Q497, respectively) contain αβα sandwich folds, resulting in a structure that loosely resembles a pseudodimer. The N-terminal segment of the Cy domain consists of a 5-stranded mixed β-sheet in which the last β-strand is donated from the adjacent C-terminal half of the protein, and the C-terminal segment contains a mixed 6-stranded β-sheet positioned almost perpendicular to the N-terminal β-sheet. This protein fold is similar to that of both NRPS condensation (16) and epimerization domains (19); EpoBcy aligns with the condensation domain of VibH from vibriobactin synthetase (PDB accession code 1L5A)(16), with an overall RMSD of 3.9 Å for 392 Cα atoms (Fig. 2*B*), and the epimerase domain of TycA from tyrocidine synthetase (PDB accession code 2XHG)(19), with an overall RMSD of 3.6 Å for 400 Cα atoms, as determined by the DALI server (20).

The N-terminal 55 residues of the EpoB protein make up the docking domain (EpoBdd), which recognizes the upstream EpoAdd to position the acetyl-*S*-EpoA-T domain for catalysis (21). EpoBdd adopts an αββαα fold, consisting of an initial α-helix, β-turn, and two final α-helices (Fig. 2*A*), and is connected to the Cy domain by a 20-residue linker (L56 – T75). This

docking domain structure is similar to the NMR structure of a docking domain at an NRPS-NRPS junction in the tubulysin system from *Angiococcus disciformis*, sharing an RMSD of 3.6 Å for 52 Cα atoms (PDB accession code 2JUG) (Fig. S3) (15).

Structures and MD simulations reveal conformational flexibility of docking domain

Three different conformations of the docking domain are observed in our structures, consistent with the presence of a flexible rather than rigid linker between the docking and Cy domains, and the existence of very little buried surface at the domain-domain interface in any of the structures (Fig. 3*A* and Fig. S4). The *R*32 crystal form reveals two conformations of the docking domain, with each molecule in the asymmetric unit adopting a different conformation. A third orientation is visible in the $P2_1$-EpoBcy structure (Fig. 3*A*). To further explore the flexibility of the docking domain, a 20 ns molecular dynamics (MD) simulation of the fully hydrated protein was run. Very little movement of the Cy domain was observed, whereas the docking domain sampled multiple positions, none of which were observed to make substantial interactions with the Cy domain (Fig. 3*B*, Movie S1). Thus, the connection between domains appears largely dependent on the covalent linker.

Structures suggest binding sites for T-domains of EpoA and EpoB

The EpoB Cy domain has the challenging task of interacting with carrier T domains from two separate proteins, each of which supply a different component to make 2-methylthiazoline **(**Fig. 1). Since EpoBdd facilitates interactions with the PKS T-domain of EpoA, the location of the docking domain in our structures identifies the approximate binding site for EpoA-T (referred to as site 1 in Fig. 4*A*). Insight into the binding location for the NRSP EpoB-T domain

comes from a structure of a NRPS T domain bound to the terminal surfactin A module (22). Structural superposition places the EpoB-T binding site on the C-terminal half of the Cy domain (Fig. S5) (referred to as site 2 in Fig. 4*A*). These putative T domain binding sites are non-overlapping, consistent with the proposal that both T domains bind EpoB at the same time (16). Intriguingly, the $P2_1$-EpoBcy structure shows an extended L shaped channel that connects putative T domain binding site 1 to the active site, and the active site to putative T domain binding site 2 (Fig. 4*A*). The distance between each putative T domain binding site and the active site is ~20 Å, the length of an extended Ppant arm. This physical relationship suggests an acetyl moiety tethered via a Ppant arm from the EpoA-T domain would sit juxtaposed to a L-cysteinyl moiety tethered via a Ppant arm from the EpoB T domain (Fig. 4*B*). Similar channels are found in both molecules that are present in the asymmetric unit of the *R*32 structure, however movement of surface loops at site 2 (the loops prior to β1 and β10), in addition to protein side chain differences near the active site result in channel constrictions (Fig. 4C,*D* and Figs. S2 and S6).

To investigate the conformational dynamics of residues near the active site that may allow for channel widening and clamping, we carried out MD simulations on the *R32* structure that had the contracted cavity (shown in Fig. 4*D*). As detailed above, in a 20 ns MD simulation the Cy domain does not move substantially, however certain protein residues located on both sides of the active site appear to undergo small movements of their backbone and slightly larger movements of side chains, with the result being that the cavity opens to resemble the contiguous channel observed in the $P2_1$ structure (Fig. 4*E*). These data suggest that small movements of residues can alternately contract and widen channels between the two T domain binding sites and the active site without the need for large movements of the protein backbone.

Structure and mutagenesis reveal unexpected location for Cy Active Site

The N- and C-terminal halves of the EpoB Cy domain form a stable interface, which in turn forms the putative substrate-binding channels that were described above. In addition to several hydrophobic residues that may mediate favorable interactions with the hydrophobic Ppant arm and acetyl moiety of substrate, the structure reveals a set of previously uncharacterized polar residues that may be involved in catalysis, including S80, Y81, D354, Q445 and D449 (Fig. 5*A,B* and Fig. S7). We mutated these residues and also N335, which was studied previously in the homologous BacA system (Table S3) (14). These six residues were mutated individually and assayed for activity using LC/MS-MS detection, monitoring formation of 2-methylthiazole-4-carboxlic acid (2MTCA) (Fig. S1). Three active site variants have severely compromised rates of product formation: D449A, N335A and Q445A have 2000-, 555- and 140-fold decreased activities compared to wild-type EpoBcy (Fig. 5*C*). The S80A, Y81F and D354A variants of EpoBcy display only moderate effects with 3-, 6- and 6-fold decreased activities, respectively. Putting these results in context with the structure has allowed us to localize the active site to the C-terminal half of the Cy structure, where the channel is lined with residues N335, Q445, and D449 (Fig 5*A,B*).

Surprisingly, these active site residues occupy a site that is distal to the previously identified (16, 17) DxxxxDxxS motif, the latter of which is on the N-terminal half of Cy (Fig. 5*A,B*). The DxxxxDxxS motif aligns well with the catalytic HHxxxDG motif of NRPS C domains as predicted (14, 17, 18)(Fig. 2*B*), but instead of playing a catalytic role, our structures suggest that the DxxxxDxxS motif (**D201-**LINV**D**LG-**S209** in EpoBcy) may be important for

maintaining the integrity of the substrate-channels. Importantly, neither Asp is free to interact with substrate. Rather, D201 is involved in a salt bridge with R85, which provides structure to one side of the channel, and D206 forms a salt bridge with R341 and a hydrogen bond to S209, supporting another side of the channel (Fig. 5*A,B*).

Discussion

Hybrid PKS/NRPSs are remarkable macromolecular assembly lines with carrier proteins delivering substrates from one enzyme module to the next and docking domains providing the intermodular communication that allows for the proper directionality (23, 15, 24). Our structures provide a visualization of the interactions between an N-terminal docking domain and a downstream enzyme within a NRPS module, and we find a bead-on-a-string type arrangement. Covalent attachment by a twenty-residue linker is all that is involved; EpoBdd makes no other contacts to the Cy domain. Thus, any docking domain could be substituted with no reengineering of the Cy domain protein surface required. Although EpoBdd is highly flexible, allowing it to search for its partner proteins (23), its attachment point to Cy appears key to its function. When EpoBdd localizes the EpoA-T to Cy through interaction with EpoAdd, this T domain will end up positioned near to one of two channel openings on Cy (site 1), allowing a substrate linked by a Ppant arm to reach down into the core of the Cy domain.

A second channel from a second T domain binding site (site 2) has been identified that is at a right angle from the site 1 channel. The existence of two channels allows for simultaneous binding of the two substrate-loaded T domains. By physically isolating the binding sites of the upstream and downstream carrier proteins, NRPS systems have developed a directionality that is important for defining the generation of a specific natural product. The length of each channel (~20 Å) matches the length of an extended Ppant arm (~20 Å), allowing us to predict that the

acetyl moiety from EpoA-T and L-cysteinyl moiety from EpoB-T will end up juxtaposed in the active site and also proximal to catalytic residues D449 and N335 (Fig. 5*B*). Notably, the EpoA channel appears a bit longer than necessary for an acetyl moiety to be accommodated (Fig. 5*A,B*), perhaps explaining the observation that larger substrates (propionyl-, isobutryl-, and benozyl-EpoA-T) can be turned over by Cy, albeit more slowly (11).

In our structures and MD simulations, we observe snapshots of more open and more closed states of the channel leading to the active site, and we expect that the binding of EpoA-T and EpoB-T domains at their respective binding sites will shift the equilibrium towards the more open state, and that T domain departure will shift toward the closed state, thus restricting active site access in the absence of substrates. Although conformational changes of protein backbone have been observed for C domain proteins that could contribute to channel opening and closing (25), here we see little to no movement of the backbone atoms of the three structures that display various degrees of channel openness. Also MD simulations show that side chain motion is sufficient to open and close the internal protein cavities. Thus for this Cy domain, there appears to be no need to invoke domain hinge motions in catalysis.

These studies have also allowed us to investigate how a Cy domain compares to a C domain. Prior to this report, many labs had performed biochemical characterizations of Cy domains with the expectation that the C and Cy domains would be very similar, yet they obtained inconsistent results from mutagenesis studies using different Cy domains (14, 17, 18). Now, with our report of the first Cy domain structure, the results of these biochemical studies are coming into focus. We do find that the EpoB Cy domain adopts a similar protein fold as the NRPS C domain validating previous predictions (16). Despite adopting a conserved protein fold, however, our results suggest that the catalytically important residues for Cy do not map to the

location of the highly conserved sequence motifs (HHxxxDG for C domains and DxxxxDxxS for Cy domains). Importantly, residues of the DxxxxDxxS sequence motif are not free to interact with substrate, instead forming salt bridges and hydrogen bonding networks that stabilize the elaborate channels that run through the core of the protein fold. Although substrate binding might cause residues of the DxxxxDxxS motif to break their interactions and be available for catalysis, we instead propose that D449 is key to catalysis, potentially serving as a catalytic base. This proposal is consistent with the 2000-fold effect on 2MTCA production when D449 is mutated to alanine.

Inspection of the proposed mechanisms in Fig. 1*B* shows a number of base-catalyzed steps that might be involved in this three step reaction: deprotonation of the amino group of cysteine for the peptide bond formation step with the acetyl moiety; deprotonation of the cysteine side chain for the cyclization reaction; and deprotonation of the ring NH for the dehydration reaction. It is possible that one residue may catalyze all three deprotonations, since the protonation state of the base could be reset after each step. Notably, the number of deprotonations equals the number of protonations, with the Ppant sulfur accepting one proton and the acetyl moiety oxygen accepting two protons as it is first reduced to a hydroxy group and then to water (Fig. 1*B*). Thus from the perspective of stoichiometry, D449 could assist in all three reactions, but a structure with substrate bound would help to evaluate the geometric prospects of a single residue catalyzing all three different steps.

We have also confirmed that N335 is catalytically important, having a 555-fold effect on 2MTCA production. Given that the corresponding asparagine residue in a chimeric engineered BacA Cy generated only an uncyclized dipeptide product when this asparagine was mutated to alanine (14), we expect that N335 is also involved in cyclization. Because N335 does not have a

titratable side chain, we do not believe that it is a catalytic base. Instead, it may serve to position the substrates appropriately for cyclization or stabilize intermediates through hydrogen bonding. In short, the structures of EpoBcy explain much of the previous biochemical work and also reveal D449 as a key residue.

Since the discovery of NRPS natural products, researchers have been interested in manipulating these modular assembly lines for the bioproduction of novel chemical compounds (26). For this to be achievable and efficient, we must increase our understanding of how these systems function at the molecular level (27). This work presents important structural information regarding two NRPS domains: the Cy domain that produces five-membered heterocycles for assembly into larger NRPS products and the docking domain that provides specificity between two different PKS/NRPS proteins to interact *in trans*.

METHODS

The EpoBcy protein construct was expressed and purified as described (18), with minor modifications detailed in the *SI Materials and Methods*. EpoBcy site specific mutagenesis was performed using standard protocols, and activity assays (18) were adapted for product detection by LC/MS-MS (*SI Materials and Methods*). Purified wild-type EpoBcy was crystallized using the vapor diffusion method, and X-ray diffraction experiments, model building and refinement are detailed in the *SI Materials and Methods*. A 20 ns MD simulation was performed in GROMACS, and parameters are described in the *SI Materials and Methods*.

Acknowledgements: This work was supported by the CEHS Center core grant (P30-ES002109); and C.L.D is a Howard Hughes Medical Institute Investigator. This work is based upon research conducted at beamline 9-2 at the Stanford Synchrotron Radiation Lightsource (SSRL) and at the Advanced Photon Source (APS) on the Northeastern Collaborative Access Team (NE-CAT) beamlines. NECAT at APS is supported by grants from the National Center for Research Resources (5P41RR015301-10) and the National Institute of General Medical Sciences at the National Institutes of Health (8 P41 GM103403-10). APS is an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science by Argonne National Laboratory, and is also supported by the U.S. DOE under Contract No. DE-AC02-06CH11357.

Author Contributions: W.L.K. and C.T.W. contributed reagents and expertise on EpoBcy purification. Y.K. purified and crystallized the protein. D.P.D. obtained monoclinic protein crystals, determined the structures, prepared mutant proteins, performed the enzyme assays and LC-MS/MS analysis. K.T. assisted with the latter. A.K.C. conducted the molecular dynamics simulations. D.P.D. and C.L.D. wrote the manuscript.

References

1. Rivera E, Gomez H (2010) Chemotherapy resistance in metastatic breast cancer: the evolving role of ixabepilone. *Breast Cancer Res* 12(Suppl 2):S2.
2. Rivera E, Lee J, Davies A (2008) Clinical development of ixabepilone and other epothilones in patients with advanced solid tumors. *Oncologist* 13(12):1207–1223.
3. Khrapunovich-Baine M, et al. (2011) Hallmarks of molecular action of microtubule stabilizing agents: effects of epothilone B, ixabepilone, peloruside a, and laulimalide on microtubule conformation. *J Biol Chem* 286(13):11765–11778.
4. Kumar A, et al. (2010) Interaction of epothilone B (patupilone) with microtubules as detected by two-dimensional solid-state NMR spectroscopy. *Angew Chemie Int Ed* 49(41):7504–7507.
5. Thompson CA (2007) FDA approves new breast cancer treatment. *Am J Heal Pharm* 64(23):2406.
6. Walsh CT, Nolan EM (2008) Morphing peptide backbones into heterocycles. *Proc Natl Acad Sci U S A* 105(15):5655–5656.
7. Hur GH, Vickery CR, Burkart MD (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* 29(10):1074.
8. Arnison PG, et al. (2013) Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Nat Prod Rep* 30(1):108–160.
9. Julien B, et al. (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from Sorangium cellulosum. *Gene* 249(1-2):153–160.
10. Tang L, et al. (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science (80-)* 287(5453):640–642.
11. Chen H, O'Connor S, Cane DE, Walsh CT (2001) Epothilone biosynthesis: assembly of the methylthiazolylcarboxy starter unit on the EpoB subunit. *Chem Biol* 8(9):899–912.
12. Molnár I, et al. (2000) The biosynthetic gene cluster for the microtubule-stabilizing agents epothilones A and B from Sorangium cellulosum So ce90. *Chem Biol* 7(2):97–109.
13. Gehring AM, Mori I, Perry RD, Walsh CT (1998) The nonribosomal peptide synthetase HMWP2 forms a thiazoline ring during biogenesis of yersiniabactin, an iron-chelating virulence factor of yersinia pestis. *Biochemistry* 37(48):11650–17104.

14. Duerfahrt T, Eppelmann K, Müller R, Marahiel MA (2004) Rational design of a bimodular model system for the investigation of heterocyclization in nonribosomal peptide biosynthesis. *Chem Biol* 11(2):261–271.
15. Richter CD, Nietlispach D, Broadhurst RW, Weissman KJ (2008) Multienzyme docking in hybrid megasynthetases. *Nat Chem Biol* 4(1):75–81.
16. Keating TA, Marshall CG, Walsh CT, Keating AE (2002) The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat Struct Biol* 9(7):522–526.
17. Marshall CG, Hillson NJ, Walsh CT (2002) Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. *Biochemistry* 41(1):244–250.
18. Kelly WL, Hillson NJ, Walsh CT (2005) Excision of the epothilone synthetase B cyclization domain and demonstration of in trans condensation/cyclodehydration activity. *Biochemistry* 44(40):13385–13393.
19. Samel S a., Czodrowski P, Essen LO (2014) Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallogr Sect D Biol Crystallogr* 70(5):1442–1452.
20. Holm L, Rosenström P (2010) Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38:W545–W549.
21. Liu F, Garneau S, Walsh CT (2004) Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis. *Chem Biol* 11(11):1533–1542.
22. Tanovic A, Samel SA, Essen LO, Marahiel MA (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science (80-)* 321(5889):659–663.
23. Weissman KJ, Müller R (2008) Protein–Protein Interactions in Multienzyme Megasynthetases. *ChemBioChem* 9(6):826–848.
24. Shen B, et al. (2001) The biosynthetic gene cluster for the anticancer drug bleomycin from Streptomyces verticillus ATCC15003 as a model for hybrid peptide-polyketide natural product biosynthesis. *J Ind Microbiol Biotechnol* 27(6):378–385.
25. Bloudoff K, Rodionov D, Schmeing TM (2013) Crystal structures of the first condensation domain of CDA synthetase suggest conformational changes during the synthetic cycle of nonribosomal peptide synthetases. *J Mol Biol* 425(17):3137–3150.
26. Kennedy J, Hutchinson CR (1999) Nurturing nature: engineering new antibiotics. *Nat*

Biotechnol 17(6):538–9.

27. Khosla C, Herschlag D, Cane DE, Walsh CT (2014) Assembly line polyketide synthases: mechanistic insights and unsolved problems. *Biochemistry* 53(18):2875–2883.
28. Walsh CT (2008) The chemical versatility of natural-product assembly lines. *Acc Chem Res* 41(1):4–10.
29. Strieker M, Tanović A, Marahiel MA (2010) Nonribosomal peptide synthetases: structures and dynamics. *Curr Opin Struct Biol* 20(2):234–240.
30. Bond CS (2003) TopDraw: a sketchpad for protein structure topology cartoons. *Bioinformatics* 19(2):311–312.
31. Ho BK, Gruswitz F (2008) HOLLOW: generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* 8:49.

Figure Legends

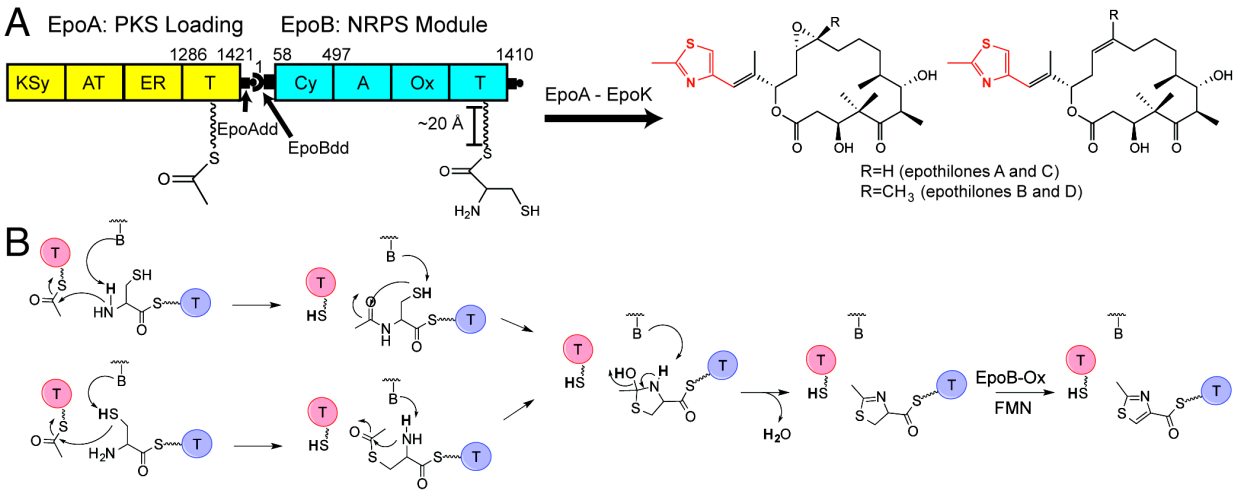


Fig. 1. Scheme for epothilone biosynthesis, focusing on thiazol(in)e formation by the NRPS cyclization domain. (*A*) Epothilones (right) are large macrocycles with a thiazole-containing side chain (red), which are produced by a hybrid PKS/NRPS EpoA-K. EpoA and EpoB are responsible for producing the thiazole side chain and have the following domains (9–12): KSy, ketosynthase-like domain; AT, acyltransferase domain; ER, enoyl reductase; and T, an acyl carrier protein. The NRPS EpoB domains include: A, adenylation domain; Cy, cyclization; Ox, a flavin-dependent oxidase domain; and T, a carrier protein domain (28, 29). The Ppant group of each T domain is represented as ∿∿S, and residue numberings for the EpoA-T, and EpoBcy constructs are indicated. (*B*) Proposed mechanisms for condensation and cyclodehydration. The EpoA-T domain and EpoB-T domains are represented as red and blue spheres, respectively.

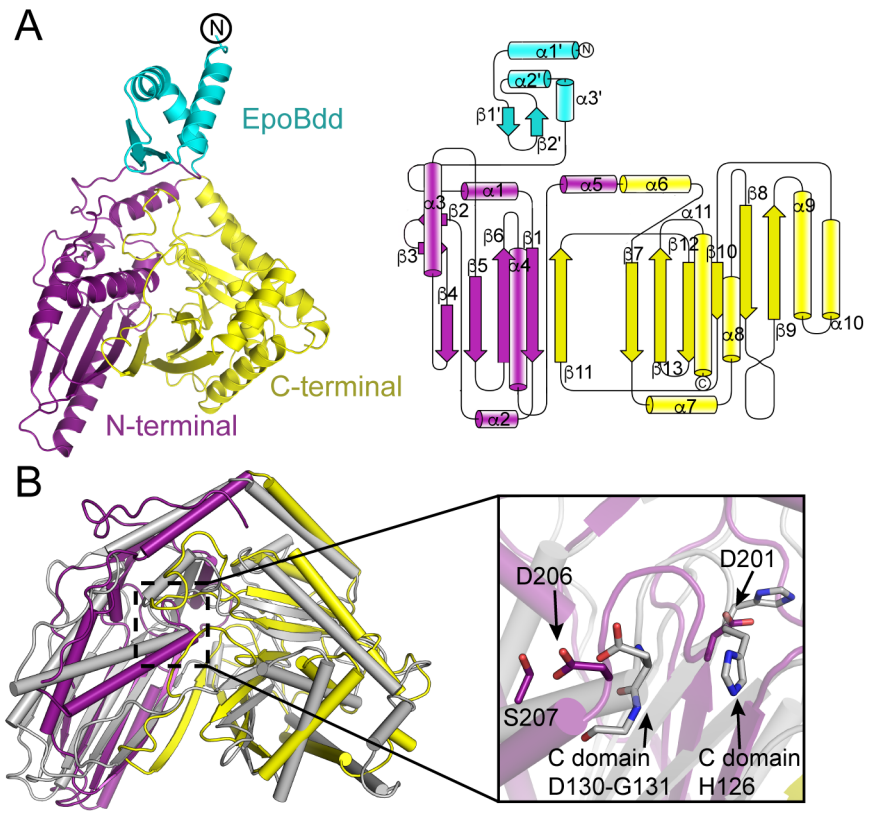


Fig. 2. Overall structure of EpoBcy. (*A*) Left: ribbon structure of monomeric EpoBcy, with the N-terminal docking domain in cyan, and the N- and C-terminal halves of the Cy domain in purple and yellow, respectively. Right: Topology diagram of EpoBcy generated using TOPDRAW (30). (*B*) Structural superposition of the NRPS Cy and C domains. The EpoB Cy domain is colored as in *A*; the C domain from VibH is colored in gray (PDB code 1L5A)(16). Helices are represented as cylinders, and the EpoB docking domain is omitted for clarity. Inset is a close-up view of the C domain H125-H126-xxx-D130-G131 motif and the respective Cy domain D201-xxxx-D206-xx-S209 sequence, depicted as sticks.

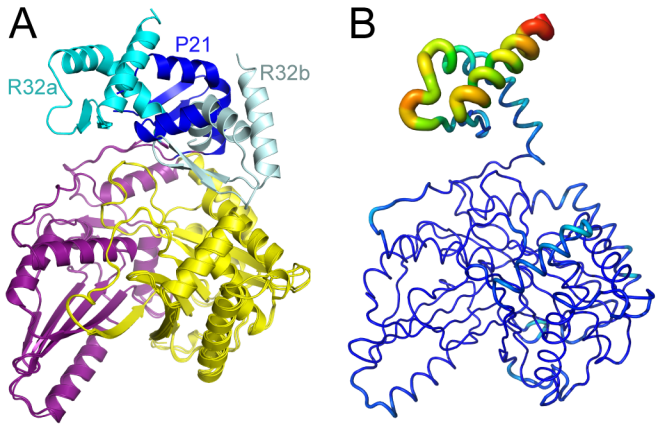


Fig. 3. The PKS/NRPS docking domain is flexibly tethered to EpoBcy. (*A*) Structural alignment of the three monomeric forms of EpoBcy observed from the $P2_1$ and $R32$ crystal structures. The Cy domain is colored as in Fig. 2, and the three observed orientations of the docking domain are colored cyan, pale cyan, and blue. (*B*) A 20 ns MD simulation of EpoBcy with increased protein movement indicated by the thickness and color of the protein trace, from thin blue (least motion) to thick red (greatest motion). Also see Movie S1.

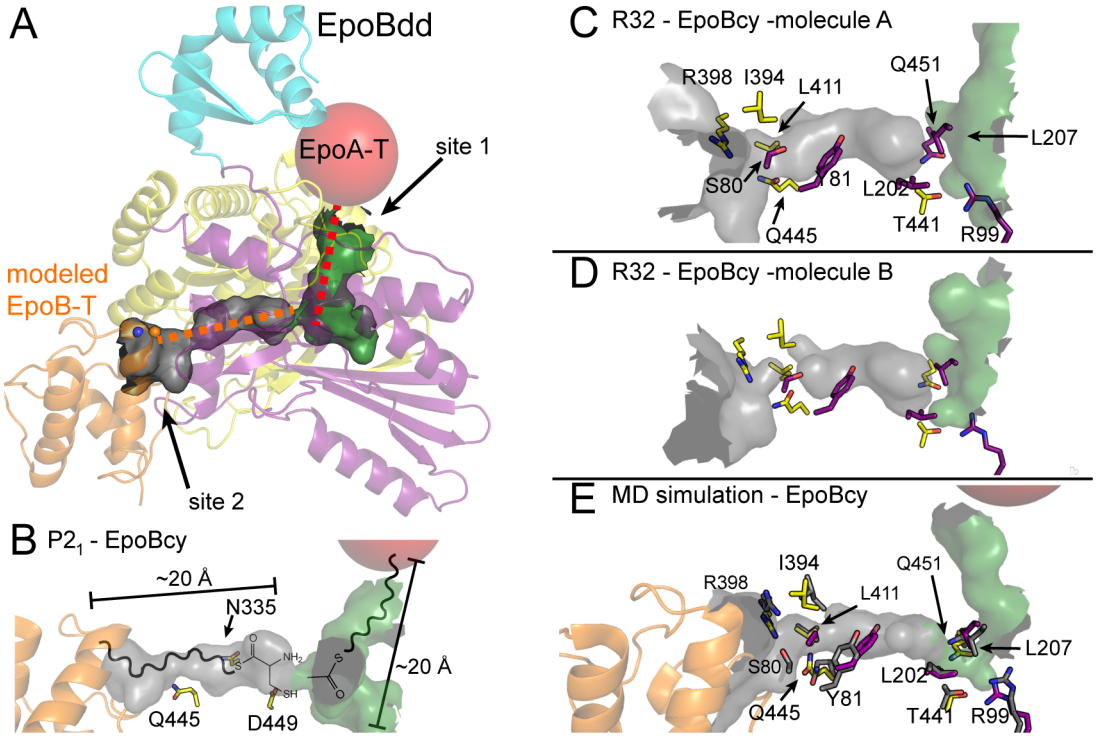
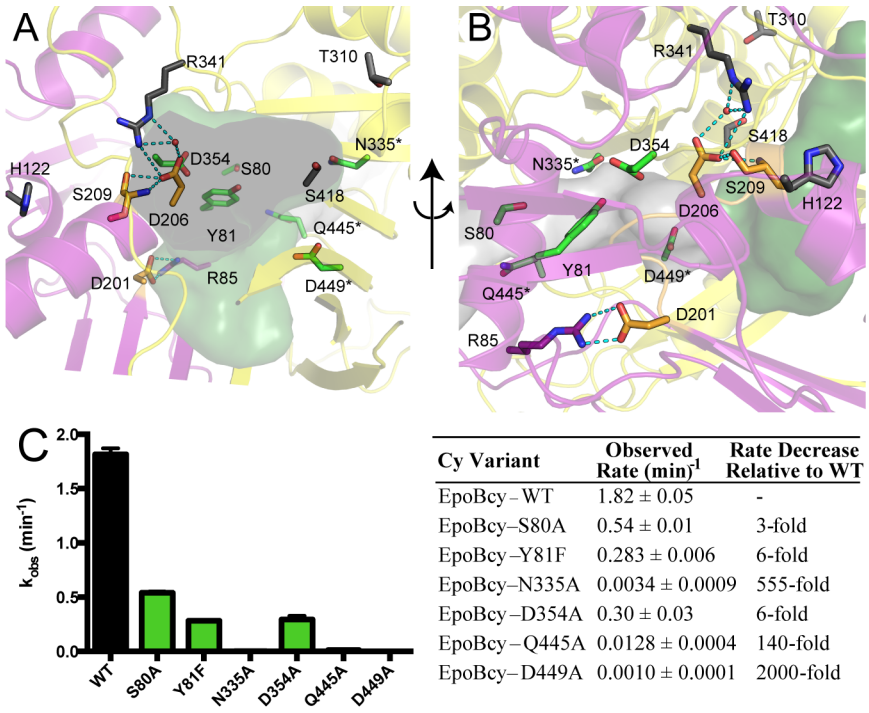


Fig. 4. Proposed binding sites for substrate T domains. (*A*) *P*2$_1$-EpoBcy (colored as in Fig. 2) is displayed in ribbons with the putative active site channels displayed in green and grey (calculated by HOLLOW) (31). The binding of EpoA-T at "site 1" is predicted by proximity to the docking domain and modeled as a red sphere; EpoB-T (orange, PDB 2VSQ) is modeled at "site 2" from an alignment of EpoBcy and SrfC S1003A variant (Fig. S5), with the site of Ppant attachment displayed in ball and stick (22). The entrance paths are the approximate length of the Ppant prosthetic group (dashed orange and red lines). (*B*) Close-up view of the observed active site channel of *P*2$_1$-EpoBcy, with locations of the EpoA-T and EpoB-T domains represented as in *A*. Ppant arms, represented as squiggly lines for acetyl-*S*-EpoA-T and cysteinyl-*S*-EpoB-T, are anticipated to meet within the active site channel where condensation and cyclization will occur. The location of the catalytically important residues D449, N335, and Q445 are indicated. (*C*) Channel calculation for *R*32-EpoBcy molecule A shows less connectivity between the site 1 channel (green) and the site 2 channel (grey) than in the most "open" channel, which is shown in *B*. Residues shown as sticks have different positions in the "open" channel; carbons are colored purple if from the N-terminal domain and yellow if from the C-terminal domain. (*D*) Channel calculation for *R*32-EpoBcy molecule B shows the most restricted internal cavities. (*E*) The 20 ns MD simulation of most "closed" EpoBcy *R32* structure (molecule B, shown in panel *D*), leads to an opening of the channel such that this channel now resembles the channel of the *P2$_1$* structure in *B*. The overlay of a representative time point towards the end of the MD simulation is displayed with sticks colored gray. *R*32- and *P*2$_1$-EpoBcy are colored as in Fig. 2*A*.



Cy Variant	Observed Rate $(min)^{-1}$	Rate Decrease Relative to WT
EpoBcy – WT	1.82 ± 0.05	-
EpoBcy–S80A	0.54 ± 0.01	3-fold
EpoBcy–Y81F	0.283 ± 0.006	6-fold
EpoBcy–N335A	0.0034 ± 0.0009	555-fold
EpoBcy–D354A	0.30 ± 0.03	6-fold
EpoBcy–Q445A	0.0128 ± 0.0004	140-fold
EpoBcy–D449A	0.0010 ± 0.0001	2000-fold

Fig. 5. Identification of EpoBcy active site residues. (*A*) Putative EpoBcy active site channels (identified by the software HOLLOW) are depicted, oriented at the EpoA-T binding channel opening, with protein shown in ribbons; the EpoA-T channel is gray surface and the EpoB-T channel is green surface. Polar residues within this cavity that were mutated in this work are shown in sticks and colored with green carbons, with asterisks denoting those mutations that most dramatically affected enzyme activity. The previously identified D201xxxxD206xxS209 sequence is colored with orange carbons, and previously mutated residues have gray carbons (see Table S3 for summary of mutational studies). R85 is colored with purple carbons and has not been mutated. Stereoviews are shown in Fig. S7. (*B*) Same as panel *A*, but rotated approximately 90° clockwise about the vertical axis. (*C*) Calculated rates of 2-methylthiazole-4-carboxylic acid formation by LC/MS-MS for EpoBcy variants in this work.