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

Daniel P. Dowlingab,1, Yan Kunga,3, Anna K. Croftc, Koli Taghizadehd, Wendy L. Kellye,4, Christopher T. Walshc, and Catherine L. Drennanb,df2

*Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139; 2Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; 3Department of Chemical and Environmental Engineering, University of Nottingham, Nottingham NG7 2RD, United Kingdom; 4Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; 5Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115; and 6Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Janet L. Smith, University of Michigan, Ann Arbor, MI, and accepted by Editorial Board Member Michael A. Marletta September 16, 2016 (received for review May 27, 2016)

Epothilones are thiazole-containing 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 solved a 2.03-Å resolution structure of this bimodal 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 20-residue linker but otherwise makes no contacts to Cy. Molecular dynamic simulations and additional crystal structures reveal a high degree of flexibility for this docking domain, emphasizing the modular nature of the components of PKS-NRPS 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 20-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 unidirectionality in the biosynthesis of natural products.


The authors declare no conflict of interest.

This article is a PNAS Direct Submission. J.L.S. is a Guest Editor invited by the Editorial Board.

Data deposition: The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5T81 and 5T72).

1Present address: Department of Chemistry, University of Massachusetts, Boston, MA 02125.

2To whom correspondence may be addressed. Email: daniel.dowling@umb.edu or cdrennan@mit.edu.

3Present address: Department of Chemistry, Bryn Mawr College, Bryn Mawr, PA 19010.

4Present address: School of Chemistry and Biochemistry and the Parker H. Petit Institute for Biotechnology and Bioengineering, Georgia Institute of Technology, Atlanta, GA 30332.

This article contains supporting information online at www.pnas.org/cgi/doi/10.1073/pnas.1608615113.

1073/pnas.1608615113/-/DCSupplemental.

Epothilones are hybrid polyketide/nonribosomal peptide natural products 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. L4) (3, 4). Azole heterocycles, such as thiazoles, are commonly found in many natural products and in different oxidation states (i.e., azolines and azolindines). They are more resistant to hydrolysis than a peptide bond, and their incorporation can increase a compound’s affinity for a target biomolecule (5). Despite their importance, there is still much to learn about the structure and mechanism of the enzymes involved in their biosynthesis (6, 7). Here we explore the structural basis of activity of the catalytic domain that generates the 2-methylthiazline precursor of epothilone natural products.

The epothilone biosynthetic gene cluster from Sorangium cellulosum is encoded by epoA-K (Fig. L4). EpoA, a polyketide synthase (PKS), and EpoB, a nonribosomal peptide synthetase (NRPS), are responsible for making the 2-methylthiazline functionality (8-11). 2-Methylthiazline derives from the condensation of an acetyl group with 1-cysteine (Fig. 1B). 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 (Ppnt) linkers (Fig. L4 and 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 (10). Two mechanisms have been proposed that differ in whether the amide bond forms first or second to the cyclization reaction (12, 13) (Fig. 1B). Subsequent oxidation by the flavin-dependent oxidase domain (EpoB-Ox) results in 2-methylthiazole (8, 10, 11). PKS-NRPS docking domains (14), referred to as EpoAdd and EpoBdld 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 (15), and early sequence alignments identified a DxxxxDxxS Cy domain sequence that replaces the C domain HHxXXDG catalytic motif (Fig. S2). Although conservation of the

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 20-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 unidirectionality in the biosynthesis of natural products.
DxxxDxxS sequence within known Cy domains suggests its importance, mutational analyses have been inconclusive as to which residues are critical for catalysis (13, 16, 17).

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. cellulosum. This 62-kDa unit, which we will call EpoBcy, was previously shown (17) 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 cognate 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 insights into the molecular basis of cyclization activity and have important implications regarding PKS-NRPS interprotein interactions.

Results

Structure of an NRPS Cy and Docking Bidomain. 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 R32, 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 P21, using R32-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. 2A). 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 five-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 six-stranded β-sheet positioned almost perpendicular to the N-terminal β-sheet. This protein fold is similar to that of both NRPS condensation domains (15, 18–21) and epimerization or E domains (22, 23). For example, EpoBcy aligns with VibH from vibriobactin synthetase (15) with an overall RMSD of 3.9 Å for 392 Cα atoms (Fig. 2B).

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 (24). EpoBdd adopts an αββαα fold, consisting of an initial α-helix, β-turn, and two final α-helices (Fig. 2A), and is connected to the Cy domain by a 20-residue linker (L56–T75). There is one other
docking domain structure with this fold, TubCdd from the Angiococcus disciformitis tubulysin system (14). This docking domain functions at a NRPS-NRPS junction and shares with EpoBdd 33% sequence identity and an RMSD of 3.6 Å for 52 Ca atoms (Fig. S3) (14). The β-hairpin and the third α-helix (α3’) of the αββαα fold show the largest deviation between these two docking domain structures (Fig. S3). In EpoBdd, the β-hairpin is tipped inward slightly toward α2’, whereas α3’ is swung slightly away from the structural core (Fig. S3D). This difference in the β-hairpin is interesting given that this region is predicted to be a binding site of the cognate recognition sequence (14, 25). However, it is not currently clear if these variations are mechanistically significant or an artifact of a difference in oligomeric state. In solution and without its cognate protein, TubCdd is a dimer, whereas EpoBdd is monomeric in our structure. Notably, the β-hairpin and the third α-helix (α3’) are both involved in the TubCdd dimer interface. Unfortunately, no structure of TubCdd with a partner protein is available that could clarify the relevance of the dimeric unit and provide additional information about protein interaction specificity.

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. 3A and Fig. S4). The R32 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 P21_EpoBcy structure (Fig. 3A). An alignment of these three docking domain structures with each other shows strong agreement (RMSD of less than 0.7 Å for all Cα atoms), indicating that this domain moves as a rigid body (Fig. S3C). To further explore the flexibility of the docking domain, a 20-ns molecular dynamics (MD) simulation of the fully hydrated protein was run. Little movement of the Cy domain was observed in comparison with the docking domain, whereas the α-3′ region of the docking domain showed substantial interactions with the Cy domain; thus, the connection between domains appears largely dependent on the covalent linker.

Structures Suggest T-Domain Binding Sites. 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). Because 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. 4A). Of the three observed conformations of EpoBdd (Fig. 3D), EpoBdd is the closest to site 1, whereas the P21 domain position is 10 Å farther removed and P32b is yet another 10 Å farther. Superimposition of EpoBcy with the recently reported structure of a holoenzyme T–E bidomain from gramicidin synthetase (23) provides further support for this upstream T domain binding site (Fig. S5A). The T domain in the gramicidin synthetase is adjacent to the R32h-position of EpoBdd when the EpoB Cy domain and gramicidin synthase E domains are superimposed.

Insight into the binding location for the NRPS EpoB-T domain comes from two structures of NRPS modules: the terminal surfactin A module (18) and the terminal holo-AB3403 module (19), in which the internal T domain of each module is positioned to interact with its C domain. Structural superimposition (Fig. S5 C and D) identifies the EpoB-T binding site as site 2 in Fig. 4A, a position that is at the interface of the N- and C-terminal halves of the Cy domain and is proximal to α1 and α10 (Fig. S2). Putative upstream and downstream T domain binding sites 1 and 2 (Fig. 4A) are nonoverlapping, consistent with the proposal that both T domains bind EpoBcy at the same time (15). Intriguingly, the P21_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. 4A). 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 cysteinyl moiety tethered via a Ppant arm from the EpoB T domain (Fig. 4B). Interestingly, a superimposition of EpoBcy with the terminal holo-AB3403 module (19) positions the Ppant arm bound in the holo-AB3403 structure into the EpoBcy site 2 channel (gray in Fig. 4B), demonstrating conservation of the channel position at least between these two structures. In contrast to this conservation, superimposition of EpoBcy with gramicidin synthetase E domain does not result in the placement of the Ppant arm bound to the E domain into the EpoBcy site 1 channel (green in Fig. 4B and Fig. S5B). Inspection of the structural superimposition (Fig. S5B) shows that four β-strands in the E domain structure are shifted with respect to the equivalent strands in EpoBcy creating an alternative channel that is positioned next to, and not overlapping with, the channel in EpoBcy. The latter provides an example of how small structural differences may influence substrate positioning.

The three EpoBcy structures show channels with different degrees of openness due to variations in surface loop positions and in side chain positions (Fig. 4 B–D and Fig. 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. 4D). As mentioned 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 P21 structure (Fig. 4E). 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.

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 P21 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 color of the protein trace, from blue (least motion) to red (greatest motion). The EpoBcy αx trace is displayed in ribbons. Also see Movie S1.
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. 5A and B and Fig. S7). We mutated these residues and also N335, which was studied previously in the homologous BacA system (Table S3) (13). 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 2,000-, 555-, and 140-fold decreased activities compared with WT EpoBcy (Fig. 5C). The S80A, Y81F, and D354A variants of EpoBcy display only moderate effects with three-, six-, and sixfold 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. 5A and B).

Surprisingly, these active site residues occupy a site that is distal to the previously identified DxxxxDxxS motif (15, 16), the latter of which is on the N-terminal half of Cy (Fig. 5A and B). The DxxxxDxxS motif aligns well with the catalytic HHxxxDG motif of NRPS C domains as predicted (13, 16, 17) (Fig. 2B), but instead of playing a catalytic role, our structures suggest that the DxxxxDxxS motif (D201-LINVDDLG-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. 5A and 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 (14, 25, 26). 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 20-residue linker is all that is involved: EpoBdd makes no other contacts to the Cy domain. Thus, any docking domain could be substituted for EpoBdd with no re-engineering of the Cy domain protein surface required. This result expands on the previous finding that switching EpoA modules can result in transfer of the different substrate unit to EpoB (27). Although EpoBdd is highly flexible, allowing it to search for its partner proteins, its attachment point to Cy appears key to its function. When EpoBdd localizes the EpoA-T to Cy through interaction with EpoAdd, the 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 (~20 Å), allowing us to predict that the acetyl moiety from EpoB-T will end up juxta-oriented with the cysteinyl moiety from EpoA-T.

Notably, the EpoA channel appears a 20 Å, and it is rotated ~90° clockwise about the vertical axis to match orientation shown in Fig. 4A. (C) Calculated rates of 2-methylthiazole-4-carboxylic acid formation by LC/MS-MS for EpoBcy variants in this work. Data are the average ± SD for three replicates.

**Methods**

The EpoBcy protein construct was expressed and purified as previously described (17), with minor modifications detailed in SI Methods. EpoBcy site-specific mutagenesis was performed using standard protocols, and activity assays (17) were adapted for product detection by LC/MS-MS (SI Methods). Purified WT EpoBcy was crystallized using the vapor diffusion method, and X-ray diffraction