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

Citation: Dowling, Daniel P.; Kung, Yan; Croft, Anna K.; Taghizadeh, Koli; Kelly, Wendy L.; Walsh, Christopher T. and Drennan, Catherine L. "Structural Elements of an NRPS Cyclization Domain and Its Intermodule Docking Domain." Proceedings of the National Academy of Sciences 113, no. 44 (October 2016): 12432–12437. © 2016 National Academy of Sciences

As Published: http://dx.doi.org/10.1073/pnas.1608615113

Publisher: National Academy of Sciences (U.S.)

Persistent URL: http://hdl.handle.net/1721.1/109039

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

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

Daniel P. Dowling[a,b,1,2], Yan Kung[b,3], Anna K. Croft[c], Koli Taghizadeh[d], Wendy L. Kelly[e,4], Christopher T. Walsh[e], and Catherine L. Drennan[a,b,d,f,2]

[a]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139; [c]Department of Chemical and Environmental Engineering, University of Nottingham, Nottingham NG7 2RD, United Kingdom; [d]Center for Environmental Health Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139; [e]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115; and [f]Department 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 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 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.

crystal structure | molecular dynamics | epothilone | natural product

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. 1*A*) (3, 4). Azole heterocycles, such as thiazoles, 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 (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-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 (8–11). 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* 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. 1*B*). 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 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 (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 uni-directionality in the biosynthesis of natural products.

Author contributions: D.P.D., Y.K., A.K.C., K.T., and C.L.D. designed research; D.P.D., Y.K., A.K.C., and K.T. performed research; W.L.K. and C.T.W. contributed new reagents/analytic tools; D.P.D., Y.K., A.K.C., K.T., and C.L.D. analyzed data; and D.P.D. and C.L.D. wrote the paper.

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 5T7Z).

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

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

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

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

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1608615113/-/DCSupplemental.

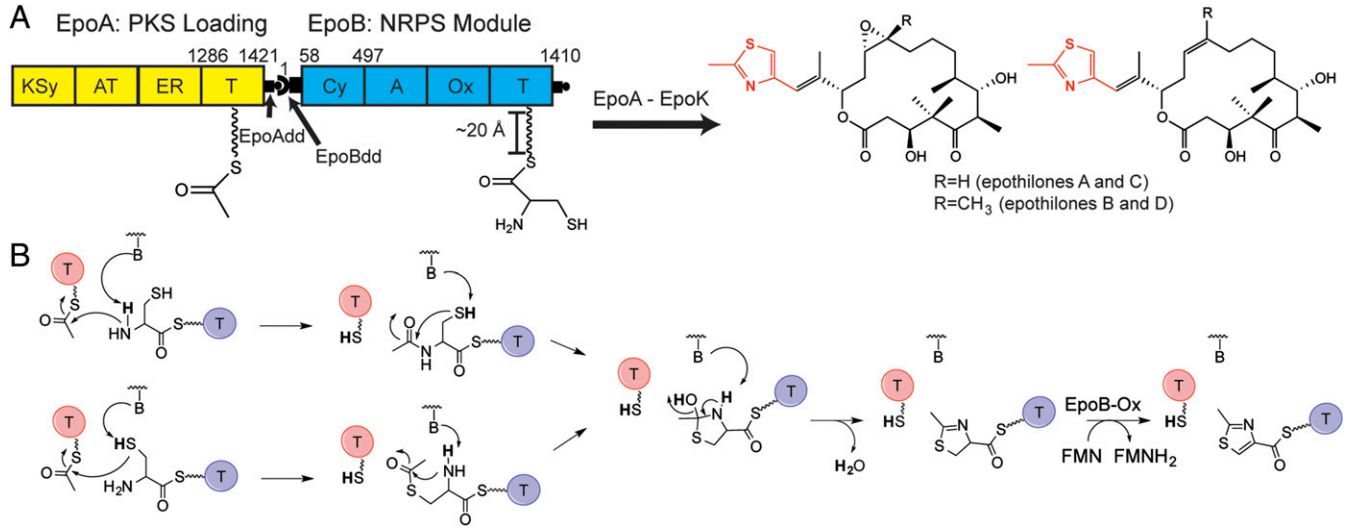


Fig. 1. Scheme for epothilone biosynthesis, focusing on thiazol(in)e formation by the NRPS cyclization domain. (*A*) Epothilones (*Right*) are large macrocycles with a thiazole-containing side chain (red), which are produced by a hybrid PKS-NRPS, EpoA-K. EpoA and EpoB are responsible for producing the thiazole side chain and have the following domains (8–11): KSy, ketosynthase-like domain; AT, acyltransferase domain; ER, enoyl reductase; and T, an acyl carrier protein. The NRPS EpoB domains include the following: A, adenylation domain; Cy, cyclization; Ox, a flavin-dependent oxidase domain; and T, a carrier protein domain (29). The Ppant group of each T domain is represented as a squiggly line 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.

DxxxxDxxS 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 *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 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. 2*B*).

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. 2*A*), and is connected to the Cy domain by a 20-residue linker (L56–T75). There is one other

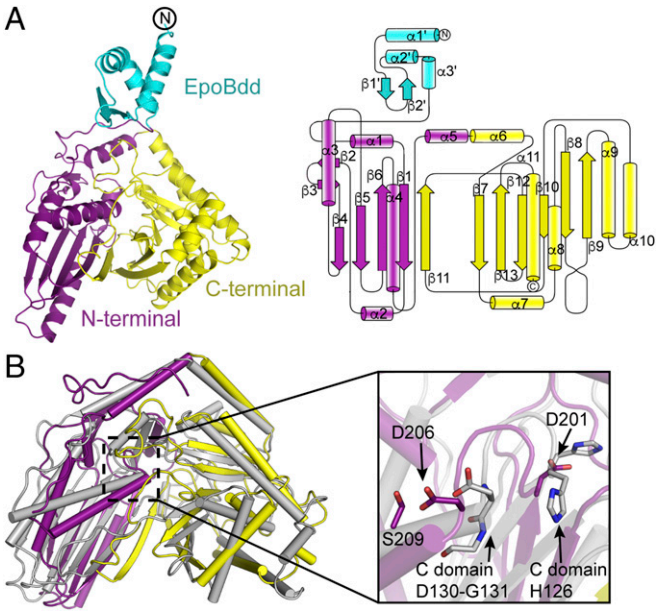


Fig. 2. Overall structure of EpoBcy. (*A*) (*Left*) Ribbon structure of monomeric EpoBcy, with the N-terminal docking domain in cyan, and the N- and C-terminal halves of the Cy domain in purple and yellow, respectively. (*Right*) Topology diagram of EpoBcy generated using TOPDRAW (31). (*B*) Structural superimposition 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 ID code 1L5A) (15). Helices are represented as cylinders, and the EpoB docking domain is omitted for clarity. (*Inset*) 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.

docking domain structure with this fold, TubCdd from the *Angiococcus disciformitus* tubulysin system (14). This docking domain functions at an NRPS-NRPS junction and shares with EpoBdd 33% sequence identity and an RMSD of 3.6 Å for 52 Cα 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. S3*D*). 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 theses 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. 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*). An alignment of these three docking domain structures with each other shows strong agreement (RMSD of less than 0.7 for all Cα atoms), indicating that this domain moves as a rigid body (Fig. S3*C*). 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, which sampled a larger range of positions especially during the first 5 ns of the simulation (Fig. 3*B* and Movie S1). None of these conformations were observed to make substantial interactions with the Cy domain; thus, the connection between domains appears largely dependent on the covalent linker.

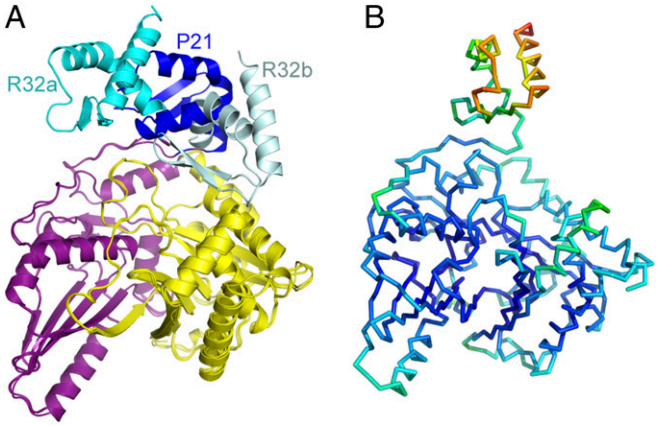


Fig. 3. The PKS-NRPS docking domain is flexibly tethered to EpoBcy. (*A*) Structural alignment of the three monomeric forms of EpoBcy observed from the $P2_1$ and *R*32 crystal structures. The Cy domain is colored as in Fig. 2, and the three observed orientations of the docking domain are colored cyan, 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 cα trace is displayed in ribbons. Also see Movie S1.

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. 4*A*). Of the three observed conformations of EpoBdd (Fig. 3*A*), R32b is the closest to site 1, whereas the $P2_1$ domain position is ∼10 Å farther removed and R32a 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. S5*A*). The T domain in the gramicidine synthetase is adjacent to the R32b-position of EpoBdd when the EpoB Cy domain and gramicidine 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. 4*A*, 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. 4*A*) are nonoverlapping, consistent with the proposal that both T domains bind EpoBcy at the same time (15). 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*). 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. 4*B*), demonstrating conservation of the channel position at least between these two structures. In contrast to this conservation, superimposition of EpoBcy with gramicidine synthase 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. 4*B* and Fig. S5*B*). Inspection of the structural superimposition (Fig. S5*B*) 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. 4*D*). 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 $P2_1$ 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.

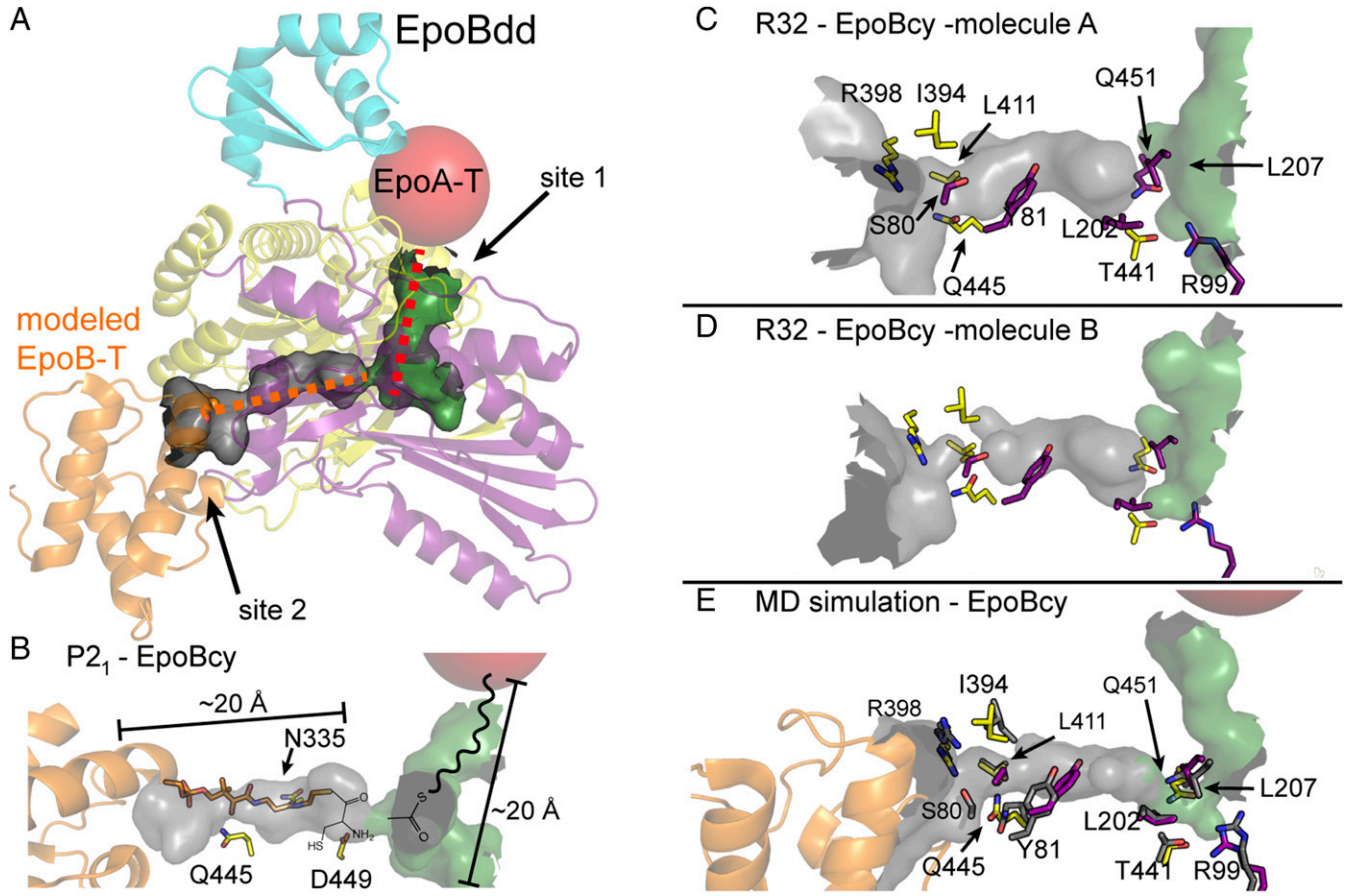


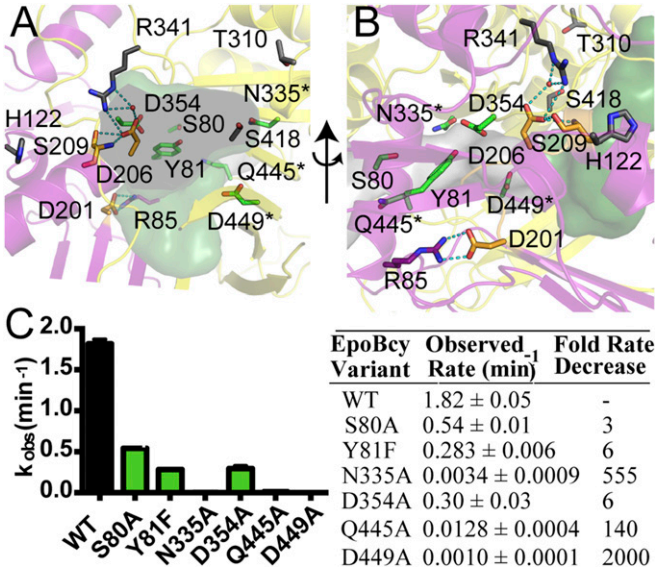
Fig. 4. Proposed binding sites for substrate T domains. (*A*) $P2_1$-EpoBcy (colored as in Fig. 2) is displayed in ribbons with the putative active site channels displayed in green and gray (calculated by HOLLOW) (32). The binding of EpoA-T at site 1 is predicted by proximity to the docking domain and is shown as a red sphere; EpoB-T (orange) is modeled at site 2 from the structural alignment of EpoBcy with AB3403 (19) that is shown in Fig. S5. Cavities are the approximate length of the Ppant prosthetic group (dashed orange and red lines). (*B*) Close-up view of active site channels from $P2_1$-EpoBcy structure. Locations of the EpoA-T and EpoB-T domains represented as in *A*. A Ppant arm is modeled in the site 2 channel (gray) based on the superimposition of EpoBcy with holo-AB3403, and a squiggly line represents the Ppant in the site 2 channel (green). 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 (gray) 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 *D*), leads to an opening of the channel such that this channel now resembles the channel of the $P2_1$ structure in *B* (residues from MD simulations are colored gray).

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* 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. 5*C*). 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. 5 *A* 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. 5 *A* and *B*). The DxxxxDxxS motif aligns well with the catalytic HHxxxDG motif of NRPS C domains as predicted (13, 16, 17) (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* 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.



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

Fig. 5. Identification of EpoBcy active site residues. (*A*) Polar residues 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. Channels colored as in Fig. 4*A* and ribbons colored as in Fig. 2*A*. Stereoview is shown in Fig. S7. (*B*) Same as *A*, but rotated ∼90° clockwise about the vertical axis to match orientation shown in Fig. 4*A*. (*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.

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 reengineering 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 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. 4*B* and S7*C*). Notably, the EpoA channel appears a bit longer than necessary for an acetyl moiety to be accommodated (Fig. 4 *A* and *B*), perhaps explaining the observation that larger substrates (propionyl-, isobutryl-, and benzoyl-EpoA-T) can be turned over by Cy, albeit more slowly (10).

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 toward 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 (20), 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. We find that the EpoB Cy domain adopts a similar protein fold as the NRPS C domain, validating previous predictions (15), but we also find 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 2,000-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, because 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 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 (13), 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, although there is still much to be learned about this fascinating cyclization chemistry.

Since the discovery of NRPS natural products, researchers have been interested in manipulating these modular assembly lines for the bioproduction of novel chemical compounds (6, 26, 28, 29). For this to be achievable and efficient, we must increase our understanding of how these systems function at the molecular level (28). 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 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

experiments, model building, and refinement are detailed in *SI Methods*. A 20-ns MD simulation was performed in GROMACS (30), and parameters are described in *SI Methods*.

ACKNOWLEDGMENTS. C.L.D. is a Howard Hughes Medical Institute Investigator. Support for this research was provided by National Institute of Environmental Health Sciences, NIH Core Center Grant P30-ES002109. This work is based on 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. NE-CAT at APS is supported by the National Institute of General Medical Sciences (NIGMS) at the NIH (P41 GM103403). APS is an Office of Science User Facility operated for the US Department of Energy (DOE) Office of Science by Argonne National Laboratory and is also supported by the US DOE under Contract DE-AC02-06CH11357. Use of SSRL is supported by DOE, Office of Science, Office of Basic Energy Sciences under Contract DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research and by NIGMS at the NIH (including P41GM103393).

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. Thompson CA (2007) FDA approves new breast cancer treatment. *Am J Health Syst Pharm* 64(23):2406.
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 Chem Int Ed Engl* 49(41):7504–7507.
5. Walsh CT, Nolan EM (2008) Morphing peptide backbones into heterocycles. *Proc Natl Acad Sci USA* 105(15):5655–5656.
6. Hur GH, Vickery CR, Burkart MD (2012) Explorations of catalytic domains in non-ribosomal peptide synthetase enzymology. *Nat Prod Rep* 29(10):1074–1098.
7. 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.
8. Julien B, et al. (2000) Isolation and characterization of the epothilone biosynthetic gene cluster from *Sorangium cellulosum*. *Gene* 249(1-2):153–160.
9. Tang L, et al. (2000) Cloning and heterologous expression of the epothilone gene cluster. *Science* 287(5453):640–642.
10. 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.
11. 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.
12. 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):17104.
13. 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.
14. Richter CD, Nietlispach D, Broadhurst RW, Weissman KJ (2008) Multienzyme docking in hybrid megasynthetases. *Nat Chem Biol* 4(1):75–81.
15. 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.
16. Marshall CG, Hillson NJ, Walsh CT (2002) Catalytic mapping of the vibriobactin biosynthetic enzyme VibF. *Biochemistry* 41(1):244–250.
17. 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.
18. Tanovic A, Samel SA, Essen LO, Marahiel MA (2008) Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science* 321(5889):659–663.
19. Drake EJ, et al. (2016) Structures of two distinct conformations of holo-non-ribosomal peptide synthetases. *Nature* 529(7585):235–238.
20. 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.
21. Samel SA, Schoenafinger G, Knappe TA, Marahiel MA, Essen L-O (2007) Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure* 15(7):781–792.
22. Samel SA, Czodrowski P, Essen L-O (2014) Structure of the epimerization domain of tyrocidine synthetase A. *Acta Crystallogr D Biol Crystallogr* 70(Pt 5):1442–1452.
23. Chen W-H, Li K, Guntaka NS, Bruner SD (2016) Interdomain and intermodule organization in epimerization domain containing nonribosomal peptide synthetases. *ACS Chem Biol* 11(8):2293–2303.
24. Liu F, Garneau S, Walsh CT (2004) Hybrid nonribosomal peptide-polyketide interfaces in epothilone biosynthesis: Minimal requirements at N and C termini of EpoB for elongation. *Chem Biol* 11(11):1533–1542.
25. Li Y, Weissman KJ, Müller R (2010) Insights into multienzyme docking in hybrid PKS-NRPS megasynthetases revealed by heterologous expression and genetic engineering. *ChemBioChem* 11(8):1069–1075.
26. Du L, Sánchez C, Shen B (2001) Hybrid peptide-polyketide natural products: Biosynthesis and prospects toward engineering novel molecules. *Metab Eng* 3(1):78–95.
27. O'Connor SE, Walsh CT, Liu F (2003) Biosynthesis of epothilone intermediates with alternate starter units: Engineering polyketide-nonribosomal interfaces. *Angew Chem Int Ed Engl* 42(33):3917–3921.
28. Khosla C, Herschlag D, Cane DE, Walsh CT (2014) Assembly line polyketide synthases: Mechanistic insights and unsolved problems. *Biochemistry* 53(18):2875–2883.
29. Strieker M, Tanović A, Marahiel MA (2010) Nonribosomal peptide synthetases: Structures and dynamics. *Curr Opin Struct Biol* 20(2):234–240.
30. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4(3):435–447.
31. Bond CS (2003) TopDraw: A sketchpad for protein structure topology cartoons. *Bioinformatics* 19(2):311–312.
32. Ho BK, Gruswitz F (2008) HOLLOW: Generating accurate representations of channel and interior surfaces in molecular structures. *BMC Struct Biol* 8:49.
33. Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248–254.
34. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
35. Terwilliger TC, et al. (2009) Decision-making in structure solution using Bayesian estimates of map quality: The PHENIX AutoSol wizard. *Acta Crystallogr D Biol Crystallogr* 65(Pt 6):582–601.
36. Vonrhein C, Blanc E, Roversi P, Bricogne G (2007) Automated structure solution with autoSHARP. *Methods Mol Biol* 364:215–230.
37. Emsley P, Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126–2132.
38. Adams PD, et al. (2010) PHENIX: A comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213–221.
39. McCoy AJ, et al. (2007) Phaser crystallographic software. *J Appl Cryst* 40(Pt 4):658–674.
40. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL (2001) Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J Phys Chem B* 105(28):6474–6487.
41. Berendsen HJC, Grigera JR, Straatsma TP (1987) The missing term in effective pair potentials. *J Phys Chem* 91(24):6269–6271.
42. Essmann U, et al. (1995) A smooth particle mesh Ewald method. *J Chem Phys* 103(19): 8577–8593.
43. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1):014101.
44. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *J Appl Phys* 52(12):7182–7190.
45. Winn MD, et al. (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* 67(Pt 4):235–242.
46. Sievers F, et al. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7(539):539.