jmb Article

# Crystal Structures of the First Condensation Domain of CDA Synthetase Suggest Conformational Changes during the Synthetic Cycle of Nonribosomal Peptide Synthetases

Kristjan Bloudoff[1], Dmitry Rodionov[1,2] and T. Martin Schmeing[1,2]

*1 - **Department of Biochemistry,** McGill University, Montréal, QC, Canada H3G 0B1*
*2 - **Groupe de Recherche Axé sur la Structure des Protéines (GRASP),** McGill University, Montréal, QC, Canada H3G 0B1*

***Correspondence to T. Martin Schmeing:** Department of Biochemistry, McGill University, Montréal, QC, Canada H3G 0B1. martin.schmeing@mcgill.ca*
http://dx.doi.org/10.1016/j.jmb.2013.06.003
***Edited by G. Schulz***

## Abstract

Nonribosomal peptide synthetases (NRPSs) are large modular macromolecular machines that produce small peptide molecules with wide-ranging biological activities, such as antibiotics and green chemicals. The condensation (C) domain is responsible for amide bond formation, the central chemical step in nonribosomal peptide synthesis. Here we present two crystal structures of the first condensation domain of the calcium-dependent antibiotic (CDA) synthetase (CDA-C1) from *Streptomyces coelicolor*, determined at resolutions 1.8 Å and 2.4 Å. The conformations adopted by CDA-C1 are quite similar in these two structures yet distinct from those seen in other NRPS C domain structures. HPLC-based reaction assays show that this CDA-C1 construct is catalytically active, and small-angle X-ray scattering experiments suggest that the conformation observed in these crystal structures could faithfully represent the conformation in solution. We have performed targeted molecular dynamics simulations, normal mode analyses and energy-minimized linear interpolation to investigate the conformational changes required to transition between the observed structures. We discuss the implications of these conformational changes in the synthetic cycle and of the observation that the "latch" that covers the active site is consistently formed in all studied C domains.



## Introduction

Nonribosomal peptide synthetases (NRPSs) are large macromolecular machines that catalyze the assembly of monomer substrates into biologically active secondary metabolites.[1–3] As the name implies, NRPS substrates are often amino acids, but over 400 monomers are known to be used as substrates, including D-amino acids, aryl acids, keto acids, hydroxy acids and fatty acids.[4] Nonribosomal peptides have important and diverse biological activity and include antifungals, antibacterials, antivirals, antitumors, siderophores and immunosuppressants,[3] including well-known compounds such as penicillin,[5] daptomycin[6] and cyclosporin.[7]

NRPSs are organized into modules of >110 kDa, with each module responsible for the addition of one specific monomer. Modules contain multiple domains, each performing specific functions in product synthesis.[8] A basic elongation module contains a condensation (C) domain, an adenylation (A) domain and a peptide carrier protein (PCP) domain. The A domain selects the cognate amino acid and adenylates it, then attaches it to a phosphopantetheinyl (PPE) group on the PCP domain. The PCP domain transports the amino acid to the C domain, which catalyzes peptide bond formation between this amino acid and the peptide attached to the PCP domain of the preceding module, thus elongating the peptide chain. Next, the PCP domain brings the

elongated peptide chain to the downstream module, where it is passed off and further elongated in the next peptidyltransferase reaction.

In this reaction cycle, numerous conformational changes are known or have been proposed to occur. As described above, the PCP domain must completely translocate to interact with different partner domains[9,10] and is also known to change conformation, depending on its functional state.[11] Large-scale movements are known to occur in the A domain between the conformation that catalyzes the adenylation reaction and the conformation that catalyzes the thiolation reaction (attachment of substrate to PCP domain).[12–15] Thus, conformational cycling is absolutely essential for NRPS function.

The C domain catalyzes the key catalytic event of NRPS function, peptide (amide) bond formation.[16,17] Three structures that include NRPS C domains have been determined by X-ray crystallography: a stand-alone C domain,[18] a C–PCP didomain complex[19] and a C-A-PCP-Te termination module.[10] The C domain comprises ~450 amino acids and has a pseudodimer configuration, with both N- and C-terminal subdomains having cores with folds in the coenzyme-A-dependent acyltransferase (CAT) superfamily. The active site sits at the bottom of a "canyon"[19] or "V"[10] formed by the two subdomains of the C domain and is covered by a "latch" that crosses over from C to N subdomain. The catalytic center includes an HHxxDG sequence motif[16,17,20] and must have binding sites for its donor and acceptor substrates. The conformations of the C domain visualized in these three structures vary somewhat, though it is unclear whether these differences stem from the fact that different proteins were used in the three studies or from the possibility that they are in different functional states. Although there are hundreds of different C domains in NRPS biosynthetic clusters, they likely all share a common mode of action.

In this study, we focus on the first C domain of the calcium-dependent antibiotic (CDA) synthetase, CDA-C1 (Fig. 1).[21,22] The CDA is part of a class of acidic lipopeptides that includes the last resort antibiotic daptomycin.[23] These antibiotics work by binding to and disrupting cytosolic membranes in bacteria.[24] The CDA biosynthetic cluster in *Streptomyces coelicolor* includes an 11-module canonical NRPS spread over 3 proteins, which adds the 11 amino acids and cyclizes the product (Fig. 1a).[22] The first monomer is not an amino acid but a fatty acid, a 2,3-epoxyhexanoyl group that is synthesized as a hexanoyl fatty acid on an acyl carrier protein (ACP) by a fatty acid synthase, then modified to the epoxy form by epoxidation enzymes HxcO and/or HcmO (Fig. 1b).[25,26] The first C domain of CDA synthetase catalyzes the transfer of the 2,3-epoxyhexanoyl group from 2,3-epoxyhexanoyl-ACP to the serinyl-PCP domain of the first module. Although normally present as part of the 799-kDa CDA PS1, excised CDA-C1 is active in catalysis.[26]

Here we present two structures of the first condensation domain of the CDA synthetase (CDA-C1), determined by X-ray crystallography at resolutions 1.8 Å and 2.4 Å, and accompanying small-angle X-ray scattering (SAXS), activity assays and computational analyses.

## Results and Discussion

### Purification, crystallization and structure determination of CDA-C1

CDA-C1 is the N-terminal domain of CdaPS1, a 20-domain, 7463-residue NRPS protein.[26] To study CDA-C1, we designed a gene construct by aligning the sequences of CDA-C1 to C domains with known structure to enable C-terminal boundary definition. The resulting construct was heterologously expressed in *Escherichia coli* with extremely high yields and purified to homogeneity using a four-column chromatographic protocol. CDA-C1 was subjected to high-throughput crystallization trials, which readily yielded crystals under several crystallization conditions. Two crystallization conditions were optimized and could be used reproducibly to produce crystals in $P2_12_12_1$ and $P2_1$ space groups that diffract to high resolution. Data collection at a rotating anode "home" source gave complete data sets with good statistics (Table S1). These were subjected to phasing trials using molecular replacement with many different search models derived from one subdomain or both subdomains of the known C domains, with no success. Selenomethionine-derivatized (SeMet) protein was then produced, which could be purified and crystallized by the same protocols. Multiwavelength data sets were collected from $P2_12_12_1$ crystals at the National Synchrotron Light Source (NSLS), and multiwavelength anomalous dispersion phasing techniques were used to determine the structure to 1.8 Å resolution (Fig. 2, Fig. S2 and Table S1). Electron density maps showed that there are two molecules in asymmetric unit with very similar conformation and allowed building of residues 4–449 in molecule A and residues 2–448 in molecule B. The use of this structure as a search model readily gave a molecular replacement solution for the $P2_1$ crystal form using home source data with resolution solved to 2.4 Å (Fig. S2b and Table S1). This final model included residues 4–222 and 234–449 in molecule A and residues 5–230, 241–410 and 420–448 in molecule B.

### The structure of CDA-C1

CDA-C1 adopts the classical C domain configuration,[10,18,19] with two subdomains each

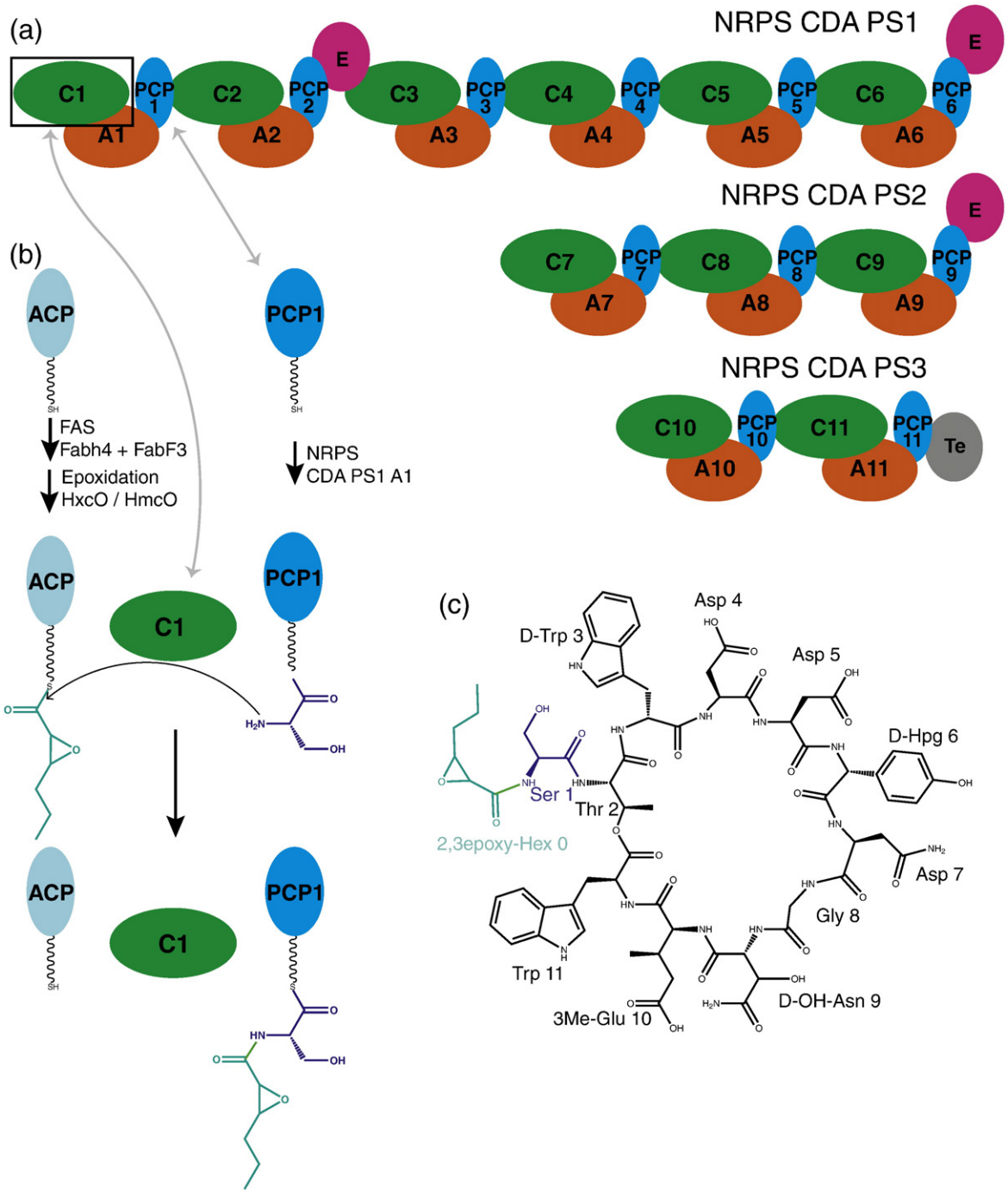



**Fig. 1.** Schematic diagrams for CDA components. (a) A schematic diagram of the proteins CDA peptide synthetase 1, 2 and 3 that make up the NRPS system for CDA synthesis in *S. coelicolor*. (b) A schematic diagram illustrating the reaction catalyzed by CDA-C1. (c) A schematic diagram of the chemical structure of CDA-4b, a representative CDA peptide synthesized in *S. coelicolor*.[4,26]

having a core CAT fold, two points of crossover from the C-terminal subdomain to the N-terminal subdomain, and the active site in the center of the "V" formed by the subdomains (Fig. 2). Superimposition of all four independent molecules determined here (two molecules from the $P2_12_12_1$ crystal form and two from the $P2_1$ crystal form) shows that, in all these molecules, the subdomains adopt approximately the same relative orientation to one another (Fig. S3). The only major difference between these structures is in the conformation of the loop 82–96. Loop 82–96 is in the linker between two β sheets of the

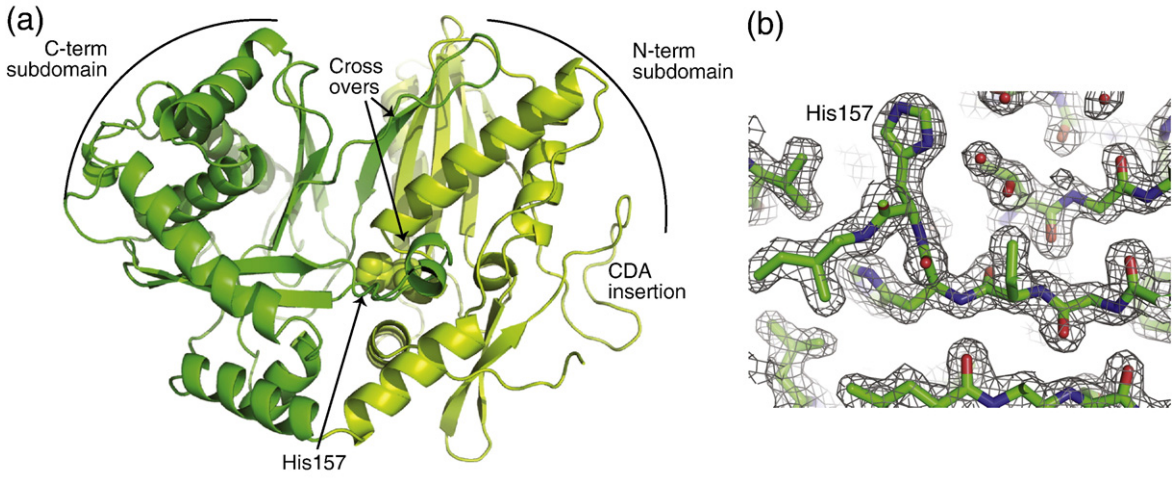



**Fig. 2.** Structure of CDA-C1. (a) Ribbon representation of CDA-C1 determined in the $P2_12_12_1$ space group. The N-terminal subdomain is chartreuse; the C-terminal subdomain is green; and the domain crossovers, CDA-specific insertion and active site H157 are all indicated. (b) An $2F_o - F_c$ electron density map contoured at 1σ.

N-terminus. Multiple alignment of over 500 C domains[27] shows that this loop is quite unique to CDA-C1, with only a protein 930752.1 from *Photorhabdus luminescens* also having an insertion in the same location (albeit with no conservation between the two sequences) (Fig. S1). Since CDA-C1 and Plu 930752 are initiator C domains, this face may have less stringent conservation requirements: the C domain does not have to pack with a full upstream module; rather, it needs to interact only with the small ACP proteins.

Comparison of the structures presented here with the structures of the three previously determined structures of C domains reveals significant conformational differences. When the C-terminal subdomains of the characterized C domains from tyrocidine synthetase III (TycC),[19] vibriobactin synthase (VibH)[18] and surfactin A synthetase C (SrfAC)[10] are superimposed upon CDA-C1, the N-terminal subdomains of these proteins are in strikingly different positions (Fig. 3 and Fig. S4). CDA-C1 is in a much more "closed" conformation, with its N- and C-terminal subdomains closer together. The N-terminal subdomain would require a shift and large rotation of ~15°, ~22° and ~25°, respectively, to assume the conformations seen with VibH, TycC and SrfAC. This is a large movement—the $C_\alpha$ positions in the N-terminal domain move by as much as 19, 20 and 23 Å and on average differ by 6, 8 and 11 Å.

## SAXS analysis of CDA-C1

This "closed" conformation of CDA-C1 seemed unlikely to be artificially forced by crystal packing. All

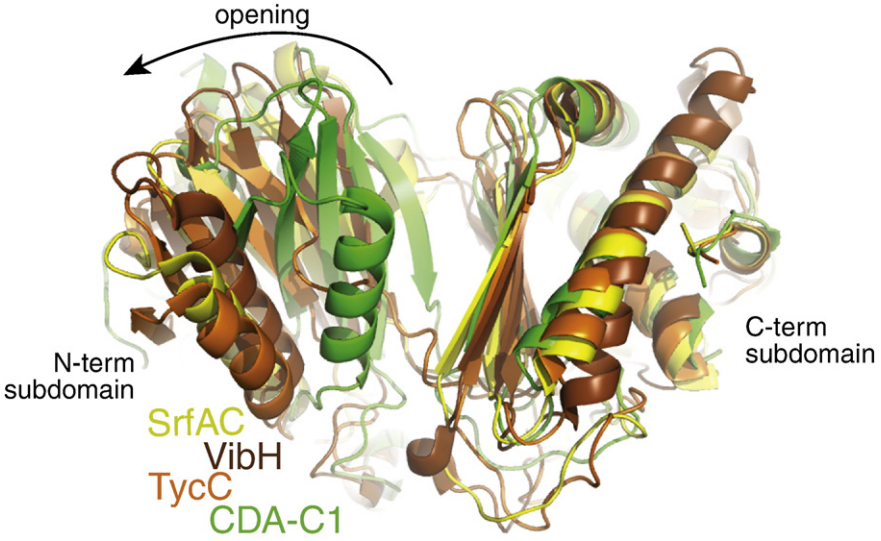



**Fig. 3.** Comparison of CDA-C1 with other C domains. Alignment of the C-terminal subdomain shows different relative subdomain–subdomain orientation in the structures of C domains. CDA-C1 displays the most "closed" conformation. CDA-C1 is in green, the C domain from SrfAC is yellow, the C domain from VibH is brown and the C domain from TycC is orange. See Fig. S4 for individual comparisons.

CDA-C1 monomers are observed in very similar conformations, even though the packing contacts between the two molecules in each asymmetric unit (crystallographic dimers) are solely through the C-terminal subdomain, and many of the other crystal contacts are different for the two space groups (Fig. S3d). To address whether this "closed" conformation is indeed adopted in solution, we undertook a SAXS analysis of CDA-C1.

SAXS data were collected on CDA-C1 at multiple concentrations with each experiment giving consistent results (a typical scattering result is shown in Fig. 4). Molecular mass of CDA-C1 in solution was found to be ~97 kDa using the Kratky plot method (comparing to bovine serum albumin and lysozyme) that agrees well with calculated molecular mass of CDA-C1 dimer (96.8 kDa) (Fig. S5). We calculated theoretical scattering curves from the structures of both a single CDA-C1 and the dimer of CDA-C1 observed in both $P2_12_12_1$ and $P2_1$ crystal forms. Comparison of the scattering data and the theoretical curves obviously indicates that CDA-C1 is a dimer in solution and that the dimer is in a conformation very similar to that of the crystallographic dimer (Fig. 4a). The envelope calculation also shows a good consistency between our crystallographic dimer and the solution conformations as assayed by SAXS (Fig. 4b). Although CDA-C1 appears to exist as a dimer in solution and in the crystal and this dimer interface is accessible in the SrfAC termination module,[10] the dimer interface does not show any sequence conservation among C domains so may not have biological significance in the intact NRPS.

To test whether the SAXS data could differentiate between the C domain conformation observed in the crystal structure reported here and those conformations seen in the previously reported structures (Fig. 3 and Fig. S4), we created models of CDA-C1 in the conformations of VibH, TycC and SrfAC. These models were used to calculate theoretical scattering curves that were fitted to the experimental SAXS scattering data, with goodness of fit assessed by calculating $\chi^2$ values. CDA-C1 had the lowest $\chi^2$ to the SAXS data (Fig. 4c), although the values for VibH and SrfAC were only slightly higher. TycC had a substantially higher $\chi^2$. Thus, the conformation we observe in both $P2_12_12_1$ and $P2_1$ crystal forms is at least consistent with solution studies.

## CDA-C1 is catalytically active

We next asked whether the construct of CDA-C1 we made, which adopts this "closed" conformation, is catalytically active. We first attempted to show activity of our construct using small molecule acyl-*N*-acetylcysteamine thioester analogues (acyl-SNACs)[28] of the two substrates (2,3-epoxy-hexanoyl-SNAC and serinyl-SNAC), but we were unable to detect product (2,3-epoxy-hexanoyl-serinyl-SNAC) formation. While we were performing our studies, Kraas *et al*. published experiments with a similar construct of CDA-C1, which showed their construct to be active in an assay where the two substrates are delivered by carrier proteins and the product 2,3-epoxy-hexanoyl-serinyl-PCP is separated by HPLC and confirmed by mass spectrometry.[26] We could demonstrate catalytic activity of our CDA-C1 construct using a version of this assay (Fig. 5). The assay is complicated by the fact that serinyl-PCP (and, in our experience, 2,3-epoxy-hexanoyl-ACP) is liable to hydrolysis and that several PCP species migrate very similarly on reverse phase columns. Nonetheless, we were able to clearly identify product formation when CDA-C1 was in the reaction mix, but not when it was replaced by CDA-C1 harboring an H157A active-site mutation (Fig. 5).[26]

Thus, the CDA-C1 construct we crystallized is catalytically active and the conformation we observe in the crystal structures is consistent with SAXS results probing its solution conformation.

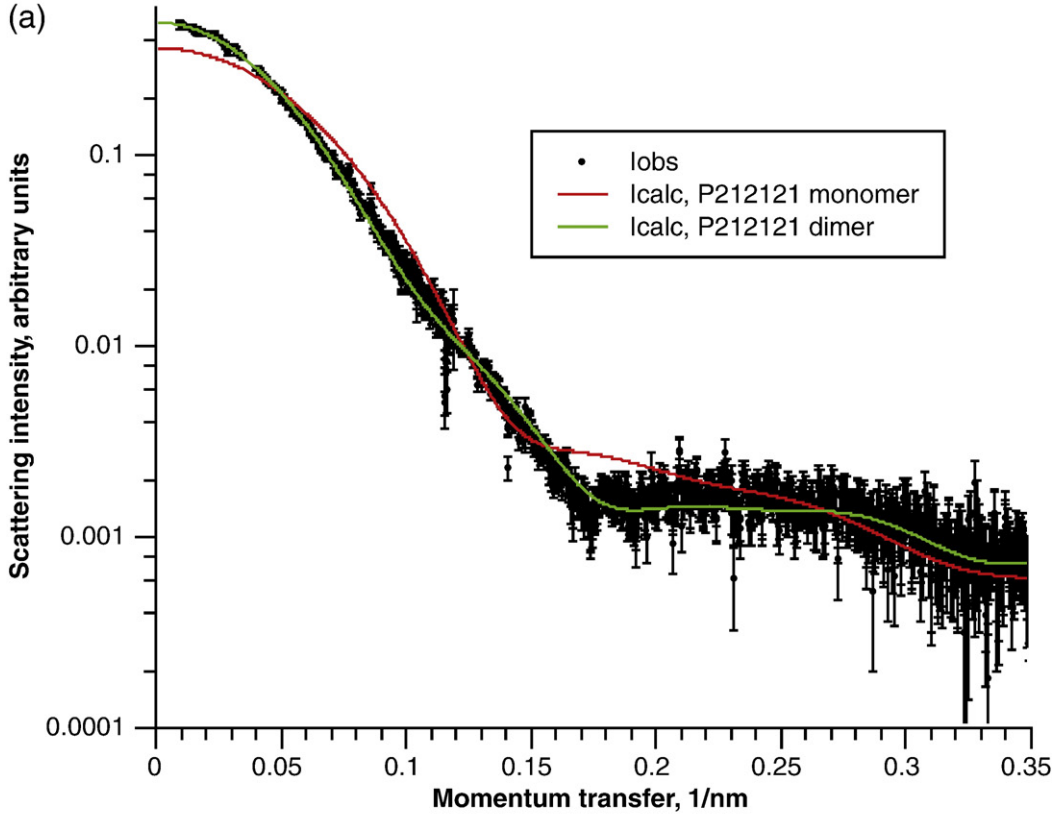
## Computational analyses of C domain movement

Although each structurally characterized C domains is seen in a different conformation, there appears to be sufficient space at the active site to bind substrate in each form. One could imagine that progressively more open C domain conformations would occur in C domains from later modules of NRPS to accommodate larger substrates, but that is unlikely to be the cause of the different conformations: VibH is, like CDA-C1, an initiator C domain but is found in an open conformation, and SrfAC is more open than TycC despite coming from an earlier module in their respective NRPS (Fig. 3, Fig. S4 and Movie S7). It is unclear if all these conformations are catalytically active or if a C domain would sample each of the observed conformations during an NRPS catalytic cycle. There is extensive precedent for using crystal structures of homologous proteins to investigate conformational changes that a protein may undergo in solution.[29–32] Therefore, we undertook analyses of the types of movement required to transition between observed conformations.

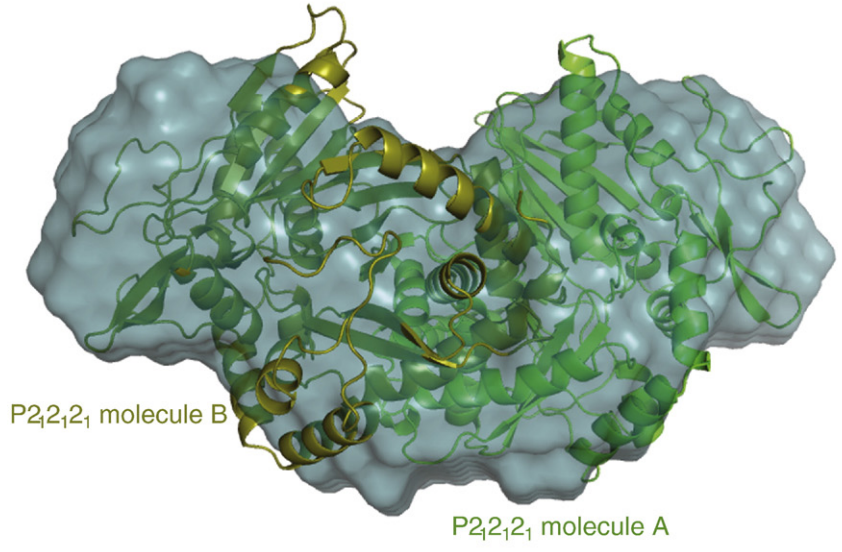
We subjected CDA-C1 to normal mode analysis to observe whether a transition from "closed" to "open" form would be replicated by normal modes, using the programs WEBnm@[33] and NOMAD-Ref.[34] Both programs reveal that such a movement is reasonable. WEBnm@ replicates this movement very well (Movie S1), although in this simulation, the amplitude is not quite enough to bring it to a fully "open" state. Likewise, several NOMAD-Ref modes appear to replicate similar movements (Movies S2 and S3).

To study and visualize the full transition between closed and open states as well as between the various open states, we carried out two further computational exercises. First, we performed linear

(a)




(b)




(c)

| PDB model | $\chi^2$ |
|---|---|
| CDA-C1 conformation (dimer) | 1.099 |
| VibH conformation (dimer) | 1.132 |
| SrfAC conformation (dimer) | 1.139 |
| TycC conformation (dimer) | 2.041 |
| CDA-C1 conformation (monomer) | 4.522 |

**Fig. 4** (*legend on next page*)

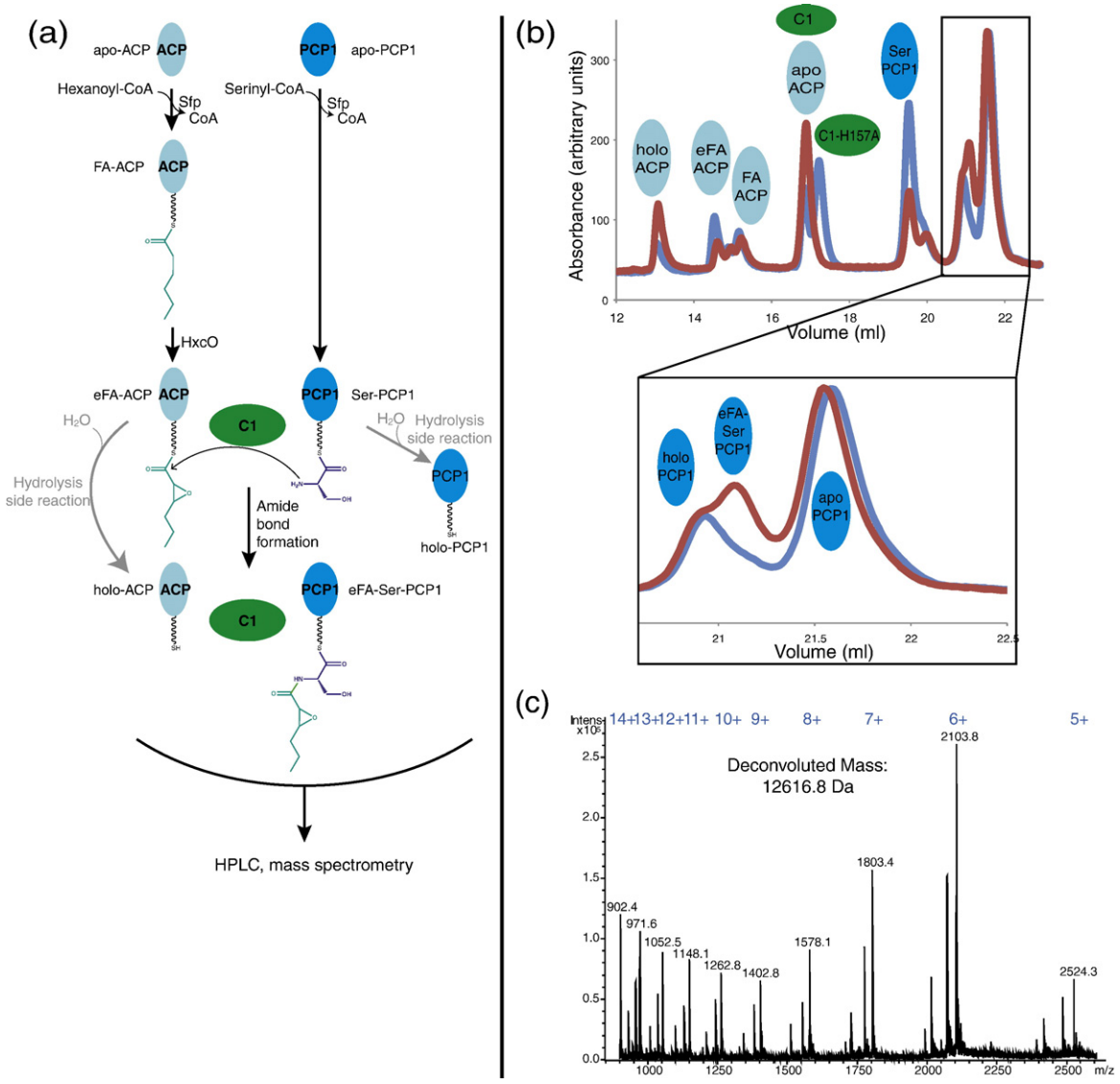



**Fig. 5.** CDA-C1 is catalytically active. (a) A schematic diagram explaining the C domain activity assay. (b) HPLC traces and (c) mass spectra of reaction assays show that product epoxyl-hexanoyl-serinyl-PCP1 is formed only when wild-type CDA-C1 (blue) is present in the reaction mixture. Reactions with CDA-C1 harboring the active-site mutation H157A (red) do not lead to product formation. Expected mass of epoxyl-hexanoyl-serinyl-PCP1 is 12,618.3 Da. Expected and observed masses for other carrier protein states are listed in Table S2.

interpolation with energy minimization refinement ("morphing") to interconvert the conformations. The resulting animations show the transition from the conformation seen in this study to those of previously published structures (Movies S4–S6), as well as the movements required to transition between these states (Movie S7). Second, we performed targeted molecular dynamics (MD) simulations on the same transitions (Movies S8–S11). The targeted MD simulation transitions appear smooth and do not seem to have to pass through any obviously unfavorable conformations. The root-mean-square

**Fig. 4.** SAXS analyses of CDA-C1. (a) Fitting of calculated scattering curves to merged experimental data indicating that CDA-C1 is a dimer in solution. (b) Superposition of averaged filtered envelope on crystallographic CDA-C1 dimer. (c) Quality of fit of various models to experimental scattering data suggesting that CDA-C1 in solution is best described by crystallographic dimer in its observed conformation.

deviations between progressive models in the simulations and the target models in SrfAC, TycC and VibH conformation smoothly decrease for residues in the N-terminal subdomain and crossover strands (Fig. S4).

Together, the normal mode analyses, the morphing and targeted MD simulations suggest that the conformation changes that would be required to interconvert the four conformations observed in NRPS C domains are reasonable.

The putative conformational changes described here involve relative movement between the two CAT fold containing subdomains, while the conformation of each CAT core fold remains unchanged. There are numerous proteins that contain CAT folds, three of which have been characterized structurally and shown to also contain two "pseudodimeric" CAT folds: murine carnitine acetyltransferase (CrAT),[35] human choline acetyltransferase (CHAT)[36] and polyketide-associated protein A5 (PapA5; which has a similar overall structure to C domains).[37] We asked whether similar opening has been seen in these three proteins. CrAT does not undergo significant conformational changes upon binding substrate[35]; CHAT undergoes small but significant conformational change (~1.5° opening) upon substrate binding,[36] and PapA5 is hypothesized to undergo some kind of opening of the active site to allow substrate to bind, as an α helix blocks the active site in the apo conformation, but this opening has yet to be observed.[37] Therefore, the opening of two CAT fold subdomains is a feature in some, but not all, proteins that contain them.

Though some C domains are expressed at stand-alone proteins, they are more usually part of a much larger protein; thus, the conformational changes must be considered in the context of a full NRPS. Within a module, an outside face of the C-terminal subdomain of the C domain forms a large binding interface with the A domain.[10] The open or closing of the C domain would not abolish this C–A interface, though subtle changes within the contact area might occur. We speculate that conformational changes such as the opening of the C domain may be part of a communication network in NRPSs by which the functional state of one domain is conveyed to other domains. This may include sensing of the $A_{sub}$ domain position, which forms part of the C–A interface in the adenylation state but rotates away in the thiolation state,[10,12–15] or PCP domain position, which has been observed bound at the acceptor site of an open C domain.[10] Such a communication network could contribute to appropriate timing and coordination of the many reactions in the NRPS synthetic cycle, increasing the efficiency and rate of small molecule synthesis.[38,39] Confirmation of this theoretical network would require multiple structures of intact NRPSs and accompanying biochemical analyses.

## The "latch" of the C domain

There are two points of crossover where a segment from the C-terminal subdomain crosses over to form part of the N-terminal subdomain. Residues ~295–309 crossover and form a small α helix, which backs against N-terminal subdomain helices h3 and h4. Residues ~367–388 cross and donate a β strand to the major β sheet in the N-terminal subdomain (formed of strands s1, s4, s5 and s6) (Fig. 2 and Fig. S7a). This segment has been described as a "latch", forming a "roof" of the active site, and has been proposed to disengage from the N-terminal subdomain during the reaction cycle.[19] However, in all structures of C domains, the crossover "latch" interaction is intact (Fig. S7). Furthermore, the interaction remains intact throughout every normal mode analysis, MD and morph simulation performed here (Movies S1–S11). Buried surface area calculations suggest that it is possible that the latch, at least in some C domains, could remain intact: In CDA-C1, SrfAC, TycC and VibH, the latch buries 954 Å$^2$, 903 Å$^2$, 641 Å$^2$ and 927 Å$^2$ of surface area, respectively, where interface on the order of 700 Å$^2$ is known to support heterodimer formation.[40,41] Finally, Samel *et al.* argue that high *B*-factors in the loop suggest that it could be a mobile element but that the latch residues in our structures display slightly lower than average *B*-factors (average *B*-factor for latch, 20.8 Å$^2$; average *B*-factor for structure, 24.1 Å$^2$).[19] Conceptually, it seems reasonable that the latch would open because if it did not, the growing peptide chain would need to be threaded through each C domain active site. However, we do not feel that there is at this time compelling evidence to support latch opening.

## An active-site tunnel and transition-state model

As no structure of a C domain with substrates has been determined, the precise binding site and approach of the PPE-bound substrates are not definitively known but can be reasonably guessed.[10,18,19] The donor and acceptor substrates must approach and bind from opposite faces and meet at the active site H157. If the latch is not open, as is the case in all structures and models to date, the active site can be described as being at the center of a tunnel in the middle of the C domain (Fig. 6). We identified this tunnel visually and also by using the program CAVER.[42] This tunnel, formed partially by the latch, stays intact through the morphing and MD simulations and in all crystal structures. The tunnel is ~30 Å long and wide enough to accommodate substrates. Indeed, the tunnel entrances correspond to what is known about the binding sites for the upstream and downstream carrier protein domains. The SrfAC module has a

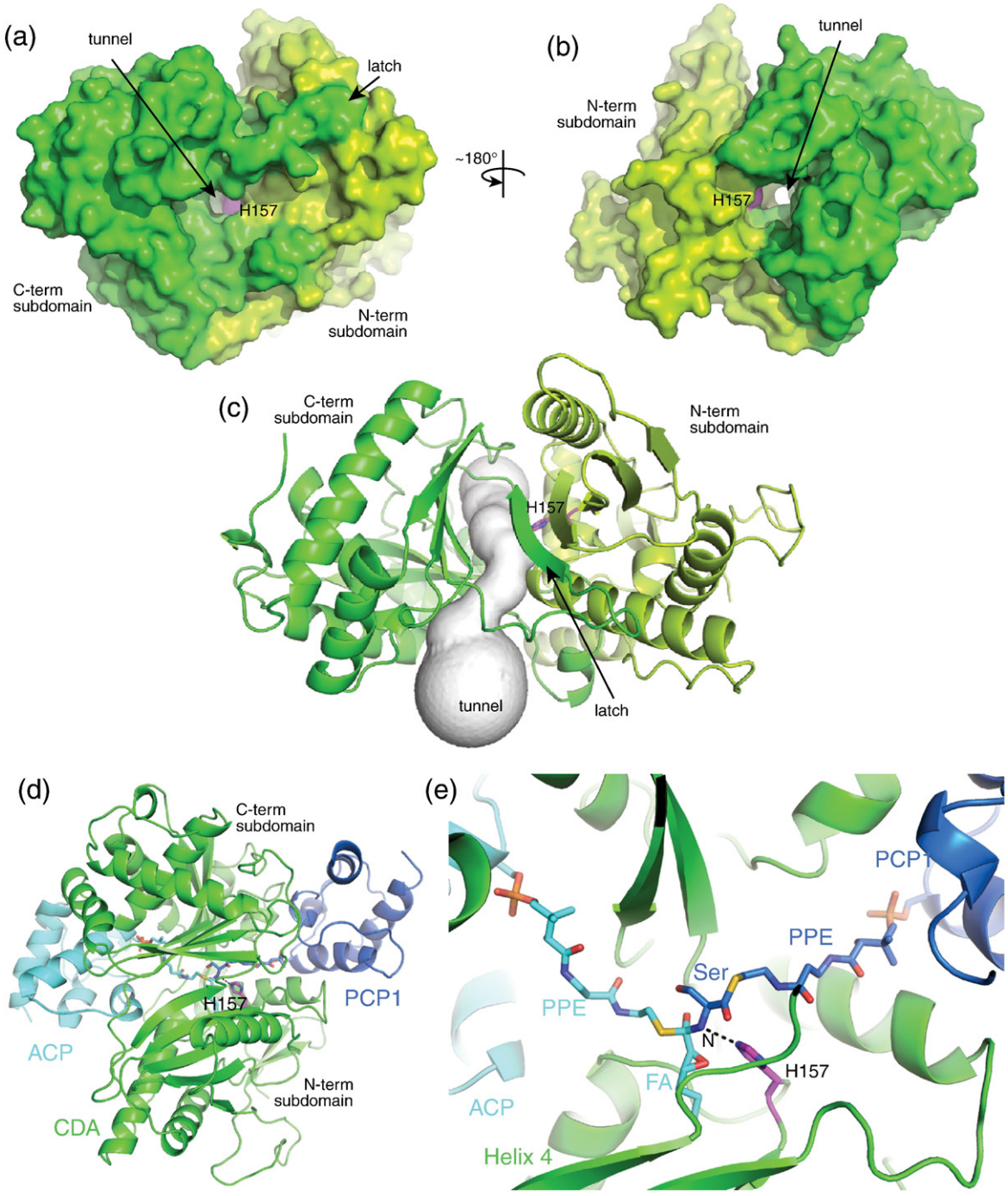



**Fig. 6.** A tunnel to the active site of CDA-C1. (a and b) Surface representations of CDA-C1 show a tunnel through CDA-C1, with the active site in its center. (c) A representation of the tunnel from CAVER[42] output. (d and e) A model of the transition state of the CDA-C1-catalyzed reaction including CDA-C1, ACP, PCP1 and covalently attached transition state. Note that the binding sites of upstream and downstream carrier proteins are not related by the same symmetry operation that relates the CAT core of the N- and C-terminal subdomains. The CAT cores are related by an ~120° rotation about an axis approximately parallel with the acceptor portion of the tunnel. Since the tunnel has a carrier protein at each end, rotation of the entire modeled complex about this axis does not interchange the positions of the carrier proteins.

substrate-less PCP domain positioned with its PPE attachment site at a reasonable distance from the C domain's catalytic histidine and positioned at the tunnel entrance shown in Fig. 6b.[19] The structure of the TycC didomain has an upstream PCP domain in an unproductive conformation, but it is generally

positioned near the tunnel entrance shown in Fig. 6c, showing that the upstream carrier protein domain can reach this site. To give a holistic view of the C domain at the point of NRPS peptide bond formation, we constructed a model of the transition state of the reaction, including the C domain, 2,3-epoxy-hexenoyl-ACP and serinyl-PCP domain (Fig. 6). The starting position of the PCP domain was taken from the SrfAC model while the ACP was docked, using the program HADDOCK, taking into account the residues on the carrier proteins required for productive donor binding data.[43–45] The pantetheinyl substrates were roughly positioned along the tunnel, and the whole complex was subject to multiple rounds of Cartesian coordinate energy minimization in the program CNS.

The model fits exactly into the described tunnel, with the nucleophilic α amino group of the serinyl-PCP domain modeled at 3 Å distance from the putative catalytic residue, as would be expected for the intermediate state of the reaction. The fatty acid 2,3-epoxy hexanoyl side chain fits nicely into a pocket lined by α helix 4 and the major β sheet in the N-terminal subdomain, whereas the serine faces the C-terminal subdomain. Although precise analysis of substrate–C domain interactions awaits successful co-complex determination, this model will be useful in efforts to dissect the substrate specificity shown by C domains.[26,28,46–49]

In summary, we have presented a structure of an active, previously underdetermined NRPS condensation domain, which adopts a novel conformation. Conformational changes in C domains such as those modeled in our computational analyses are likely to occur in the catalytic cycle of NRPS C domains and may be important for peptide synthesis.

## Methods

### Cloning and expression of CDA-C1

The CDA-C1 construct was designed by aligning the sequence of first C domain of CDA peptide synthetase I of *S. coelicolor* A3(2) (NCBI NP_627443.1) to C domains of known structure (Fig. S1).[50] The CDA-C1 construct was synthesized by DNA 2.0, Inc. (Menlo Park, CA, USA), featuring an N-terminal octahistidine tag and tobacco etch virus (TEV) protease cleavage site of sequence MHHHHHHHHENLYFQG and CDA residues Met1-Thr449. Protein production was performed in BL21 (DE3) *E. coli* cells grown in LB medium supplemented by 300 μg $ml^{-1}$ kanamycin (LB-kan) at 37 °C. CDA-C1 expression was induced at an $OD_{600}$ of ~0.5–0.6 with the addition of 1 mM IPTG, and expression continued overnight at 16 °C. SeMet-CDA-C1 was expressed in the same cell strain grown in kanamycin-containing M9 minimal medium supplemented with 0.2% glucose, 2 μM $MgSO_4$ and 0.1 μM $CaCl_2$. After reaching an $OD_{600}$ of 0.5, the medium was further supplemented with the amino acids K, F, T (100 mg $l^{-1}$ each) and I, L, V (50 mg $ml^{-1}$ each) and seleno-L-methionine (60 mg $l^{-1}$).[51] Fifteen minutes after supplementation, cultures were induced with 1 mM IPTG and grown overnight at 25 °C.

### Protein purification of CDA-C1

CDA-C1 cell pellets were resuspended in a buffer of 50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 2 mM β-mercaptoethanol (βME), 1 mM imidazole (pH 8.0) and 1 mM PMSF. Cells were lysed by sonication and centrifuged for 30 min at 40,000***g*** and 4 °C. The supernatant was pooled and applied onto a 5-ml HiTrap IMAC FF column (GE Healthcare) charged with $Ni^{2+}$. CDA-C1 was eluted with 50 mM Tris–HCl (pH 7.5), 100 mM NaCl, 2 mM βME, 250 mM imidazole (pH 8.0) and 1 mM PMSF; purity was assessed using SDS-PAGE. The fractions containing CDA-C1 were diluted 5-fold with a buffer of 50 mM Tris–HCl (pH 7.5) and 2 mM βME and were applied to a MonoQ 5/50 GL column (GE Healthcare) and eluted with a gradient of 0–1 M NaCl. Relevant fractions were pooled and digested overnight at 4 °C with N-His-TEV protease, using a ratio of 1 mg N-His-TEV protease per 40 mg CDA-C1. The cleaved sample was reapplied onto the 5-ml HiTrap IMAC FF column charged with $Ni^{2+}$, and the flow through was collected. The sample was brought to 1 M $(NH_4)_2SO_4$, applied to 2× of 1 ml HiTrap phenyl HP column (GE Healthcare) and eluted with a gradient of 1–0 M $(NH_4)_2SO_4$. After dialysis into 50 mM Tris–HCl (pH 7.5), 100 mM NaCl and 2 mM dithiothreitol, CDA-C1 was concentrated to 10 mg $ml^{-1}$, flash-frozen and stored in liquid nitrogen.

### Crystallization and data collection

Sparse-matrix crystallization trials of CDA-C1 and subsequent optimization identified two crystallization solutions that allowed growth of diffraction-quality crystals in sitting-drop format using 2 μl of 10 mg $ml^{-1}$ protein sample and 2 μl of crystallization solution in the drop and a 400 μl in the reservoir: (a) 25–27% polyethylene glycol (PEG) 3000, 0.2–0.25 M lithium sulfate and 0.1 M Hepes (pH 7.5) and (b) 9–13% PEG 10,000 and 0.1 M Hepes (pH 7.1–7.3). Crystals in condition (a) were directly mounted in cryoloops and flash-cooled in liquid nitrogen, while crystals in (b) were cryoprotected in a solution containing 20% PEG 10,000 prior to mounting. SeMet-CDA-C1 crystals were grown in 26–31% PEG 3000, 0.2–0.26 M lithium sulfate and 0.1 M Hepes (pH 7.5).

Data from SeMet-CDA-C1 crystals were collected at Se peak, inflection and remote wavelengths at X6A beamline at the NSLS, Brookhaven National Laboratory, Brookhaven, New York, with a 0.5° frame width (Table S1). Native data sets were collected using a Rigaku RUH-3R rotating copper-anode source equipped with a R-AXIS IV++ image plate detector at McGill University, Montreal, Canada (GRASP) with a 1° frame width.

### Structure determination

Data were indexed and scaled using the program *HKL-2000*.[52] The structure of SeMet-CDA-C1 was

determined by a three-wavelength multiwavelength anomalous dispersion phasing experiment using the program PHENIX[53] (Table S1). The structure of CDA-C1 in the crystal form grown under a condition containing PEG 10,000 was determined by molecular replacement with SeMet-CDA-C1 as the reference model, using the program Phaser.[54] Structures were subjected to iterative rounds of modeling and refinement to give the final model (Table S1).

### SAXS analyses

Purified CDA-C1 was subjected to gel filtration on Superdex200 10/300 GL column equilibrated with 25 mM 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol (pH 7.0), 150 mM NaCl and 10% glycerol. The most concentrated fraction (0.843 mg $ml^{-1}$) was used unaltered for SAXS data collection. The remaining fractions were pooled and concentrated on an Amicon(R) Ultra concentrator with a 10-kDa-cutoff membrane to produce samples at 3.48, 9.82 and 23.04 mg $ml^{-1}$. Samples were filtered through a Millipore Ultrafree-MC VV 0.1-µm filter prior to loading into SAXS cell. SAXS data were collected on an Anton Paar SAXSess MC2 CCD system on a PANalytical PW3830 generator with a Cu LFF tube (GRASP). Beam was collimated to 8 mm length. Data points from 0.35 to 0.4 1 $nm^{-1}$ were covered by a CCD blemish and thus omitted. Primary data were processed and desmeared (Lake algorithm) using SAXSquant software. Scattering curves were merged using the program PRIMUS.[55] Theoretical scattering was calculated and fitted to experimental data using the program CRYSOL.[56] $P(r)$ plots were calculated with the program GNOM.[56] For ab initio shape reconstruction, 50 models generated with the program DAMMIF[57] were averaged with the program DAMAVER[58] assuming 2-fold symmetry. The resulting averaged and filtered envelope was superimposed onto crystal structure using the program SUPCOMB.[59]

### Expression and purification of proteins required for reaction assay

ACP SCO3249 (ACP) (NCBI NP_627461.1, residues Met1-Ala89), DptA-PCP1 (PCP1) (NCBI WP_006122820.1, residues Asn929-Thr1030) and HxcO (NCBI NP_627459.1, residues Thr2-Pro600) constructs were designed based on the sequences reported by Kopp *et al.*[25] and Kraas *et al.*[26] and synthesized by DNA 2.0, Inc. Both ACP and HxcO featured the N-terminal tag sequence MHHHHHHHHENLYFQG, and PCP1 featured the C-terminal tag sequence ENLYFQGHHHHHH. Protein production of HxcO, DptA and ACP was performed as described here for CDA-C1. Sfp from *Bacillus subtilis* was expressed and purified based on the protocol of Quadri *et al.*[60]

The first step in purification of ACP, PCP1 and HxcO was nickel affinity chromatography as described here for CDA-C1. HxcO was further purified using a Q Sepharose column (GE Healthcare) with a gradient of 0–1 M NaCl in a buffer of 50 mM Tris–HCl (pH 7.5) and 1 mM tris(2-carboxyethyl)phosphine (TCEP). PCP1 and ACP were cleaved with TEV protease and reapplied to the 5-ml HiTrap IMAC FF column charged with $Ni^{2+}$, then subjected to anion-exchange chromatography using a 5-ml HiTrap Q Sepharose column (GE Healthcare) and a gradient of 0–1 M NaCl in a buffer of 50 mM Tris–HCl (pH 7.5), 1 M NaCl and 2 mM βME. PCP1 was placed into a buffer of 25 mM Hepes (pH 7.0), 50 mM NaCl and 1 mM TCEP, then concentrated and frozen. ACP was applied to a HiPrep 16/60 Sephacryl S-200 HR in buffer containing 25 mM Hepes (pH 7.0) and 50 mM NaCl, then concentrated and frozen.

### Assay for CDA-C1 activity

Assay for CDA-C1 activity was adapted from Kraas *et al.*[26] ACP (200 µM) was incubated for 60 min at room temperature with hexanoyl-coenzyme A (800 µM) (Sigma-Aldrich, Oakville, Canada) and Sfp (40.2 µM), in buffer containing 25 mM Hepes (pH 7.0), 50 mM NaCl and 1 mM TCEP. HxcO (5 µM), FAD (25 µM) and HCl (3.75 mM) were added; the reaction was incubated for a further 60 min. In parallel, PCP1 was incubated at room temperature with serinyl-coenzyme A (800 µM) (Zamboni Chem Solutions, Montreal, Canada) and Sfp (40.2 µM) in buffer containing 25 mM Hepes (pH 7.0), 50 mM NaCl and 1 mM TCEP for 20 min. At the completion of loading of the carrier domains, the reactions were mixed, CDA-C1 (6.2 µM) was added and the solution was further incubated for 1 h.

Separation of differentially loaded carrier proteins was performed using reverse phase chromatography (C4, 4.6 mm × 250 mm Microsorb 300 Å, 5 µM; Varian) at a flow rate of 0.5 ml $min^{-1}$ using gradient elution from 40% to 50% B over 25 min, followed by a gradient up to 55% B over another 25 min (buffer A: 0.2% trifluoroacetic acid with $H_2O$ and buffer B: 0.2% trifluoroacetic acid with acetonitrile). Protein masses were determined offline using ESI-MS (Esquire HCT Ultra; Bruker Daltonics) in positive-ion mode with an ESI nebulizer (Mass Spectrometry Core Facility, Bellini Life Science Complex, Montreal, Canada). The mass spectrometer was set to acquire spectra in the mass range 900–3000 $m/z$ for an average of 3 min, and the protein was infused through a syringe pump at 240 µl $h^{-1}$. Acquired spectra were averaged, and the charge states of the protein were determined with the aid of a charge state ruler (Esquire Data Analysis; Bruker Daltonics).

### Computational analyses

Models of CDA-C1 in the conformation observed in the structures SrfAC, TycC and VibH were produced to allow energy-minimized refined linear interpolation, targeted MD simulations and analysis of SAXS data using the SWISS-MODEL server[61] in alignment mode. Energy-minimized refined linear interpolations ("morphings") were performed using the rigimol and refine functions in the program iPyMOL (Incentive PyMOL 2006 release; DeLano Scientific). Targeted MD simulations were performed with the program NAMD[62] using a hydrated sphere of 47.2 Å radius, a temperature of 310 K and an elastic constant TMDk of 200 kcal/mol/$Å^2$ for between 40 ps and 1 ns. Morphing and targeted MD were performed between the CDA-C1 structure and models with sequence of CDA-C1

in the conformations of SrfAC, TycC and VibH, but note that the reciprocal transition using the structure of the more "open" C domains and a model with its sequence in the CDA-C1 conformation works equally well. Normal mode analyses were undertaken using the Web-based programs WEBnm@[33] and NOMAD-Ref.[34] Buried surface areas were calculated using the Lee and Richards buried surface accessibility calculation[63] implemented in the program CNS.[64] The active-site tunnel was identified with the program CAVER,[42] using the coordinates of H157 as the search seed.

For the model of the transition state of the C-domain-catalyzed reaction, the server SWISS-MODEL was used to produce homology models of upstream ACP starting from the structure of ACP from *Staphylococcus aureus* (Protein Data Bank ID 4DXE; Center for Structural Genomics of Infectious Diseases, unpublished results) and the downstream PCP from the structure of the SrfAC module.[10] The initial position of ACP was derived from molecular docking using the program HADDOCK,[65] with outputs vetted by whether the resulting model agreed with experimental protein–protein contact data.[43,44] The initial position of PCP1 was achieved by superimposition of the SrfAC C domain on CDA-C1. Restraint parameters for the acyl-PPE transition state (TS) were generated using the program PRODRG.[66] Atoms of the TS were arbitrarily placed along the tunnel, and the entire system was subjected to conjugate gradient minimization with no experimental energy terms in the program CNS, to yield the final holo ACP–C–PCP1 model.

### Accession numbers

The coordinates and structure factors for SeMet-CDA-C1 in $P2_12_12_1$ space group and CDA-C1 in $P2_1$ space group have been deposited in the Protein Data Bank with accession codes 4JN3 and 4JN5.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jmb.2013.06.003

## Acknowledgements

We are indebted to Vivian Stojanoff and Edwinto Lazo for data collection at beamline X6A of NSLS, Brookhaven National Laboratory, and to Kurt Dejgaard for advice and training in mass spectrometry, as well as Janice Reimer for purification of Sfp protein; Michael Tarry, Diego Alonzo and Fabien Bergeret for advice and discussion; and Albert Berghuis for advice, critical reading of the manuscript and support of D.R. The plasmid containing Sfp was a kind gift of Nathan Magarvey (McMaster University). We thank Alex Deiters, Qingyang Liu and Yan Zou (North Carolina State University) for preparation of SNAC analogues. We are indebted to John Colucci, Helmi Zaghdane and Robert Zamboni (Zamboni Chem Solutions) for an extremely generous gift of serinyl-coenzyme A. This research was supported by the Canadian Institutes of Health Research Operating Grant No. MOP 106615 awarded to T.M.S., a Human Frontiers Science Program Organization Career Development Award to T.M.S., a Tier 2 Canada Research Chair in Macromolecular Machines held by T.M.S. and a Canadian Institutes of Health Research Strategic Training Initiative in Chemical Biology studentship held by K.B.

*Received 14 March 2013;*
*Received in revised form 26 May 2013;*
*Accepted 3 June 2013*
Available online 10 June 2013

***Keywords:***
nonribosomal peptide synthetases;
condensation domain;
conformational changes;
X-ray crystallography;
calcium-dependent antibiotic

***Abbreviations used:***
NRPS, nonribosomal peptide synthetase; CDA, calcium-dependent antibiotic; SAXS, small-angle X-ray scattering; CAT, coenzyme-A-dependent acyltransferase; ACP, acyl carrier protein; PCP, peptide carrier protein; PPE, phosphopantetheinyl; MD, molecular dynamics; TEV, tobacco etch virus; βME, β-mercaptoethanol; PEG, polyethylene glycol; NSLS, National Synchrotron Light Source; TCEP, tris(2-carboxyethyl)phosphine.

## References

1. Schwarzer, D., Finking, R. & Marahiel, M. A. (2003). Nonribosomal peptides: from genes to products. *Nat. Prod. Rep.* **20**, 275–287.
2. Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D. & Thomas, M. G. (2008). Nonribosomal peptide synthetases involved in the production of medically relevant natural products. *Mol. Pharm.* **5**, 191–211.
3. Konz, D. & Marahiel, M. A. (1999). How do peptide synthetases generate structural diversity? *Chem. Biol.* **6**, R39–R48.
4. Caboche, S., Pupin, M., Leclere, V., Fontaine, A., Jacques, P. & Kucherov, G. (2008). NORINE: a database of nonribosomal peptides. *Nucleic Acids Res.* **36**, D326–D331.
5. van Liempt, H., von Dohren, H. & Kleinkauf, H. (1989). δ-(L-α-Aminoadipyl)-L-cysteinyl-D-valine synthetase from *Aspergillus nidulans*: the first enzyme in penicillin biosynthesis is a multifunctional peptide synthetase. *J. Biol. Chem.* **264**, 3680–3684.
6. Wessels, P., von Dohren, H. & Kleinkauf, H. (1996). Biosynthesis of acylpeptidolactones of the daptomycin type. A comparative analysis of peptide synthetases forming A21978C and A54145. *Eur. J. Biochem.* **242**, 665–673.

7. Zocher, R., Nihira, T., Paul, E., Madry, N., Peeters, H., Kleinkauf, H. & Keller, U. (1986). Biosynthesis of cyclosporin A: partial purification and properties of a multifunctional enzyme from *Tolypocladium inflatum*. *Biochemistry*, **25**, 550–553.
8. Weber, T. & Marahiel, M. A. (2001). Exploring the domain structure of modular nonribosomal peptide synthetases. *Structure*, **9**, R3–R9.
9. Frueh, D. P., Arthanari, H., Koglin, A., Vosburg, D. A., Bennett, A. E., Walsh, C. T. & Wagner, G. (2008). Dynamic thiolation–thioesterase structure of a non-ribosomal peptide synthetase. *Nature*, **454**, 903–906.
10. Tanovic, A., Samel, S. A., Essen, L. O. & Marahiel, M. A. (2008). Crystal structure of the termination module of a nonribosomal peptide synthetase. *Science*, **321**, 659–663.
11. Koglin, A., Mofid, M. R., Lohr, F., Schafer, B., Rogov, V. V., Blum, M. M. *et al.* (2006). Conformational switches modulate protein interactions in peptide antibiotic synthetases. *Science*, **312**, 273–276.
12. Gulick, A. M., Starai, V. J., Horswill, A. R., Homick, K. M. & Escalante-Semerena, J. C. (2003). The 1.75 Å crystal structure of acetyl-CoA synthetase bound to adenosine-5′-propylphosphate and coenzyme A. *Biochemistry*, **42**, 2866–2873.
13. Reger, A. S., Wu, R., Dunaway-Mariano, D. & Gulick, A. M. (2008). Structural characterization of a 140 degrees domain movement in the two-step reaction catalyzed by 4-chlorobenzoate:CoA ligase. *Biochemistry*, **47**, 8016–8025.
14. Sundlov, J. A., Shi, C., Wilson, D. J., Aldrich, C. C. & Gulick, A. M. (2012). Structural and functional investigation of the intermolecular interaction between NRPS adenylation and carrier protein domains. *Chem. Biol.* **19**, 188–198.
15. Mitchell, C. A., Shi, C., Aldrich, C. C. & Gulick, A. M. (2012). The structure of PA1221, a non-ribosomal peptide synthetase containing adenylation and peptidyl carrier protein domains. *Biochemistry*, **51**, 3252–3263.
16. De Crecy-Lagard, V., Marliere, P. & Saurin, W. (1995). Multienzymatic non ribosomal peptide biosynthesis: identification of the functional domains catalysing peptide elongation and epimerisation. *C. R. Acad. Sci., Ser. III*, **318**, 927–936.
17. Stachelhaus, T., Mootz, H. D., Bergendahl, V. & Marahiel, M. A. (1998). Peptide bond formation in nonribosomal peptide biosynthesis. Catalytic role of the condensation domain. *J. Biol. Chem.* **273**, 22773–22781.
18. Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. (2002). The structure of VibH represents nonribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526.
19. Samel, S. A., Schoenafinger, G., Knappe, T. A., Marahiel, M. A. & Essen, L. O. (2007). Structural and functional insights into a peptide bond-forming bidomain from a nonribosomal peptide synthetase. *Structure*, **15**, 781–792.
20. Bergendahl, V., Linne, U. & Marahiel, M. A. (2002). Mutational analysis of the C-domain in nonribosomal peptide synthesis. *Eur. J. Biochem.* **269**, 620–629.
21. Chong, P. P., Podmore, S. M., Kieser, H. M., Redenbach, M., Turgay, K., Marahiel, M. *et al.* (1998). Physical identification of a chromosomal locus encoding biosynthetic genes for the lipopeptide calcium-dependent antibiotic (CDA) of *Streptomyces coelicolor* A3(2). *Microbiology*, **144**, 193–199.
22. Hojati, Z., Milne, C., Harvey, B., Gordon, L., Borg, M., Flett, F. *et al.* (2002). Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem. Biol.* **9**, 1175–1187.
23. Strieker, M. & Marahiel, M. A. (2009). The structural diversity of acidic lipopeptide antibiotics. *ChemBioChem*, **10**, 607–616.
24. Lakey, J. H., Lea, E. J., Rudd, B. A., Wright, H. M. & Hopwood, D. A. (1983). A new channel-forming antibiotic from *Streptomyces coelicolor* A3(2) which requires calcium for its activity. *J. Gen. Microbiol.* **129**, 3565–3573.
25. Kopp, F., Linne, U., Oberthür, M. & Marahiel, M. A. (2008). Harnessing the chemical activation inherent to carrier protein-bound thioesters for the characterization of lipopeptide fatty acid tailoring enzymes. *J. Am. Chem. Soc.* **130**, 2656–2666.
26. Kraas, F. I., Giessen, T. W. & Marahiel, M. A. (2012). Exploring the mechanism of lipid transfer during biosynthesis of the acidic lipopeptide antibiotic CDA. *FEBS Lett.* **586**, 283–288.
27. Rausch, C., Hoof, I., Weber, T., Wohlleben, W. & Huson, D. H. (2007). Phylogenetic analysis of condensation domains in NRPS sheds light on their functional evolution. *BMC Evol. Biol.* **7**, 78.
28. Ehmann, D. E., Trauger, J. W., Stachelhaus, T. & Walsh, C. T. (2000). Aminoacyl-SNACs as small-molecule substrates for the condensation domains of nonribosomal peptide synthetases. *Chem. Biol.* **7**, 765–772.
29. Vonrhein, C., Schlauderer, G. J. & Schulz, G. E. (1995). Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases. *Structure*, **3**, 483–490.
30. Yonus, H., Neumann, P., Zimmermann, S., May, J. J., Marahiel, M. A. & Stubbs, M. T. (2008). Crystal structure of DltA. Implications for the reaction mechanism of non-ribosomal peptide synthetase adenylation domains. *J. Biol. Chem.* **283**, 32484–32491.
31. Van Wynsberghe, A., Li, G. & Cui, Q. (2004). Normal-mode analysis suggests protein flexibility modulation throughout RNA polymerase's functional cycle. *Biochemistry*, **43**, 13083–13096.
32. Marechal, J. D. & Perahia, D. (2008). Use of normal modes for structural modeling of proteins: the case study of rat heme oxygenase 1. *Eur. Biophys. J.* **37**, 1157–1165.
33. Hollup, S. M., Salensminde, G. & Reuter, N. (2005). WEBnm@: a Web application for normal mode analyses of proteins. *BMC Bioinformatics*, **6**, 52.
34. Lindahl, E., Azuara, C., Koehl, P. & Delarue, M. (2006). NOMAD-Ref: visualization, deformation and refinement of macromolecular structures based on all-atom normal mode analysis. *Nucleic Acids Res.* **34**, W52–W56.
35. Hsiao, Y. S., Jogl, G. & Tong, L. (2006). Crystal structures of murine carnitine acetyltransferase in

ternary complexes with its substrates. *J. Biol. Chem.* **281**, 28480–28487.
36. Kim, A. R., Rylett, R. J. & Shilton, B. H. (2006). Substrate binding and catalytic mechanism of human choline acetyltransferase. *Biochemistry*, **45**, 14621–14631.
37. Buglino, J., Onwueme, K. C., Ferreras, J. A., Quadri, L. E. & Lima, C. D. (2004). Crystal structure of PapA5, a phthiocerol dimycocerosyl transferase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* **279**, 30634–30642.
38. Hahn, M. & Stachelhaus, T. (2006). Harnessing the potential of communication-mediating domains for the biocombinatorial synthesis of nonribosomal peptides. *Proc. Natl Acad. Sci. USA*, **103**, 275–280.
39. Hur, G. H., Meier, J. L., Baskin, J., Codelli, J. A., Bertozzi, C. R., Marahiel, M. A. & Burkart, M. D. (2009). Crosslinking studies of protein–protein interactions in nonribosomal peptide biosynthesis. *Chem. Biol.* **16**, 372–381.
40. Janin, J. & Chothia, C. (1976). Stability and specificity of protein–protein interactions: the case of the trypsin–trypsin inhibitor complexes. *J. Mol. Biol.* **100**, 197–211.
41. Amit, A. G., Mariuzza, R. A., Phillips, S. E. & Poljak, R. J. (1986). Three-dimensional structure of an antigen–antibody complex at 2.8 Å resolution. *Science*, **233**, 747–753.
42. Chovancova, E., Pavelka, A., Benes, P., Strnad, O., Brezovsky, J., Kozlikova, B. *et al.* (2012). CAVER 3.0: a tool for the analysis of transport pathways in dynamic protein structures. *PLoS Comput. Biol.* **8**, e1002708.
43. Lai, J. R., Fischbach, M. A., Liu, D. R. & Walsh, C. T. (2006). A protein interaction surface in nonribosomal peptide synthesis mapped by combinatorial mutagenesis and selection. *Proc. Natl Acad. Sci. USA*, **103**, 5314–5319.
44. Lai, J. R., Fischbach, M. A., Liu, D. R. & Walsh, C. T. (2006). Localized protein interaction surfaces on the EntB carrier protein revealed by combinatorial mutagenesis and selection. *J. Am. Chem. Soc.* **128**, 11002–11003.
45. Trivedi, O. A., Arora, P., Vats, A., Ansari, M. Z., Tickoo, R., Sridharan, V. *et al.* (2005). Dissecting the mechanism and assembly of a complex virulence mycobacterial lipid. *Mol. Cell*, **17**, 631–643.
46. Belshaw, P. J., Walsh, C. T. & Stachelhaus, T. (1999). Aminoacyl-CoAs as probes of condensation domain selectivity in nonribosomal peptide synthesis. *Science*, **284**, 486–489.
47. Stachelhaus, T. & Marahiel, M. A. (1995). Modular structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *J. Biol. Chem.* **270**, 6163–6169.
48. Linne, U. & Marahiel, M. A. (2000). Control of directionality in nonribosomal peptide synthesis: role of the condensation domain in preventing misinitiation and timing of epimerization. *Biochemistry*, **39**, 10439–10447.
49. Clugston, S. L., Sieber, S. A., Marahiel, M. A. & Walsh, C. T. (2003). Chirality of peptide bond-forming condensation domains in nonribosomal peptide synthetases: the C5 domain of tyrocidine synthetase is a $^{D}C_{L}$ catalyst. *Biochemistry*, **42**, 12095–12104.
50. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.
51. Van Duyne, G. D., Standaert, R. F., Karplus, P. A., Schreiber, S. L. & Clardy, J. (1993). Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124.
52. Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
53. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N. *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **66**, 213–221.
54. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674.
55. Konarev, Petr V., Volkov, Vladimir V., Sokolova, Anna V., Koch, Michel H. J. & Svergun, Dmitri I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282.
56. Svergun, D., Barberato, C. & Koch, M. H. J. (1992). CRYSOL—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Crystallogr.* **25**, 495–503.
57. Franke, D. & Svergun, D. (2009). DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **42**, 342–346.
58. Volkov, V. & Svergun, D. I. (2003). Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr.* **36**, 860–864.
59. Kozin, M. B. & Svergun, D. I. (2001). Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr.* **34**, 33–41.
60. Quadri, L. E., Weinreb, P. H., Lei, M., Nakano, M. M., Zuber, P. & Walsh, C. T. (1998). Characterization of Sfp, a *Bacillus subtilis* phosphopantetheinyl transferase for peptidyl carrier protein domains in peptide synthetases. *Biochemistry*, **37**, 1585–1595.
61. Arnold, K., Bordoli, L., Kopp, J. & Schwede, T. (2006). The SWISS-MODEL workspace: a Web-based environment for protein structure homology modelling. *Bioinformatics*, **22**, 195–201.
62. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E. *et al.* (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.
63. Lee, B. & Richards, F. M. (1971). The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol.* **55**, 379–400.
64. Brunger, A. T. (2007). Version 1.2 of the crystallography and NMR system. *Nat. Protoc.* **2**, 2728–2733.
65. de Vries, S. J., van Dijk, M. & Bonvin, A. M. (2010). The HADDOCK Web server for data-driven biomolecular docking. *Nat. Protoc.* **5**, 883–897.
66. Schuttelkopf, A. W. & van Aalten, D. M. (2004). PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **60**, 1355–1363.