

Molecular basis for carrier protein-dependent amide bond formation in the biosynthesis of lincosamide antibiotics

Received: 4 September 2022

Accepted: 12 May 2023

Published online: 12 June 2023

Check for updates

Takahiro Mori ^{1,2,3,5}✉, Stanislav Kadlcik ^{1,4,5}, Shuang Lyu ^{1,5}, Zdenek Kamenik ⁴, Kosuke Sakurada ¹, Aninda Mazumdar ⁴, Huibin Wang ¹, Jiri Janata ⁴✉ & Ikuro Abe ^{1,2}✉

In the biosynthesis of the lincosamide antibiotic celesticetin, the condensation enzyme CcbD generates the lincosamide pharmacophore by forming an amide bond between the carrier protein (CP)-tethered proline and ergothioneine-conjugated thiooctose. Although the function of CcbD has been investigated, its substrate specificity, structures and catalytic mechanisms remain unclear. Here we show the structure–function analyses of CcbD. Our biochemical analysis revealed that CcbD exhibits promiscuous substrate specificity towards CP-tethered acyl substrates to generate unnatural lincosamides. Furthermore, structural analyses indicated that CcbD possesses an unusual overall fold, while the N-terminal region shows weak similarity to cysteine proteases. Thus, CcbD, like cysteine proteases, utilizes the Cys-His-Glu catalytic triad to form amide bonds in a CP-dependent manner, which is significantly different from other known amide bond-forming enzymes. Furthermore, the structures of the CcbD/thiooctose complex and the cross-linked CcbD/CcbZ-CP complex, as well as structure-based mutagenesis, revealed the intimate structural details of the CP-dependent amide bond formation reaction.

Lincosamides, including lincomycins A and B (**1** and **2**) (refs. 1,2), Bu-2545 (**3**) (ref. 3) and celesticetin (**4**) (ref. 4), are antibacterial compounds effective against Gram-positive bacteria (Fig. 1a)^{5–8}. Lincomycin A and its semisynthetic derivative clindamycin are clinically used to treat severe bacterial infections in people allergic to penicillin⁹. The lincosamide structures feature an unusual thiooctose core connected via an amide bond with *N*-methyl-*trans*-4-propyl-L-proline (*N*-methyl-PPL; in **1**) or *N*-methyl-L-proline (*N*-methyl-Pro; in **3** and **4**) (ref. 10). Lincosamides block bacterial protein synthesis at the early stage of the elongation cycle by targeting the peptidyltransferase domain of the 50S ribosomal subunit, because of their structural similarity to the 3'-end

of L-Pro-Met-transfer RNA (tRNA) and deacylated tRNA^{11,12}. The structure–activity relationship of lincosamide compounds indicated that the connection between the *N*-methyl-PPL/Pro and the thiooctose is essential for the antibacterial activity and the amino acid selectivity (PPL versus Pro) is also important for the level of antibacterial activity¹³.

In the biosynthesis of **1** and **4**, the connection between PPL/Pro and thiooctose is catalysed by LmbC, LmbN and LmbD, and CcbC, CcbZ and CcbD, respectively^{5–7,14–16}. First, the stand-alone adenylation enzymes LmbC/CcbC catalyse the loading of PPL or Pro onto the phosphopantetheine arm of the carrier protein (CP) domain of LmbN/CcbZ^{14,17}. The condensation enzymes LmbD/CcbD then catalyse the

¹Graduate School of Pharmaceutical Sciences, The University of Tokyo, Tokyo, Japan. ²Collaborative Research Institute for Innovative Microbiology, The University of Tokyo, Tokyo, Japan. ³PRESTO, Japan Science and Technology Agency, Saitama, Japan. ⁴Institute of Microbiology, The Czech Academy of Sciences, Prague, Czech Republic. ⁵These authors contributed equally: Takahiro Mori, Stanislav Kadlcik, Shuang Lyu. ✉ e-mail: tmori@mol.f.u-tokyo.ac.jp; janata@biomed.cas.cz; abei@mol.f.u-tokyo.ac.jp

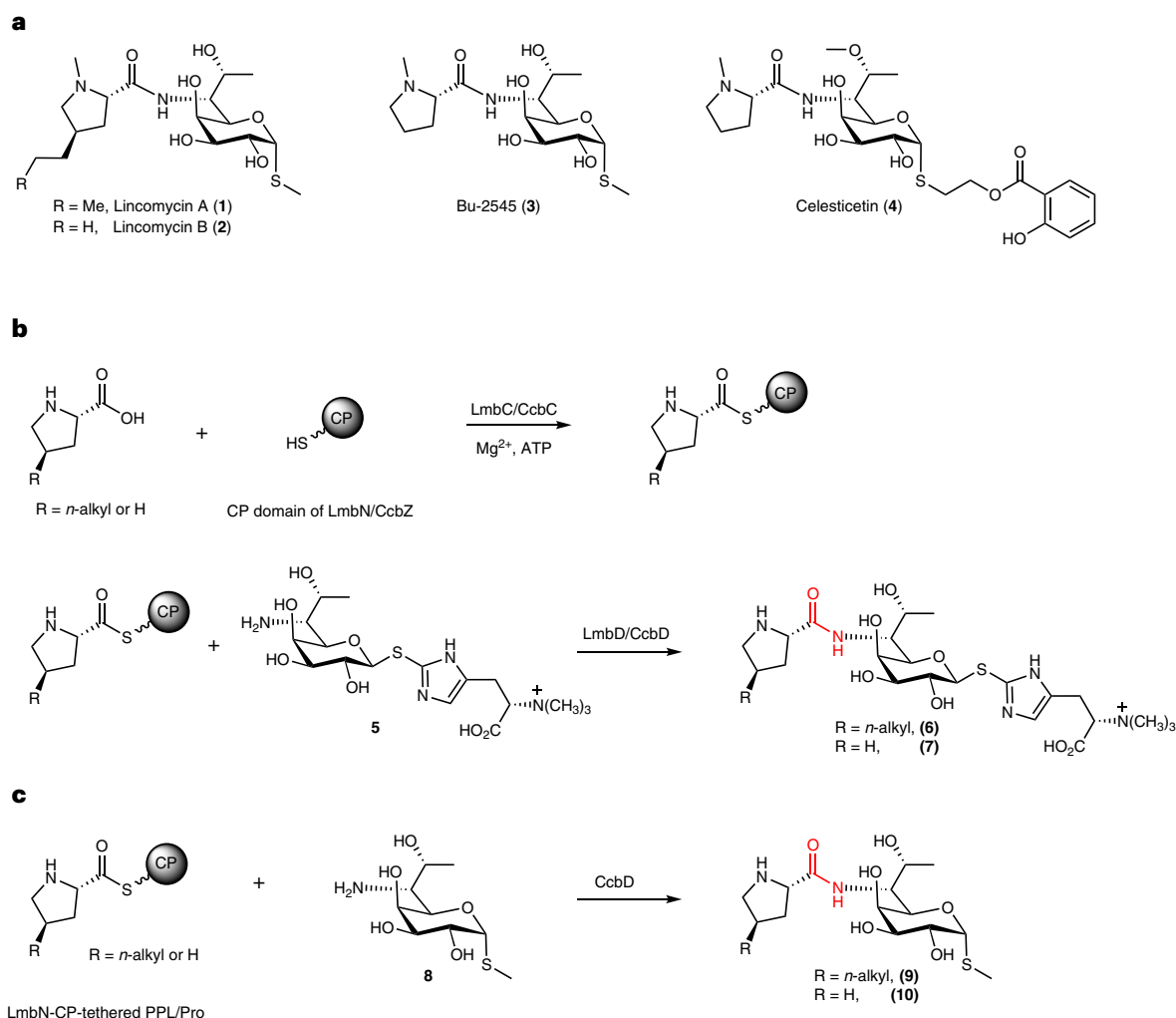


Fig. 1 | Biosynthetic pathway of lincosamide antibiotics. a, Structures of lincosamide antibiotics. **b**, Amide bond formation in the biosynthesis of **1** and **4**. **c**, The enzyme reaction of CcbD with **8**.

condensation of CP-tethered PPL, or Pro and ergothioneine (EGT) S-conjugated thiooctose **5**, to form an amide bond in **6** and **7**, respectively (Fig. 1b). The amino acid selectivity is determined by the LmbC and CcbC specificities, which strictly recognize PPL and Pro to form LmbN-CP-tethered PPL and CcbZ-CP-tethered Pro, respectively¹⁷. In contrast, CcbD shows relaxed substrate specificity towards both of its substrates, CP-tethered Pro and thiooctose^{14,17}. Specifically, as the aminoacyl substrate, CcbD can utilize both CcbZ-CP-tethered Pro (its natural substrate) and LmbN-CP-tethered PPL (intermediate from **1** biosynthesis), and it also accepts methylthiolincosamide (MTL: **8**), in which EGT is substituted with a C1-S-methyl group bearing the opposite stereochemistry compared with the natural substrate **5**, to generate **9** and **10**, respectively (Fig. 1c)¹⁵.

Although this CP-dependent condensation system is similar to those of non-ribosomal peptide synthetases (NRPSs), LmbD/CcbD show low sequence similarity to any functionally characterized amide bond-forming enzymes (below 15%)^{13,14}. Furthermore, the model structure of CcbD predicted by AlphaFold2 (ref. 18) does not resemble those of other amide bond-forming enzymes (Supplementary Fig. 1)^{19–25}. These observations suggest that LmbD/CcbD possess an unusual fold and employ a different mechanism for amide bond formation from other enzymes. Furthermore, although the function of CcbD has been investigated by *in vivo* and *in vitro* analyses, its substrate specificities towards CP-tethered acyl-substrates, overall structures, catalytic residues and catalytic mechanisms remain unclear.

In this Article, we performed structure–function analyses of CcbD by *in vitro* characterization, crystallization and structure-based mutagenesis. The enzyme reaction analyses with L-prolyl-CoA and unnatural acyl substrates revealed that the enzyme recognizes the CP moiety of the substrate, but shows broad substrate specificity towards the acyl moiety to generate various unnatural lincosamide compounds. The crystallographic analyses of CcbD indicated that its overall structure is not related to known structurally characterized amide bond-forming enzymes, while its N-terminal region is weakly similar to those of cysteine proteases²⁶. Furthermore, site-directed mutagenesis experiments indicated that CcbD utilizes Cys-His-Glu residues as a catalytic triad. Lastly, the complex structures of CcbD with **8** and CcbD with CP revealed the molecular basis for the recognition of substrates and CP, and the mechanism of the CcbD-catalysed CP-dependent amide bond formation reaction.

Results

Substrate specificity of CcbD

First, we compared the CcbD reaction between **5** and **8** to understand the substrate preference of CcbD for the C1-S-alkyl moiety. The yield of the product from **8** was comparable to that from **5**, indicating that CcbD does not recognize the structure of C1-S-alkyl moiety in the enzyme reaction (Extended Data Fig. 1). Second, while CcbD shows substrate tolerance towards the alkyl-chain length of PPL¹⁷, the requirement of CP and the substrate scope for non-proline acyl-substrates have

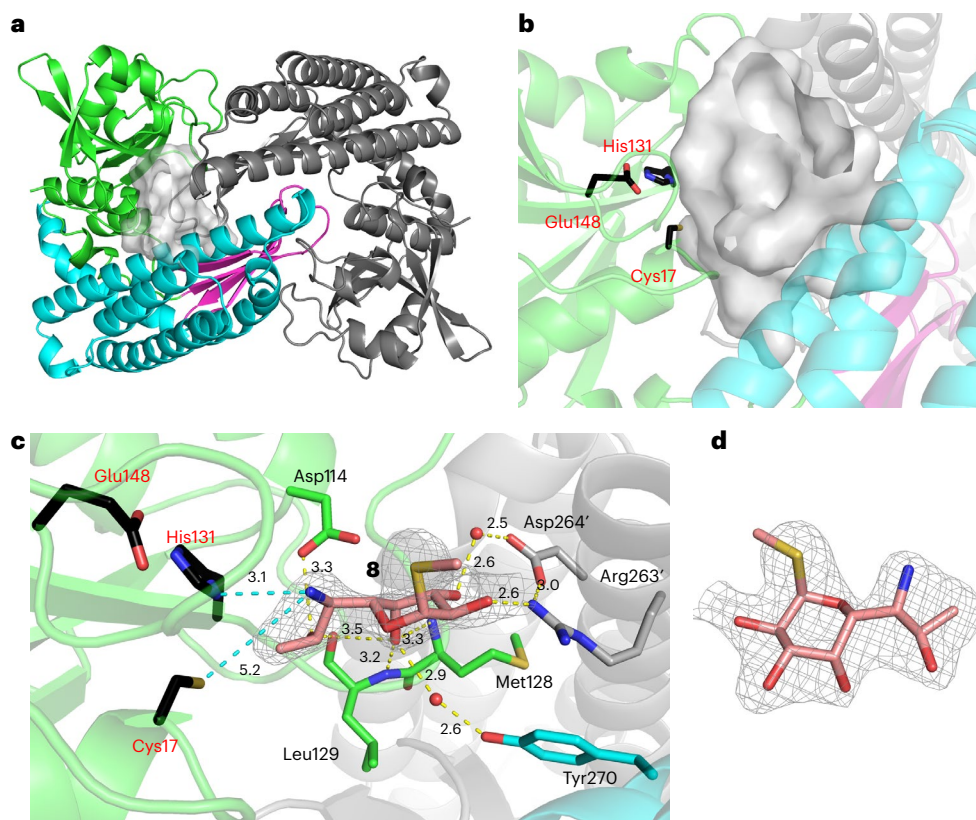


Fig. 2 | The structure of CcbD. **a**, The homodimer structure of CcbD. The α/β -fold domain, four anti-parallel β -sheets and C-terminal five α -helix domain are coloured green, magenta and cyan, respectively. **b**, The active site pocket of CcbD and the catalytic triad. **c**, The active site architecture and binding mode of substrate **8**. Dashed yellow lines represent hydrogen bonds. Dashed cyan lines

show distances between the amino group of **8** and catalytic residues. Water molecules are shown with red non-bonded spheres. **d**, Fo–Fc polder map of **8**. The electron density maps of **8** (contoured at $+3.0\sigma$) in **c** and **d** are represented by a grey mesh. The catalytic residues are highlighted with red.

not been clarified. Therefore, to understand the importance of CP in the enzyme reaction, L-Pro-CoA was incubated with CcbD and **8**. We determined that CcbD also uses L-Pro-CoA as a substrate, instead of CP-tethered L-Pro, although the reaction efficiency is significantly lower than that with CP-tethered L-Pro (Extended Data Fig. 2). We next tested various acyl-CPs, including acetyl-CP (**9**), butanoyl-CP (**10**), hexanoyl-CP (**11**), octanoyl-CP (**12**), alanoyl-CP (**13**), pipercolyl-CP (**14**) and phenylalanoyl-CP (**15**), as substrates of CcbD. Interestingly, CcbD accepted all of them as substrates to generate **16–24** with yields of ~ 8 – 34% (Extended Data Fig. 3 and Supplementary Table 1). CcbD produced two compounds **18/19** and **20/21** from **11** and **12**, respectively. The structure of **19** was determined to be *N*-hexanoyl-**8** by comparison with authentic compound (Supplementary Fig. 2a–c and Supplementary Table 2). Further MS/MS analysis suggested that **21** is *N*-octanoyl-**8**, and **18** and **20** are products resulting from ester bond formation at the C7-hydroxyl group (Supplementary Fig. 2c,d). Moreover, we also tested various amino-containing compounds **25–36** as substrates of CcbD (Extended Data Fig. 4). Surprisingly, CcbD accepted **32–36** as substrates to generate dipeptide compounds **37–41**, although the yields are quite low, suggesting that CcbD recognizes the ethanolamine moiety of substrates.

These results indicated that the CP moiety is important for the precise substrate recognition by CcbD, while CcbD shows more relaxed specificity for the proline moiety and amino substrates to generate unnatural-type lincosamides and dipeptides.

Overall structure and active site of CcbD

To understand the structural basis for the CP-dependent amide bond-forming reaction by CcbD, we solved the structure of

selenomethionine-labelled CcbD at 2.5 \AA resolution (Fig. 2 and Supplementary Table 3). CcbD exists as a homodimer and the overall structure of CcbD consists of three domains, including the α/β -fold domain, which contains five α -helices and six β -strands, four anti-parallel β -sheets and the C-terminal five α -helix domain (Fig. 2a). The CcbD monomer possesses a large cleft between the α/β -fold domain and the C-terminal α -helix domain. The four anti-parallel β -sheet domain of the adjoining monomer inserts into this cleft to form a dimer. The dimer is stabilized through electrostatic, hydrophilic and hydrophobic interactions between the amino acid residues on each inserted β -sheet domain, and the dimer interface buries $\sim 2,850 \text{ \AA}^2$. The monomers are nearly identical to each other, with root-mean-square deviations (RMSDs) of 0.3 \AA . The dimerization creates a large cavity between the dimer surface, consisting of $\alpha 1$, $\alpha 5$, $\alpha 10$ and a loop between $\alpha 6$ and $\beta 5$ from monomer A, and $\alpha 10$ and a loop between $\beta 1$ and $\beta 2$ from monomer B. The estimated total volumes and areas of the active-site cavity are 790 \AA^3 and 920 \AA^2 , respectively.

Interestingly, despite the lack of sequence similarity, the α/β -fold domain of CcbD shares weak structural similarity to those of cysteine proteases, including a putative C39-like peptidase protein with unknown function from *Bacillus anthracis* (PDB code 3ERV) and the phytochelatin synthase (PCS)-like enzyme NsPCS from *Nostoc* sp. pcc 7120 (PDB code 2BTW), with RMSDs of 3.3 \AA and 3.3 \AA for the C α atoms, respectively (13% and 10% identity)²⁶. However, the four anti-parallel β -sheets and the C-terminal five α -helix domain, which are important for the homodimerization of CcbD, were not observed in cysteine proteases and NsPCS (Extended Data Fig. 5a–c). Comparison of the active site between CcbD and NsPCS indicated that the active site shape and residues are not conserved except for the catalytic triad as

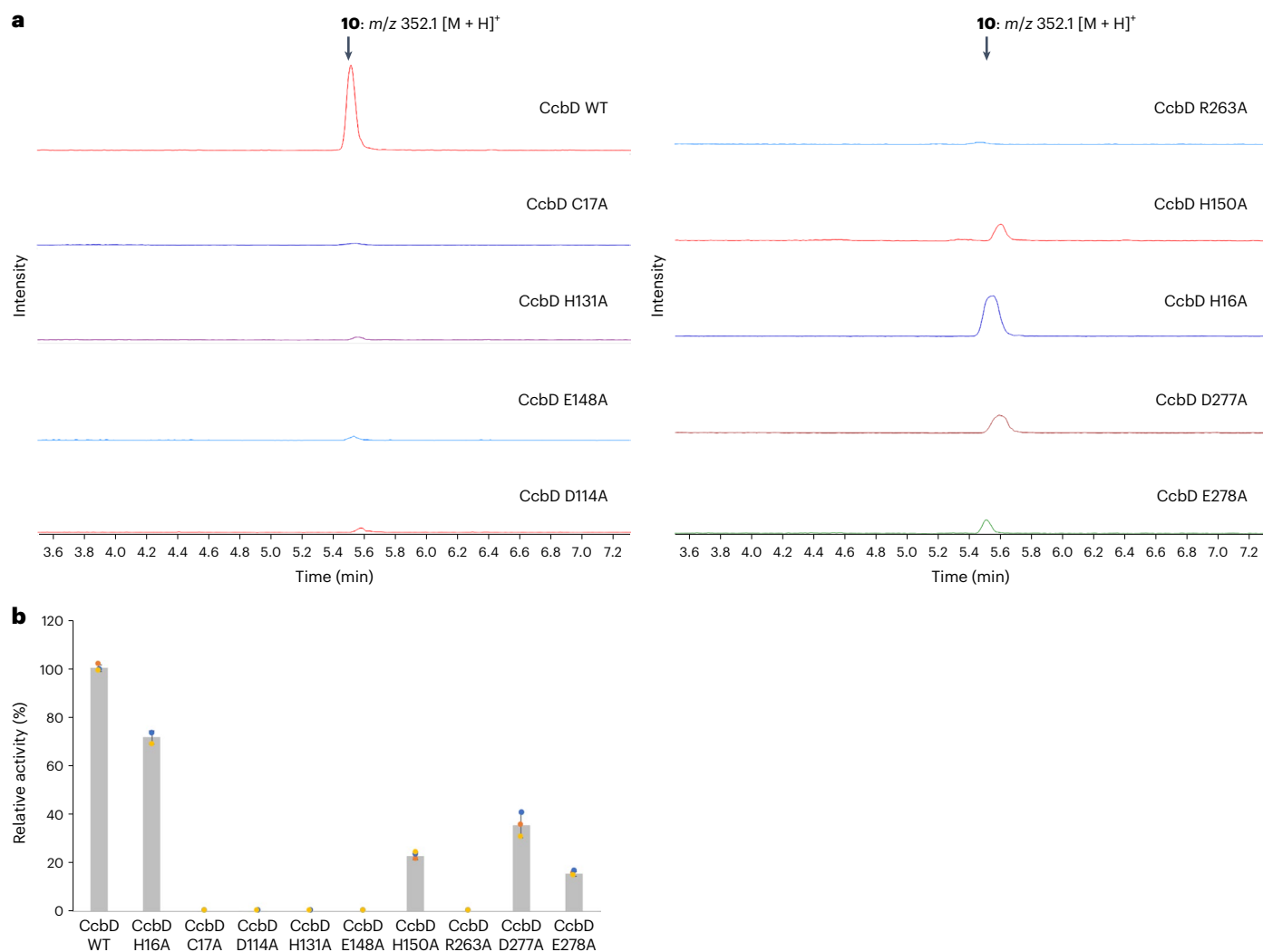


Fig. 3 | Mutagenesis experiment of residues in the active site of CcbD. a, LC-MS charts of the reaction products of CcbD WT and variants. **b**, Relative activities of CcbD variants. The bars are means of $n = 3$ independent experiments, and the error bars indicate s.d. All experiments were repeated independently three times with similar results.

described below. The ligand binding site of NsPCS is located on the surface of monomer, while CcbD forms large pocket to specifically accepts acyl-tethered CP (Extended Data Fig. 5d,e).

Identification of catalytic residues

Cysteine proteases and PCSs utilize a Cys-His-Asn/Asp catalytic triad for the enzyme reaction^{27–30}. CcbD also possesses a similar triad composed of residues Cys17, His131 and Glu148 in the putative active site cavity (Fig. 2b). The C17A, H131A and E148A variants completely lost the amide bond-forming activity, indicating that these residues play a catalytic role in the enzyme reaction of CcbD (Fig. 3a,b). The structural similarities and conservation of the catalytic triad between cysteine proteases and CcbD suggested that the first step of the CcbD reaction is the formation of the CcbD–PPL/Pro complex, by the nucleophilic attack of Cys17 to the thioester group of CP-tethered PPL/Pro.

Binding mode of the thiooctose substrate

We solved the complex structure of CcbD with **8** at 2.25 Å resolution to investigate the binding mode of thiooctose. Compound **8** binds near Cys17 in the catalytic centre through hydrogen bond interactions (Fig. 2c,d). A C2-hydroxy group interacts with Arg263' from the

adjoining monomer, and the C3-hydroxy group also forms a hydrogen bond network with Arg263' and Asp264' via a water molecule. Furthermore, the C4-hydroxy group interacts with Tyr270 via a water molecule and the main chain amides of Met128 and Leu129. The position of the C7-hydroxy group is fixed by the hydrogen bond with the Asp114 side chain and the main chain carbonyl of Leu129. The C6-amino group is positioned close to the catalytic residues Cys17 and His131, at distances of 5.2 Å and 3.1 Å, respectively, suggesting that His131, once activated by Glu148, abstracts the hydrogen atom to activate the amino group of **8**, and the activated amino attacks the thioester of the Cys-tethered PPL/Pro intermediate. The D114A and R263'A variants, whose side chains interact with the sugar substrate, abolished the amide bond-forming activity, suggesting that the hydrogen bond networks with the sugar moiety of the substrate are important for substrate recognition (Fig. 3). The S-methyl group of **8** is oriented towards the entrance of the active site cavity, and there is sufficient space to accept the EGT moiety of the natural substrate **5**. The ethanolamine moiety of **32–36** would bind in similar positions to those of the C6-amino and C7-hydroxy groups of **8** and is likely to be recognized by Asp114, Leu129 and the catalytic His131 as in the **8**-binding. Furthermore, this sugar binding site is significantly large enough to bind other amino-containing compounds, including bulky phenylalanol and tryptophanol.

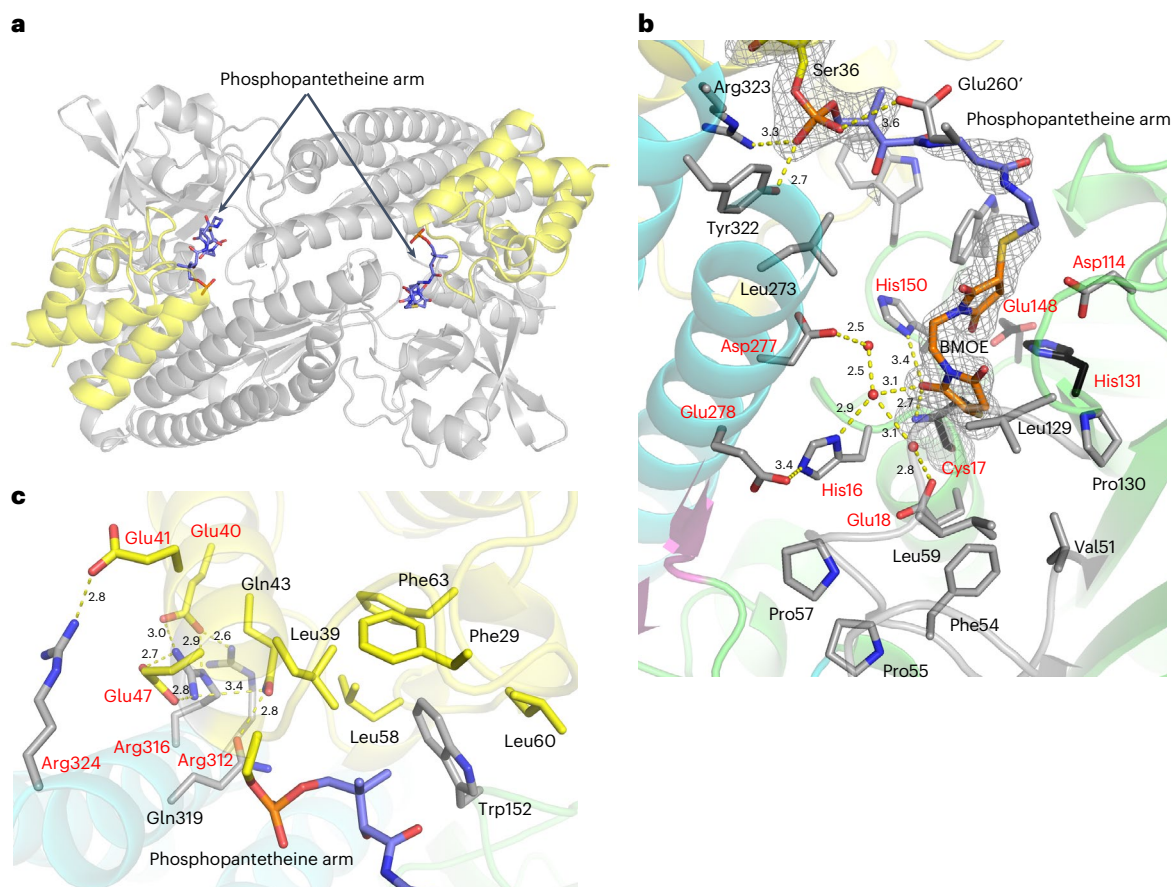


Fig. 4 | The structure of CcbD/CcbZ-CP complex. a, The overall structure of the CcbD/CcbZ-CP complex. The CcbD and CcbZ-CP domains are represented by grey and yellow cartoons. **b**, The active site of CcbD in the complex structure. **c**, The interaction between CcbD and CcbZ-CP. The residues on CcbD and CcbZ-CP

are coloured grey and yellow, respectively, in **b** and **c**. Dashed yellow lines represent hydrogen bonds. The Fo-Fc polder map of covalently bounded Cys17 in CcbD, Ser36 in CcbZ-CP and 4'-phosphopantetheine arm (contoured at +3.0 σ) is represented by a grey mesh. The mutated residues are highlighted with red.

Cross-linking reaction of CcbD and CcbZ-CP

To understand how CcbD selectively recognizes CP during the reaction, we attempted to crystallize CcbD and the CcbZ-CP complex. First, we performed a site-specific cross-linking reaction of CcbD and CcbZ-CP by using a bifunctional maleimide reagent, 1,2-bis(maleimido) ethane (BMOE), to obtain covalently cross-linked CcbD/CcbZ-CP³¹. While CcbZ-CP does not have any cysteine residues, CcbD possesses five, with Cys216 and Cys249 located on the surface of the enzyme (Supplementary Fig. 3). Therefore, we substituted these cysteines with serine to prevent non-specific cross-linking reactions, and the resulting C216S/C249S variant was used for cross-linking to the sulfhydryl group of the 4'-phosphopantetheine in CcbZ-CP. The incubation of the C216S/C249S variant and holo-CcbZ-CP in the presence of BMOE efficiently afforded the covalent complex (Extended Data Fig. 6). In contrast, no apparent cross-linking was observed between the C17S/C216S/C249S variant and CcbZ-CP (Extended Data Fig. 6). These results clearly suggested that the specific cross-linking reaction occurred between the sulfhydryl groups of the 4'-phosphopantetheine of CcbZ-CP and Cys17 via BMOE in the active site of CcbD.

We purified and crystallized the cross-linked CcbD/CcbZ-CP complex. The structure of the CcbD/CcbZ-CP complex was solved at 2.3 Å resolution. In the complex structure, CcbZ-CP is located between the α/β -fold domain and the C-terminal α -helical domain, and covers the active site entrance of CcbD (Fig. 4a). The structure of CcbD in the complex is almost identical to that in the wild type (WT), with an RMSD of 0.6 Å for the C α atoms. The interface area between CcbD and CcbZ-CP

comprises ~ 770 Å², which is 4.8% of the surface area of CcbD and 17% of the surface area of CcbZ-CP. This contact area is similar to those of other complex structures of CPs and enzymes, such as the polyketide synthase and NRPS machineries^{32–34}.

The 4'-phosphopantetheine moiety, which is covalently tethered to Ser36 of CcbZ-CP, is located in the same cavity as **8**. The maleimide groups of BMOE form covalent bonds with the sulfhydryl group of the catalytic Cys17 of CcbD and the sulfhydryl group of the 4'-phosphopantetheine of CcbZ-CP (Fig. 4b). The pantetheine moiety does not form tight interactions with the active site residues, probably because BMOE is larger than the natural substrate, proline. The phosphate group of the 4'-phosphopantetheine forms a hydrogen bond network with Tyr322, Arg323 and Glu260' of CcbD. The C1-carbonyl group of the maleimide, which forms a covalent bond with Cys17, also interacts with the main chain amide of Cys17 and the side chain of His150. Considering that the Cys17-tethered PPL/Pro should be in a similar position to that of maleimide, His150 and the main chain amide group of Cys17 would form an oxyanion hole to stabilize the tetrahedral intermediate during the thioester bond formation and subsequent amide bond formation reactions. The activity of the H150A variant is significantly reduced to 23%, indicating the importance of the oxyanion hole (Fig. 3 and Supplementary Table 4). The hydrophilic residues, including His16, Asp277 and Glu278, are located close to the catalytic Cys17 and form a hydrogen bond network via water molecules. Their H16A, D277A and E278A variants also had dramatically decreased activity, suggesting that the amino group of PPL/Pro would interact with this hydrophilic environment (Figs. 3 and 4b and Supplementary Table 4).

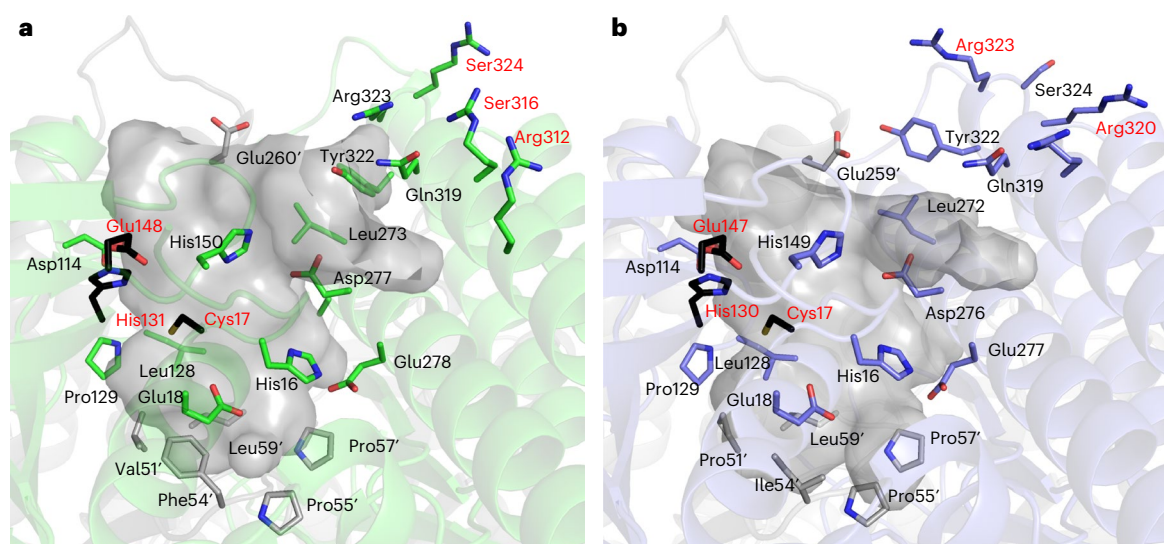


Fig. 5 | Comparison of the active sites of CcbD and LmbD. a, The active site of CcbD crystal structure. **b**, The active site of LmbD model structure. The catalytic residues and the important residues for interaction with CP are highlighted with red.

Notably, a large space is observed deeper inside the cavity, beyond the maleimide binding site (Fig. 4b and Extended Data Fig. 7). This space is constructed by hydrophobic residues and large enough to accept long alkyl-chains and large side chains of amino acids. These observations suggest that the alkyl-chain of PPL and the other acyl substrates bind in this space, and explain why CcbD shows broad substrate specificity. The model structure of LmbD suggests that the catalytic triad, active site residues and hydrophobic space are well conserved in LmbD, suggesting that LmbD would show a similar property to CcbD (Fig. 5).

The superimposition of the CcbD/8 and CcbD/CcbZ-CP structures and the docking model of 5 in CcbD/CcbZ-CP suggest that the binding sites for 5/8 and 4'-phosphopantetheine overlap and there is no space for the simultaneous binding of 5/8 and Pro-tethered 4'-phosphopantetheine (Extended Data Fig. 8). To further investigate the formation of CcbD-proline intermediate, we also incubated 8 with the presumed Pro-tethered CcbD, which was prepared by incubating CcbD, CcbC, holo-CcbZ-CP and proline (without 8) and removing the substrate by desalting. As a result, even after removal of the proline substrates, 10 was still produced (Supplementary Fig. 4a). Moreover, when Pro-CoA was incubated with or without CcbD, the non-enzymatic release of proline was not significantly increased in the presence of CcbD, suggesting that CcbD does not catalyse the hydrolysis of either Pro-CoA or Pro-conjugated Cys to release proline in the active site (Supplementary Fig. 4b). This could be due to the accessibility of water molecules in the active site. In the active site of NsPCS, which catalyses deglycination of glutathione, a water molecule interacting with catalytic His183 is located close to the acylated Cys70 (Extended Data Fig. 5f). In contrast, no activated water molecules interacting with basic or acidic amino acids were observed within 4 Å of Cys17 in the CcbD structures. On the basis of these observations, we believe that the formation of the CcbD-PPL/Pro complex and the ping-pong mechanism are more likely.

Interaction between CcbD and CcbZ-CP

CcbD recognizes CcbZ-CP through salt bridges, hydrogen bond interactions and hydrophobic interactions (Fig. 4c). The Arg312, Arg316 and Arg324 residues on α 12 in CcbD form salt bridges with Glu47, Glu40 and Glu41 on α 2 in CcbZ-CP, respectively, and Arg323 and Gln319 in CcbD interact with Gln43 in CcbZ-CP via a hydrogen bonding network. Moreover, Trp152 in CcbD is inserted into the hydrophobic region of CcbZ-CP (Fig. 4c and Supplementary Fig. 5).

To confirm the importance of these interactions, we constructed the CcbD R312A, R316A and R324A variants and their triple variant. The variants of the counterpart residues in CcbZ-CP, including E40A, E41A and E47A and their triple variant, were also constructed. Cross-linking assays to evaluate the interactions of each CcbD variant with CcbZ-CP and CcbD with CcbZ-CP variants revealed that CcbD R316A and the triple variant of CcbD, and E40A, E47A and the triple variant of CcbZ-CP significantly reduced the cross-linking efficiency (Extended Data Fig. 9). Furthermore, the amide bond formation activities of the CcbD R316A and CcbD R312A/R316A/R324A variants and the CcbZ-CP E40A/E41A/E47A variant were significantly decreased by 50%, 5% and 38%, respectively, while the activities of CcbD R312A and R324A were comparable to that of the WT (Extended Data Fig. 10 and Supplementary Table 4). Arg316 forms salt bridges with both Glu40 and Glu47 and a hydrogen bond interaction with Gln43, which is the reason why Arg316 is the most important residue for the activity. These results clearly indicated that the salt bridge and hydrophilic interactions between CcbD and CcbZ-CP are important for CP recognition. The sequence alignment between LmbN-CP and CcbZ-CP (68% identity) revealed that the hydrophilic and hydrophobic residues, observed in the interactions between CcbD and CcbZ-CP, are conserved except for Leu60. This would be the reason why CcbD accepts LmbN-CP-tethered PPL in addition to its natural substrate (Supplementary Fig. 6).

Discussion

Amide bond formation is one of the most ubiquitous reactions in nature, and a variety of amide bond-forming enzyme families have been identified, including the condensation (C)-domain of NRPS, ATP-dependent amide bond synthetase, *N*-acyltransferase, aminoacyl-tRNA synthetase, PCS and ATP-grasp enzymes³⁵. Among them, the CP-dependent amide bond formation is known only in the C-domains of NRPSs and some Gcn5-related *N*-acetyltransferase (GNAT) enzymes^{35–38}. The structure of the NRPS C-domain consists of two domains that form V-shaped pseudo-dimers (Supplementary Fig. 1c)^{19,39–41}. The active site, formed in the interface tunnel of the two domains, contains the conserved HHxxxDG catalytic motif for positioning the α -amino group of the acceptor substrate to facilitate the nucleophilic attack to the thioester in the donor substrate. Donor and acceptor peptide CPs bind on the front and back faces of the active site tunnel. Although the structures of CP-dependent GNAT enzymes have not been reported, GNAT

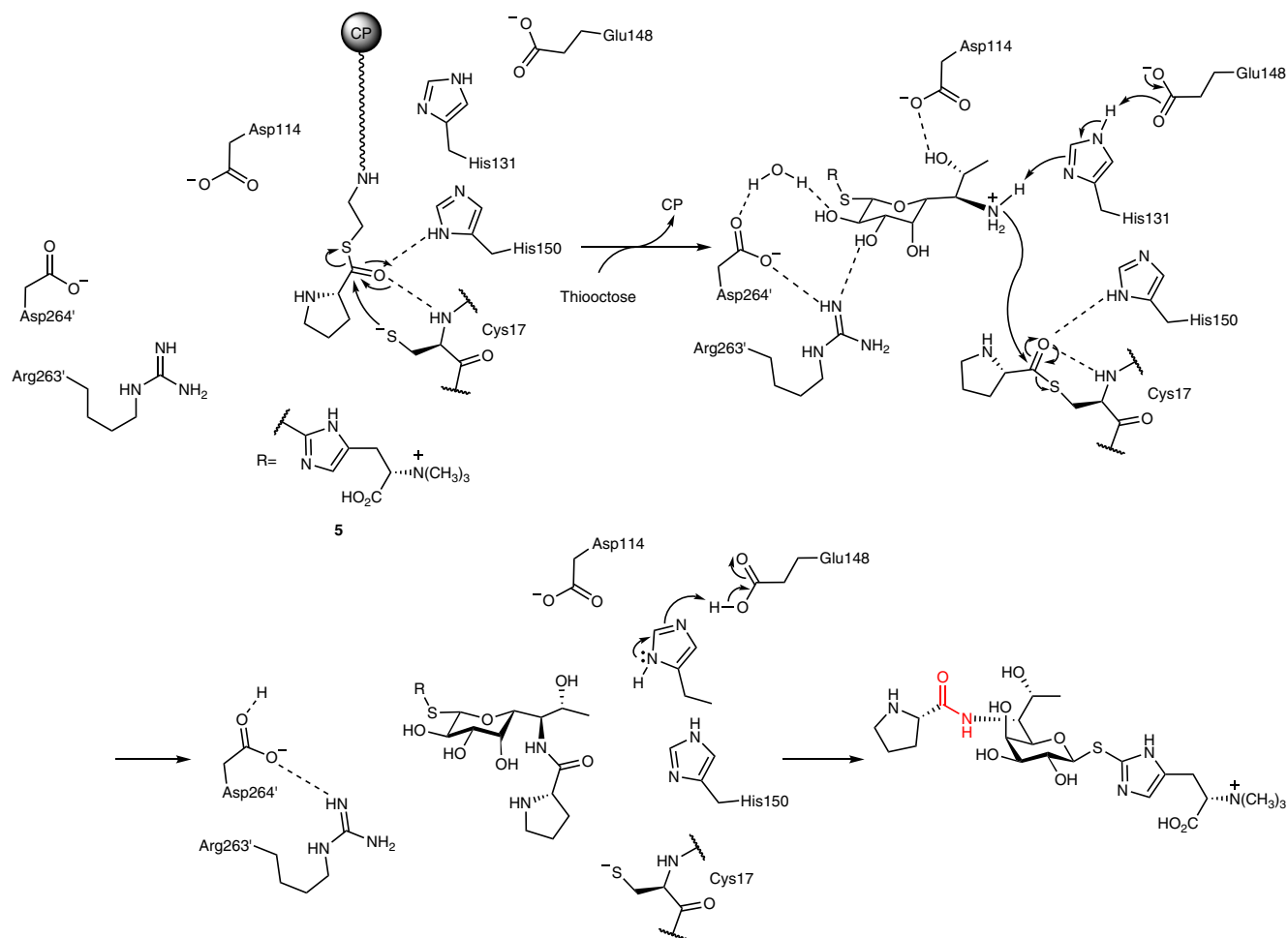


Fig. 6 | Proposed mechanism of the CcbD-catalysed amide bond-formation reaction. Dashed lines represent hydrogen-bond interactions. The reaction is initiated by loading of Pro onto catalytic Cys17 to generate the CcbD-Pro complex. Further nucleophilic attack from the amino group of thiooctose, which is activated by His131 and Glu148, generates an amide bond.

superfamily enzymes share the conserved GNAT-fold comprising the central β -strands flanked by α -helices (Supplementary Fig. 1h)⁴².

In contrast, the structure of the N-terminal domain of CcbD is weakly related to those of cysteine proteases and PCs. The active site of CcbD is constructed on the dimer interface, while those of other cysteine proteases and PCs are formed in the monomer cavity²⁶. The CcbD-specific C-terminal α -helix domain is involved in dimerization, active site formation and interactions with CP, mainly through hydrophilic interactions and salt bridges. Furthermore, the C-terminal α -helix domain in the adjoining monomer also forms the active site and interacts with the 4'-phosphopantetheine.

Regardless of the lack of sequence similarity and active site architectures, CcbD has a Cys17-His131-Glu148 catalytic triad similar to the catalytic triads of cysteine proteases and PCs^{26–30,43–45}. In the reaction of PCs, which catalyses deglycination of glutathione and condensation between γ -Glu-Cys and glutathione to generate phytochelatins, the catalytic Cys residue is initially acylated by the first glutathione and form enzyme-associated γ -Glu-Cys. Then, the amino group of another glutathione attacks the thioester bond of the enzyme- γ -Glu-Cys intermediate to generate phytochelatins. Similarly, the chemoenzymatic peptide synthesis by cysteine proteases has been demonstrated as the reverse reaction of hydrolysis, through thermodynamically or kinetically controlled synthesis^{46,47}. In these peptide syntheses, the acyl donor substrate, activated with an ester, amide or nitrile bond, is attacked by a nucleophilic cysteine residue to form a thioester bond

with the protease. Subsequently, the thioester bond is cleaved by the amino acid substrate to generate the amide bond.

Considering the similarity of the catalytic residues, CcbD is expected to share a catalytic mechanism comparable to those of cysteine proteases and PCs. On the basis of these observations, we propose the following mechanisms for the condensation between thiooctose and CP-tethered Pro/PPL (Fig. 6). The enzyme reaction is initiated by the binding of the carbonyl group of Pro in the CP-tethered substrate into the oxyanion hole of the active site, through interactions with the backbone of Cys17 and the imidazole ring of His150. The nucleophilic attack from Cys17 to the thioester bond in the Pro-connected CP then generates the CcbD-Pro complex, with the release of CP. Subsequently, the thiooctose binds to the active site of the CcbD-Pro complex, and the thioester at Cys17 is attacked by the amino group of thiooctose, with nucleophilicity enhanced by the proximity of His131 and Glu148, to generate an amide bond. The high sequence similarity (56% identity) and the conservation of the active site architectures between CcbD and LmbD suggest that the reaction mechanism of LmbD is the same as that of CcbD (Fig. 5). Thus, the CcbD/LmbD-catalysed condensation reaction between the amino sugar substrate and Pro/PPL is structurally and mechanistically different from those of the NRPS system but resembles those of cysteine proteases and PCs.

In conclusion, our structural analysis of CcbD revealed that CcbD/LmbD catalyses the formation of the pharmacophore of lincosamide antibiotics via a mechanism that is structurally and mechanistically

significantly distinct from the previously identified CP-dependent amide bond-forming enzymes, and thus provided insights into the diversity of amide bond-forming reactions in nature. Future structure-based engineering of CcbD/LmbD will develop the enzymes as biocatalysts to generate unnatural lincosamides with various acyl and sugar moieties for future drug discovery.

Methods

General

Oligonucleotide primers (Supplementary Table 5 and Supplementary Data 1) and DNA sequencing services were provided by Eurofins Genomics. The restriction enzymes and PrimeSTAR GXL DNA polymerase were purchased from Takara Bio Solvents and chemicals were purchased from Wako Chemicals, Merck KGaA and Hampton Research, unless noted otherwise. PCR was performed using a TaKaRa PCR Thermal Cycler Dice Gradient (Takara Bio). The nuclear magnetic resonance spectra of compounds were recorded on ECX-500 MHz (JEOL) spectrometers.

Construction of plasmids for CcbD expression

The *ccbD* gene was PCR amplified from the chromosomal DNA of the celesticetin-producing type strain *Streptomyces caelestis* ATCC 15084, using the primers Fw: CCGCATATGGCCCAATCCAAGGGTTCGGTTGAT and Rv: CCGCTCGAGGAGTTCCTTGAGCAATCGCCG. The *ccbD* gene was inserted into the pET42b vector (Novagen) via the NdeI and XhoI restriction sites. The resulting plasmid was used to produce C-terminally His₈-tagged CcbD. The plasmid for CcbC expression was used as previously reported¹⁷.

Construction of plasmids for expression of CcbD and CcbZ-CP variants

The primers used for the construction of plasmids for site-directed mutagenesis studies are listed in Supplementary Table 5. The plasmid for the expression of WT CcbD or WT CcbZ-CP was used as the template for PCR-based site-directed mutagenesis, which was performed with a QuikChange Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's protocol.

Expression and purification of CcbD, CcbZ-CP, holo-CcbZ-CP and their variants

The pET42a plasmids for the expression of CcbD and its variants were transformed into *Escherichia coli* BL21(DE3). The resulting strains were cultured in LB medium supplemented with 34 mg l⁻¹ chloramphenicol and 50 mg l⁻¹ kanamycin sodium at 37 °C, with shaking at 160 rpm. Plasmids for the expression of CcbZ-CP and its mutants and *sfp* were transformed into *E. coli* BL21(DE3) or *E. coli* BLR. The resulting strains were cultured in LB medium supplemented with 100 mg l⁻¹ ampicillin sodium and 50 mg l⁻¹ kanamycin sodium, respectively, at 37 °C, with shaking at 160 rpm. When the OD₆₀₀ reached 0.6, the cell cultures were cooled on ice for 30 min, and then isopropyl β-D-1-thiogalactopyranoside (0.3 mM) was added to induce the target protein expression and the cultures were continued at 16 °C, 160 rpm. After 18 h of post-induction incubation, cells were collected by centrifugation at 5,500g for 10 min and suspended in lysis buffer, containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 5 mM imidazole and 5% glycerol. The cell suspension was sonicated for 5 min on ice. After the cell debris was removed by centrifugation at 20,000g for 30 min, the supernatant was mixed with 1 ml Ni-NTA agarose resin and loaded onto a gravity flow column. Unbound proteins were removed with 50 ml lysis buffer containing 30 mM imidazole, and then the His-tagged protein was eluted with lysis buffer containing 300 mM imidazole. Buffers for CcbD H16A, C17A, D114A, H131A, E148A, H150A, R263A, D277A and E278A purification were supplemented with 1 mM dithiothreitol. For the in vitro assay, the eluates of CcbD and its variant proteins were concentrated to 10 mg ml⁻¹ after the removal of imidazole, using a

30 kDa Amicon Ultra-15 filtration unit (Merck Millipore). The eluates containing holo-CcbZ-CP and its variant proteins were concentrated to 20 mg ml⁻¹ after the imidazole was removed, using a 3 kDa or 10 kDa Amicon Ultra-15 filtration unit (Millipore). For crystallization, the eluate from the Ni-NTA agarose was applied to a 6 ml RESOURCE Q anion exchange chromatography column (4 °C, Cytiva) and a HiLoad 16/60 Superdex 200 pre-packed gel filtration column (4 °C, Cytiva), and eluted with a solution containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 5% glycerol and 1 mM dithiothreitol. The resulting eluate was concentrated to 15 mg ml⁻¹, using an Amicon Ultra-4 (molecular weight cut-off 30 kDa) filter at 4 °C. The purity of the proteins was monitored by SDS-PAGE, and the protein concentrations were determined with a SimpliNano microvolume spectrophotometer.

In vitro assays of CcbD and its variants

The standard enzymatic reaction of CcbD with substrate was pre-incubated for 30 min in a 50 μl reaction mixture, containing 50 mM Tris-HCl (pH 7.5), 2 mM L-proline, 2 mM dithiothreitol, 5 mM ATP (pH 7.5), 10 mM MgCl₂, 2 μM CcbC, 50 μM holo-CcbZ-CP or its variants, and 2 mM substrate **5** or **8**. After this pre-incubation, 10 μM CcbD (WT or variants) was added to the reaction and incubated for 5 min at 30 °C. To measure the relative activity between **5** and **8**, the consumption of substrates was calculated by liquid chromatography-mass spectrometry (LC-MS) analysis. For the analysis of substrate specificity towards acyl substrates, 10 μM CcbD was incubated in 50 mM HEPES buffer (pH 7.5) buffer, containing 10 mM MgCl₂, 100 μM acyl-CoA, 100 μM **8**, 30 μM Sfp and 50 μM apo-CcbZ-CP on a 25 μl scale, overnight at 30 °C. For the substrate specificity analysis for amino compounds, the reaction buffer containing 50 mM HEPES (pH 7.5), 100 μM L-proline, 5 mM ATP (pH 7.5), 5 mM MgCl₂, 2 μM CcbC, 100 μM holo-CcbZ-CP and 10 μM CcbD was pre-incubated for 30 min in a 25 μl reaction mixture, then 100 μM of amino substrates were added and reacted for overnight at 30 °C.

To measure total turnover number and conversion rate towards L-proline reaction, the reaction buffer containing 50 mM HEPES (pH 7.5), 1 mM L-proline, 1 mM **8**, 5 mM ATP (pH 7.5), 5 mM MgCl₂, 20 μM CcbC and 100 μM CcbZ-CP was pre-incubated for 30 min in a 25 μl reaction mixture. After this pre-incubation, 2 μM CcbD was added to the reaction and incubated for 13 h at 30 °C. For the total turnover number and conversion rate analysis towards acyl-substrates reactions, the reaction buffer containing 5 mM MgCl₂, 100 μM acyl-CoA, 100 μM MTL, 30 μM Sfp and 100 μM apo-CcbZ-CP was pre-incubated for 30 min in a 25 μl reaction mixture. After this pre-incubation, 10 μM CcbD was added to the reaction and incubated for 13 h at 30 °C. Each sample was then subjected to LC-MS analyses and the consumption of **8** was calculated.

LC-MS methods used for analysing in vitro reactions. For the analysis of the CcbD WT with various substrates LC-MS samples were injected into a Shimadzu LabSolutions LCMS 8045 system. A HILICpak VG-50 2D column (Shodex) or C18-MS-II column (Nacal Tesque) was used for separation. The gradient elution was performed with solvent A (50 mM NH₄Ac) and solvent B (CH₃CN), with a flow rate of 0.2 ml min⁻¹ (*T* = 0 min, 95% B; *T* = 3 min, 95% B; *T* = 20 min, 50% B; *T* = 25 min, 95% B; and *T* = 40 min, 95% B: for Hilic column) or with solvent A (1 mM NH₄ formate) and solvent B (CH₃CN), with a flow rate of 0.2 ml min⁻¹ (*T* = 0 min, 2.5% B; *T* = 4 min, 2.5% B; *T* = 20 min, 100% B: for C₁₈ column). For the analysis of the CcbD R312A, R316A, R324A and R312A/R316A/R324A variants and the E40A/E41A/E47A variant of CcbZ-CP, LC-MS samples were injected into a Shimadzu LabSolutions LCMS 8045 system. A HILICpak VG-50 2D column was used for separation. The gradient elution was performed with solvent A (50 mM NH₄Ac) and solvent B (CH₃CN), with a flow rate of 0.2 ml min⁻¹ (*T* = 0 min, 95% B; *T* = 3 min, 95% B; *T* = 20 min, 50% B; *T* = 25 min, 95% B; and *T* = 40 min, 95% B). For the analysis of CcbD C17A, D114A, H131A, E148A, H150A, R263A, Asp277 and Glu278, LC-MS analyses were performed on an Agilent

1260 Infinity II LC coupled to an Agilent 6546 LC/Q-TOF MS equipped with a Jet Stream electrospray ion source (Agilent Technologies). A 3 μl portion of the sample was loaded on the ACQUITY PREMIER BEH Amide (2.1 mm \times 100 mm, 1.7 μm) LC column, which was maintained at 40 °C. A two-component mobile phase, B and A, containing acetonitrile (ACN) with 20 mM ammonium formate, pH 4.75, at 9:1 (v/v), and ACN with 20 mM ammonium formate, pH 4.75, at 1:1 (v/v), was used for separation. The elution was performed at a flow rate of 0.4 ml min⁻¹ with the following conditions (min/%B) 0/99.0; 1/99.0; 7/1.0; 9/1.0; and 10/99.0, with 1.0 min column clean-up (99.0% B) and 2 min equilibration (99% B). The mass spectrometer was used in the positive mode and the data storage was centroid, with a dual AJS electrospray ion source ion source under the following conditions: the drying gas temperature, 250 °C; drying gas, 8 l min⁻¹; nebulizer, 35 psi; sheath gas temperature, 400 °C; sheath gas flow, 12 l min⁻¹; capillary voltage set at +3,500 V, nozzle voltage +200 V; acquisition rate 5 spectra s⁻¹ and time 200 ms per spectrum. MS chromatograms for ions corresponding to [M + H]⁺ were extracted with a 0.05 Da tolerance window. The data acquisition and post-run processing were performed with the MassHunter Workstation software, version 10.1 (Agilent). The Q-TOF MS was tuned in the positive mode for the mass range of 50–1,700 *m/z*, using Swarm Autotune. Purine (121.050873 *m/z*) and Agilent HP-0921 (922.009798 *m/z*) were used as the references to maintain the mass accuracy. The data were further processed with the Qualitative Analysis 10.0 software (Agilent).

Analysis of CcbD-Pro-conjugated intermediate. To investigate the generation of CcbD-Pro intermediate, the substrates were removed after incubation of enzyme reaction components without **8**. The reactions were performed at 30 °C for 1 h in a 100 μl reaction mixture containing 50 mM HEPES (pH 7.0), 5 mM ATP, 5 mM MgCl₂, 1 mM proline, 10 μM CcbC, 50 μM holo-CcbZCP and 50 μM CcbD. Then, the small molecules were removed by Zeba Micro Spin Desalting Columns, 7K molecular weight cut-off, three times. After removing substrates, final 200 μM of MTL was added and further incubated at 30 °C for 1 h. As a negative control, a reaction without CcbD was also performed, and 10 μM CcbD and 200 μM MTL were added after the desalting process to confirm that no proline nor Pro-tethered CcbZ-CP remained in the solution. Each sample was then subjected to LC–MS analyses.

Hydrolysis of Pro-conjugated cysteine in the reaction. To analyse whether the CcbD generates proline via hydrolysis of Pro-conjugated catalytic cysteine, the Pro-CoA was incubated with or without CcdD. The reactions were performed by the incubation of in the buffers containing 50 mM HEPES (pH 7.0), 200 μM L-proline and (20 μM CcbD) for 2 h at 30 °C in a 25 μl reaction mixture. Each sample was then subjected to LC–MS analyses.

Production and purification of natural thiooctose substrate of CcbD (5)

The seed culture of *S. lincolnensis* *lmbN* Δ CP strain¹³ was prepared by inoculating spores into 50 ml of YEME medium in 500 ml flat-bottom boiling flasks, which were incubated at 28 °C for 30 h, 180 rpm. A 2 ml portion of the seed culture was then inoculated into 40 ml of AVM medium¹³ in 500 ml flat-bottom boiling flasks and incubated at 28 °C for 120 h. The supernatants from 30 flat-bottom boiling flasks were used in the next steps. The cells were centrifuged at 4,000g at 4 °C for 10 min, and the supernatant was stored at –20 °C. The supernatant was adjusted to pH 2–3 with formic acid and extracted in two steps. First, Amberlite XAD-4 in a glass column was used, and the amount of the sorbent was approximately 5 cm in diameter and 10 cm in height. Methanol (MeOH) followed by 0.1% formic acid was used to equilibrate the column before loading the supernatant. The eluate-containing compounds not interacting with Amberlite and including the compound of interest were collected. Second, MCX 35 cc (6 g) (Waters) cartridges were used. The cartridge was conditioned and equilibrated with

MeOH followed by 2% formic acid, and then the collected eluate from the previous extraction was applied on the cartridge, which was washed with 2% formic acid and MeOH. Thereafter, 200 ml MeOH with 5% of an aqueous solution of ammonium hydroxide (29%) was loaded to elute the compound of interest. For further purification of the natural substrate, two-step preparative high-performance liquid chromatography was performed. The first step was carried out using a Triart Diol-HILIC column (250 \times 20 mm, 5 μm particles; YMC) with a two-component mobile phase: A, 50 mM ammonium acetate, pH 4.7, and B, ACN. The flow rate was 8 ml min⁻¹ with a linear gradient (min/% of B) 0/95; 5/95; 28.6/30; and 32/95 with equilibration before the next analysis. The second step was performed on a Luna column (C18, 250 \times 15 mm, 5 μm ; Phenomenex) using a two-component mobile phase, with A as 0.1% formic acid and B as MeOH. The flow rate was 3 ml min⁻¹ with a gradient of 5% to 90% B, for 70 min. The fractions with the natural thiooctose substrate were monitored using an Acquity UPLC system with a 2996 PDA detection system (194–600 nm), connected to an LCT Premier XE time-of-flight mass spectrometer (Waters). The sample (5 μl) was loaded onto the ACQUITY PREMIER BEH Amide (2.1 mm \times 50 mm, 1.7 μm) LC column, maintained at 40 °C. A two-component mobile phase, B and A, containing ACN with 20 mM ammonium formate, pH 4.75, 9:1 (v/v) and ACN with 20 mM ammonium formate, pH 4.75, 1:1 (v/v), respectively, was used for separation. The elution was performed at a flow rate of 0.4 ml min⁻¹ with the following gradient (min/%B) 1/99.0; 7/1.0; 9/1.0; and 10/99.0, with 1.0 min column clean-up (99.0% B) and 2 min equilibration (99% B). The mass spectrometer was used in the positive mode (W) with the cone voltage, +40 V; the capillary voltage, +2,800 V; ion source block temperature, 120 °C; desolvation gas temperature, 350 °C; desolvation gas flow, 800 l h⁻¹; and cone gas flow, 50 l h⁻¹; with an inter-scan delay of 0.01 s and a scan time of 0.15 s. The mass accuracy was maintained using lock spray technology with leucine-enkephalin as the reference compound (2 ng μl^{-1} , 5 μl min⁻¹). The diode array detector detection technique and the same conditions were used to quantify the natural substrate. The chromatograms were extracted at the absorption maxima of the compounds, and the area of the peaks was used to construct a five-point calibration curve ($R = R^2 = 0.9999$) of the EGT standard (Cayman Chemical), which was used to quantify the natural thiooctose substrate (5).

Preparation and purification of the cross-linked CcbD-CcbZ-CP complexes

For the cross-linking assays of CcbD-CcbZ-CP and their variants, 20 μM CcbD or CcbD variant protein was mixed with 60 μM holo-CcbZ-CP or holo-CcbZ-CP variant proteins and 0.2 mM BMOE, in buffer containing 40 mM NaH₂PO₄/K₂HPO₄ (pH 7.0), 5 mM EDTA and 5% glycerol. At 10, 30 and 60 min after the reaction was initiated, a 50 μl aliquot of the reaction was removed and the reaction was quenched with SDS–PAGE loading buffer and monitored by SDS–PAGE. To obtain the CcbD-CcbZPCP complex proteins, 60 μM His-tag free CcbD C249S/C216S variant was mixed with 600 μM His-tagged holo-CcbZPCP protein, in buffer containing 40 mM NaH₂PO₄/K₂HPO₄ (pH 7.0), 5 mM EDTA and 5% glycerol. The cross-linking reaction was initiated by adding 0.2 mM BMOE, and incubated on ice for 1 h. The BMOE was removed by passage through a PD-10 column, and the complexes were monitored by SDS–PAGE. The remaining CcbD variant proteins and holo-CcbZPCP proteins were removed by chromatography on nickel affinity resin and a HiLoad 16/60 Superdex 200 pre-packed gel filtration column (4 °C, Cytiva).

Crystallization and structure determination

Crystals of selenomethionine-labelled CcbD were obtained after 3 days at 20 °C, while crystals of CcbD-apo and CcbD-CcbZPCP complex were obtained after 3 days at 10 °C. All crystals were obtained by using the sitting-drop vapour-diffusion method with the following reservoir solutions: selenomethionine-labelled CcbD: 0.02 M Tris–HCl (pH 6.7), 0.16 M MgCl₂ and 10% w/v PEG 8000; and CcbD WT: 0.02 M Tris–HCl

(pH 6.7), 0.16 M MgCl₂, and 10% w/v PEG 8000. The complex structures were prepared by incubating CcbD crystals at 10 °C for 1 h with 50 mM **8** in the crystallization drop. The crystallization conditions to prepare the CcbD-CcbZ-CP complex were 0.1 M MES (pH 6.7), 0.16 M MgCl₂, 0.2 M sodium thiocyanate and 8% w/v PEG 8000. The crystals were transferred into the cryoprotectant solution (reservoir solution with 25% (v/v) glycerol), and then flash-cooled at -173 °C in a nitrogen gas stream. The X-ray diffraction datasets were collected at BL-1A (Photon Factory), using a beam wavelength of 1.1 Å. The diffraction datasets for selenomethionine-labelled CcbD, CcbD with **8** and CcbD-CcbZ-CP were processed and scaled using the XDS program package⁴⁸ and Aimless in CCP4 (ref. 49). The determination of Se sites and the generation of the initial model were performed with Crank2 in CCP4 (ref. 50). The dataset for selenomethionine-labelled CcbD was slightly twinned. The analysis by Xtriage in PHENIX⁵¹ suggested that the correlation between the intensities related by the twin law h, -k, -h-l, with an estimated twin fraction of 0.15 is most likely due to a non-crystallographic symmetry axis parallel to the twin axis. Further processes were carried out without detwin. The initial phases of CcbD with **8** and the CcbD-CcbZ-CP structures were determined by molecular replacement, using the selenomethionine-labelled CcbD structure as the search model. Molecular replacement was performed with Phaser⁵² in PHENIX⁵¹. The initial phases were further calculated with AutoBuild in PHENIX⁵¹. The structures were modified manually with Coot⁵³ and refined with PHENIX.refine⁵⁴. The cif parameters of **8** for the energy minimization calculations were obtained by using the PRODRG server⁵⁵. The final crystal data and intensity statistics are summarized in Supplementary Table 3. The Ramachandran statistics are as follows: 97.7% favoured, 2.3% allowed for selenomethionine-labelled CcbD; 98.0% favoured, 1.9% allowed, 0.1% outliers allowed for CcbD with **8**; and 98.4% favoured, 1.6% allowed for CcbD-CcbZ-CP. A structural similarity search was performed, using the Dali program server⁵⁶. All crystallographic figures were prepared with PyMOL (DeLano Scientific, <http://www.pymol.org>). For docking model structure with **5**, initial docking models were constructed using the AutoDock Vina plugin in UCSF Chimera⁵⁷. Then, the conformation of ligands was manually modified on the basis of the binding mode of **8** in the complex structure of CcbD/**8** to avoid the close contacts between the ligands and the active site residues, with Coot.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data supporting the findings of this study are available within this article and its Supplementary Information files and source files or from the corresponding authors on request. The crystallographic data that support the findings of this study are available from the Protein Data Bank (<http://www.rcsb.org>). The coordinates and the structure factor amplitudes for the selenomethionine-labelled CcbD, complex structures of CcbD with **8** and CcbD with CcbZ-CP were deposited under accession code 7YNI⁵⁸, 7YN2⁵⁹ and 7YN3⁶⁰, respectively. Source data are provided with this paper.

References

- Owen, S. P., Bhuyan, B. K. & Kupiecki, F. P. Discovery and biological properties of cis-beta-carboxyacrylamidine, a new cytotoxic agent. *Antimicrob. Agents Chemother.* **5**, 808–811 (1965).
- Mason, D. J. & Lewis, C. Biological activity of the lincomycin-related antibiotics. *Antimicrob. Agents Chemother.* **10**, 7–12 (1964).
- Hanada, M., Tsunakawa, M., Tomita, K., Tsukiura, H. & Kawaguchi, H. Antibiotic Bu-2545, a new member of the celesticetin-lincomycin class. *J. Antibiot.* **33**, 751–753 (1980).
- Hoeksema, H. Celesticetin. V. The structure of celesticetin. *J. Am. Chem. Soc.* **90**, 755–757 (1968).
- Spizek, J. & Rezanka, T. Lincosamides: chemical structure, biosynthesis, mechanism of action, resistance, and applications. *Biochem. Pharmacol.* **133**, 20–28 (2017).
- Zhang, D., Tang, Z. & Liu, W. Biosynthesis of lincosamide antibiotics: reactions associated with degradation and detoxification pathways play a constructive role. *Acc. Chem. Res.* **51**, 1496–1506 (2018).
- Janata, J., Kamenik, Z., Gazak, R., Kadlcik, S. & Najmanova, L. Biosynthesis and incorporation of an alkylproline-derivative (APD) precursor into complex natural products. *Nat. Prod. Rep.* **35**, 257–289 (2018).
- Lin, C. I., McCarty, R. M. & Liu, H. W. The biosynthesis of nitrogen-, sulfur-, and high-carbon chain-containing sugars. *Chem. Soc. Rev.* **42**, 4377–4407 (2013).
- Spizek, J. & Rezanka, T. Lincomycin, clindamycin and their applications. *Appl. Microbiol. Biotechnol.* **64**, 455–464 (2004).
- Hoeksema, H. et al. Chemical studies on lincomycin. I. The structure of lincomycin. *J. Am. Chem. Soc.* **86**, 4223–4224 (1964).
- Tenson, T. & Ehrenberg, M. Regulatory nascent peptides in the ribosomal tunnel. *Cell* **108**, 591–594 (2002).
- Schlünzen, F. et al. Structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* **413**, 814–821 (2001).
- Janata, J. et al. Lincosamide synthetase—a unique condensation system combining elements of non-ribosomal peptide synthetase and mycothiol metabolism. *PLoS ONE* **10**, e0118850 (2015).
- Zhao, Q., Wang, M., Xu, D., Zhang, Q. & Liu, W. Metabolic coupling of two small-molecule thiols programs the biosynthesis of lincomycin A. *Nature* **518**, 115–119 (2015).
- Wang, S. A. et al. Studies of lincosamide formation complete the biosynthetic pathway for lincomycin A. *Proc. Natl Acad. Sci. USA* **117**, 24794–24801 (2020).
- Wang, M., Zhao, Q. & Liu, W. The versatile low-molecular-weight thiols: beyond cell protection. *Bioessays* **37**, 1262–1267 (2015).
- Kadlcik, S. et al. Adaptation of an L-proline adenylation domain to use 4-propyl-L-proline in the evolution of lincosamide biosynthesis. *PLoS ONE* **8**, e84902 (2013).
- Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- Keating, T. A., Marshall, C. G., Walsh, C. T. & Keating, A. E. The structure of VibH represents non-ribosomal peptide synthetase condensation, cyclization and epimerization domains. *Nat. Struct. Biol.* **9**, 522–526 (2002).
- Vetting, M. W., Hegde, S. S., Javid-Majd, F., Blanchard, J. S. & Roderick, S. L. Aminoglycoside 2'-N-acetyltransferase from *Mycobacterium tuberculosis* in complex with coenzyme A and aminoglycoside substrates. *Nat. Struct. Biol.* **9**, 653–658 (2002).
- Retailleau, P. et al. Interconversion of ATP binding and conformational free energies by tryptophanyl-tRNA synthetase: structures of ATP bound to open and closed, pre-transition-state conformations. *J. Mol. Biol.* **325**, 39–63 (2003).
- Noike, M. et al. A peptide ligase and the ribosome cooperate to synthesize the peptide pheganomycin. *Nat. Chem. Biol.* **11**, 71–76 (2015).
- Petchey, M. et al. The broad aryl acid specificity of the amide bond synthetase McbA suggests potential for the biocatalytic synthesis of amides. *Angew. Chem. Int. Ed.* **57**, 11584–11588 (2018).
- Matsuda, K. et al. Heterochiral coupling in non-ribosomal peptide macrolactamization. *Nat. Catal.* **3**, 507–515 (2020).
- Winn, M. et al. Discovery, characterization and engineering of ligases for amide synthesis. *Nature* **593**, 391–398 (2021).
- Vivares, D., Arnoux, P. & Pignol, D. A papain-like enzyme at work: native and acyl-enzyme intermediate structures in phytochelatin synthesis. *Proc. Natl Acad. Sci. USA* **102**, 18848–18853 (2005).

27. Harrison, M. J., Burton, N. A. & Hillier, I. H. Catalytic mechanism of the enzyme papain: predictions with a hybrid quantum mechanical/molecular mechanical potential. *J. Am. Chem. Soc.* **119**, 12285–12291 (1997).
28. Buller, A. R. & Townsend, C. A. Intrinsic evolutionary constraints on protease structure, enzyme acylation, and the identity of the catalytic triad. *Proc. Natl Acad. Sci. USA* **110**, E653–E661 (2013).
29. Harrison, M. J., Burton, N. A., Hillier, I. H. & Gould, I. R. Mechanism and transition state structure for papain catalysed amide hydrolysis, using a hybrid QM/MM potential. *Chem. Commun.* **24**, 2769–2770 (1996).
30. Elsasser, B. et al. Distinct roles of catalytic cysteine and histidine in the protease and ligase mechanisms of human legumain as revealed by DFT-based QM/MM simulations. *ACS Catal.* **7**, 5585–5593 (2017).
31. Miyanaga, A., Iwasawa, S., Shinohara, Y., Kudo, F. & Eguchi, T. Structure-based analysis of the molecular interactions between acyltransferase and acyl carrier protein in vicenistatin biosynthesis. *Proc. Natl Acad. Sci. USA* **113**, 1802–1807 (2016).
32. Evans, S. E. et al. An ACP structural switch: conformational differences between the apo and holo forms of the actinorhodin polyketide synthase acyl carrier protein. *ChemBioChem* **9**, 2424–2432 (2008).
33. Goodrich, A. C., Meyers, D. J. & Frueh, D. P. Molecular impact of covalent modifications on non-ribosomal peptide synthetase carrier protein communication. *J. Biol. Chem.* **292**, 10002–10013 (2017).
34. Greule, A. et al. Kistamicin biosynthesis reveals the biosynthetic requirements for production of highly crosslinked glycopeptide antibiotics. *Nat. Commun.* **10**, 2613 (2019).
35. Winn, M., Richardson, S. M., Campopiano, D. J. & Micklefield, J. Harnessing and engineering amide bond forming ligases for the synthesis of amides. *Curr. Opin. Chem. Biol.* **55**, 77–85 (2020).
36. Kozakai, R. et al. Acyltransferase that catalyses the condensation of polyketide and peptide moieties of goadivionin hybrid lipopeptides. *Nat. Chem.* **12**, 869–877 (2020).
37. Hu, Z., Awakawa, T., Ma, Z. & Abe, I. Aminoacyl sulfonamide assembly in SB-203208 biosynthesis. *Nat. Commun.* **10**, 184 (2019).
38. Barra, L. et al. β -NAD as a building block in natural product biosynthesis. *Nature* **600**, 754–758 (2021).
39. Izoré, T. & Cryle, M. J. The many faces and important roles of protein–protein interactions during non-ribosomal peptide synthesis. *Nat. Prod. Rep.* **35**, 1120–1139 (2018).
40. Challis, G. L. & Naismith, J. H. Structural aspects of non-ribosomal peptide biosynthesis. *Curr. Opin. Struct. Biol.* **14**, 748–756 (2004).
41. Miller, B. R. & Gulick, A. M. Structural biology of non-ribosomal peptide synthetases. *Methods Mol. Biol.* **1401**, 3–29 (2016).
42. Vetting, M. W. et al. Structure and functions of the GNAT superfamily of acetyltransferases. *Arch. Biochem. Biophys.* **433**, 212–226 (2005).
43. Rea, P. A. Phytochelatin synthase: of a protease a peptide polymerase made. *Physiol. Planta* **145**, 154–164 (2012).
44. Rea, P. A. Phytochelatin synthase. eLS <https://doi.org/10.1002/9780470015902.a0028220> (2020).
45. Gisdon, F. J. et al. Structural and biophysical analysis of the phytochelatin-synthase-like enzyme from *Nostoc* sp. shows that its protease activity is sensitive to the redox state of the substrate. *ACS Chem. Biol.* **17**, 883–897 (2022).
46. Hernández, J. G., Ardila-Fierro, K. J., Crawford, D., James, S. L. & Bolm, C. Mechanoenzymatic peptide and amide bond formation. *Green. Chem.* **19**, 2620–2625 (2017).
47. Gimenez-Dejoo, J., Tsuchiya, K. & Numata, K. Insights into the stereospecificity in papain-mediated chemoenzymatic polymerization from quantum mechanics/molecular mechanics simulations. *ACS Chem. Biol.* **14**, 1280–1292 (2019).
48. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
49. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
50. Ness, S. R., de Graaff, R. A., Abrahams, J. P. & Pannu, N. S. CRANK: new methods for automated macromolecular crystal structure solution. *Structure* **12**, 1753–1761 (2004).
51. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
52. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
53. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
54. Afonine, P. V. et al. Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
55. Schüttelkopf, A. W. & van Aalten, D. M. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. D* **60**, 1355–1363 (2004).
56. Holm, L. & Rosenström, P. Dali server: conservation mapping in 3D. *Nucleic Acids Res.* **38**, W545–W549 (2010).
57. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).
58. Mori, T., Kadlick, S., & Abe, I. Crystal structure of selenomethionine labeled CcbD. *Protein Data Bank* <https://doi.org/10.2210/pdb7YN1/pdb> (2023).
59. Mori, T., Lyu, S. & Abe, I. Crystal structure of CcbD with methylthiolincosamide. *Protein Data Bank* <https://doi.org/10.2210/pdb7YN2/pdb> (2023).
60. Mori, T., Lyu, S. & Abe, I. Crystal structure of CcbD complex with CcbZ carrier protein domain. *Protein Data Bank* <https://doi.org/10.2210/pdb7YN3/pdb> (2023).

Acknowledgements

The synchrotron radiation experiments were performed at the BL-1A of the Photon Factory. This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology, Japan (JSPS KAKENHI grant number JP18F18779 JP19K15703, JP20H00490, JP20KK0173, JP21K18246, JP23H00393 and JP23H02641), the New Energy and Industrial Technology Development Organization (NEDO, grant number JPNP20011), AMED (grant number JP21ak0101164), the PRESTO programme from Japan Science and Technology Agency (JPMJPR20DA), the project 20-09811Y from the Czech Science Foundation, Czech Academy of Sciences (Lumina Quaeruntur Program project no. LQ200202002), and the project National Institute of Virology and Bacteriology (Programme EXCELES, ID project no. LX22NPO5103), funded by the European Union—Next Generation EU.

Author contributions

T.M., I.A. and J.J. designed the experiments. T.M., S.K., S.L., K.S., A.M., Z.K. and H.W. performed in vitro analysis and crystallization experiments. T.M., S.K., S.L., K.S., A.M., Z.K. and I.A. analysed the data. T.M., S.K., Z.K. and I.A. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41929-023-00971-y>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41929-023-00971-y>.

Correspondence and requests for materials should be addressed to Takahiro Mori, Jiri Janata or Ikuro Abe.

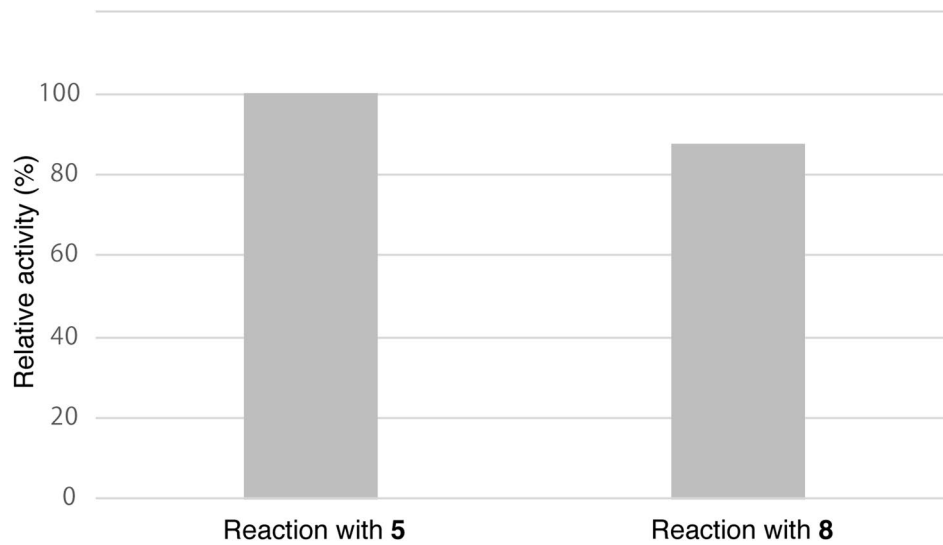
Peer review information *Nature Catalysis* thanks the anonymous reviewers for their contribution to the peer review of this work

Reprints and permissions information is available at www.nature.com/reprints.

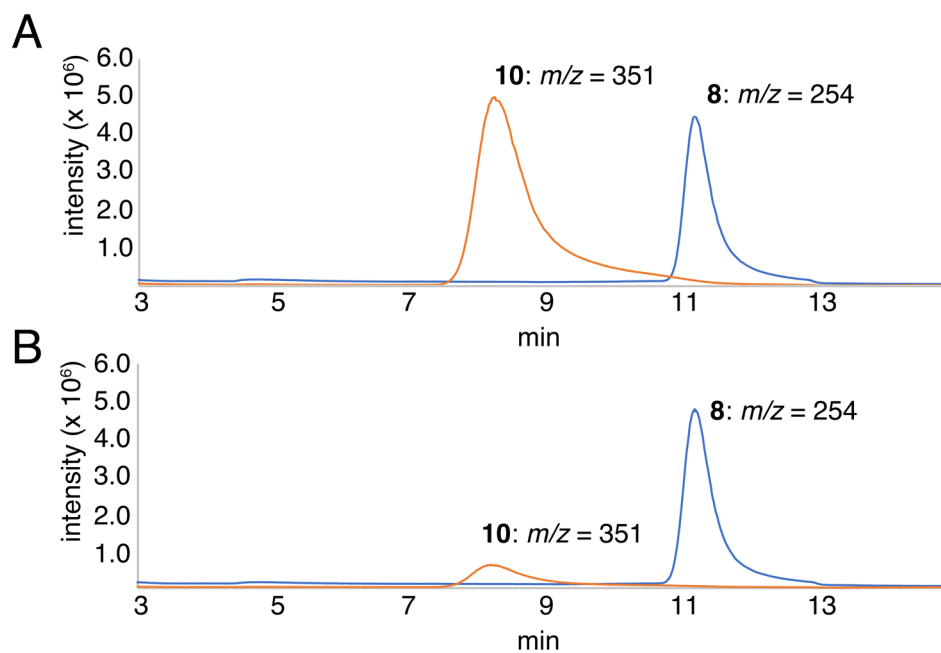
Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

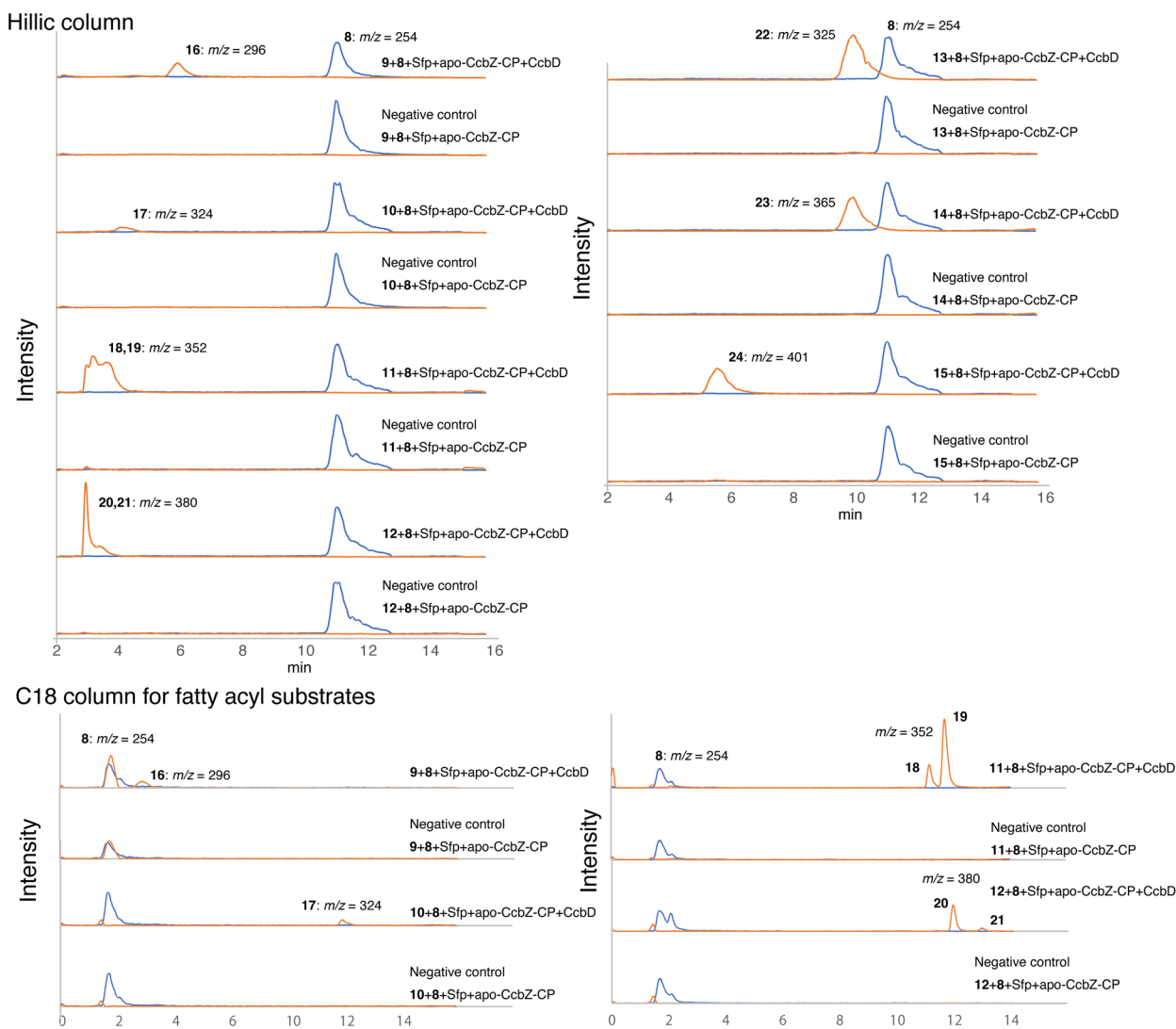


Extended Data Fig. 1 | Comparison of the enzyme reactions between 5 and 8. All reactions were performed in duplicate. Data are presented as mean values.

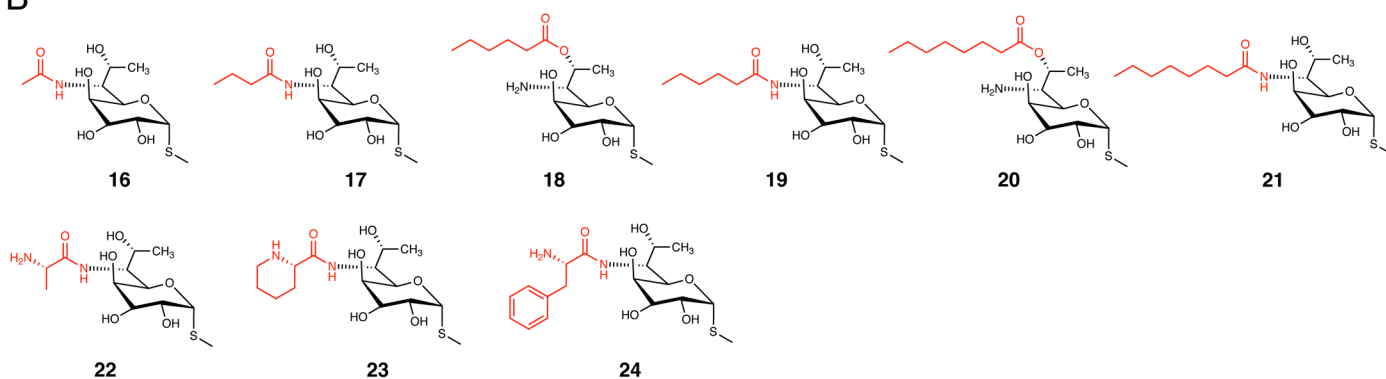


Extended Data Fig. 2 | In vitro enzyme reaction of CcbD. LC-MS chart of the reactions of CcbD with (A) CP-tethered L-Pro and (B) L-Pro-CoA. The reactions were performed at 30 °C for 5 min. Blue and orange charts represent EIC = 254 and EIC = 351, respectively.

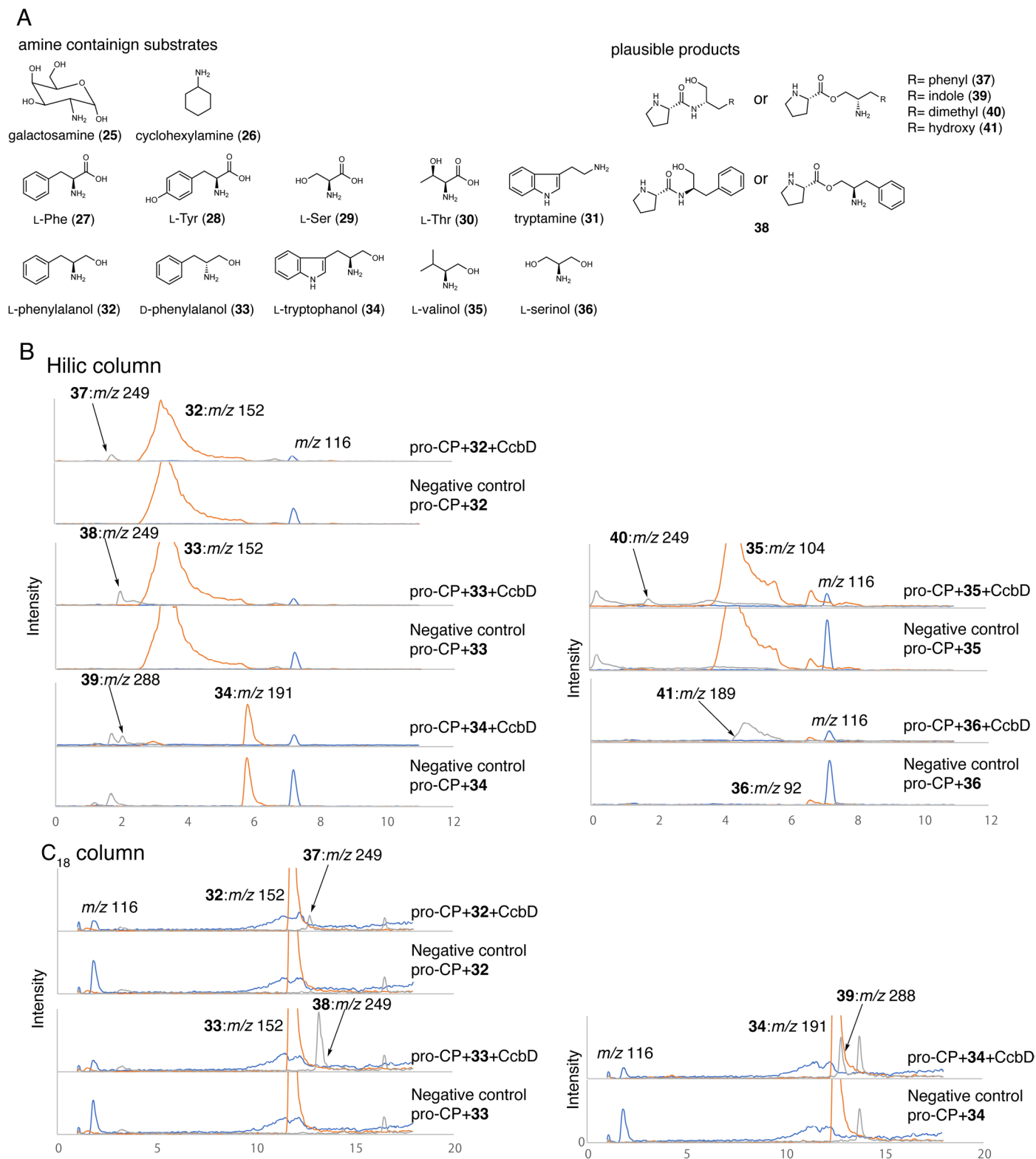
A



B

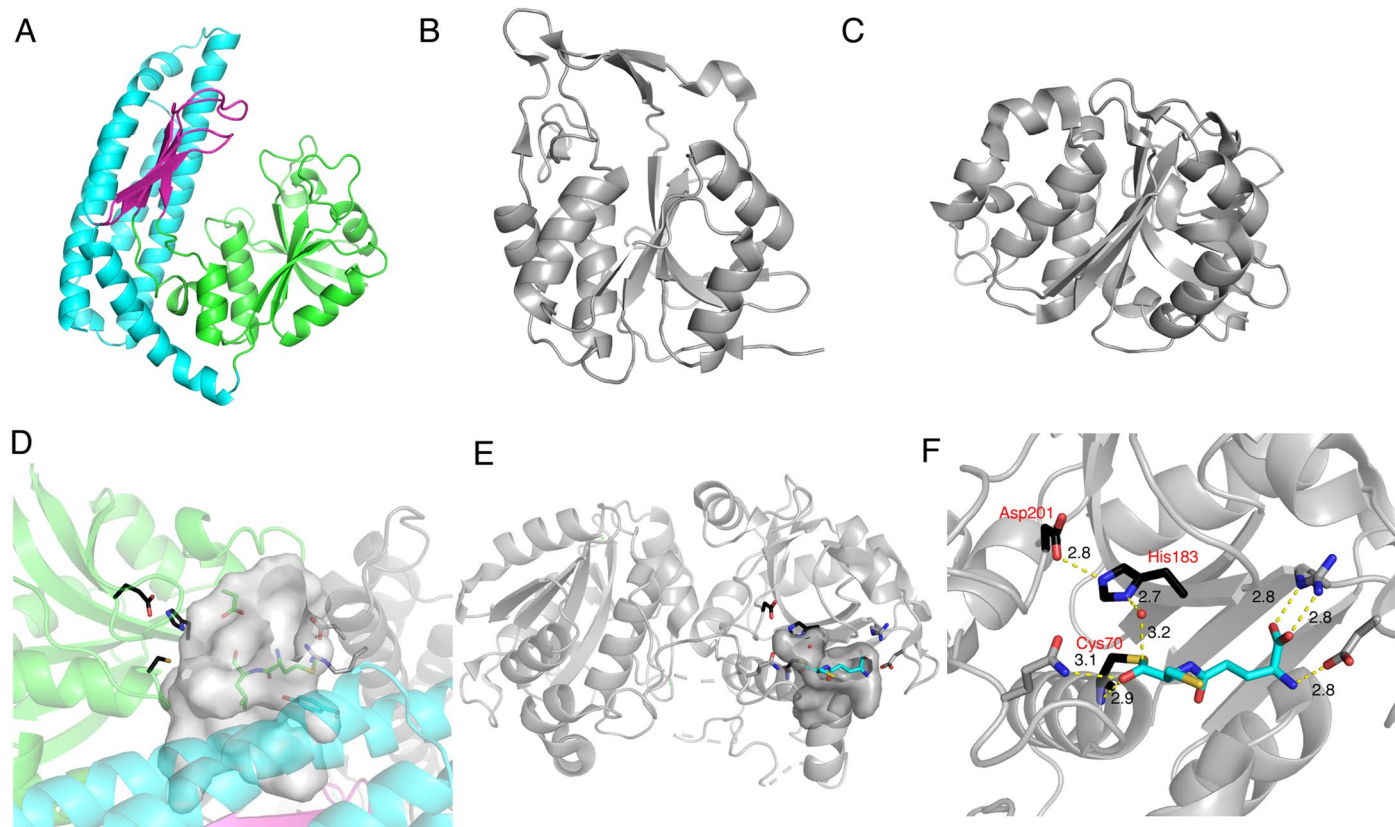


Extended Data Fig. 3 | In vitro enzyme reaction of CcbD with non-proline acyl-substrates. (A) LC-MS charts of the enzyme reactions of CcbD with various acyl-CPs. **(B)** Structures of products. Blue and orange charts represent EIC = 254 and EICs of product masses, respectively.

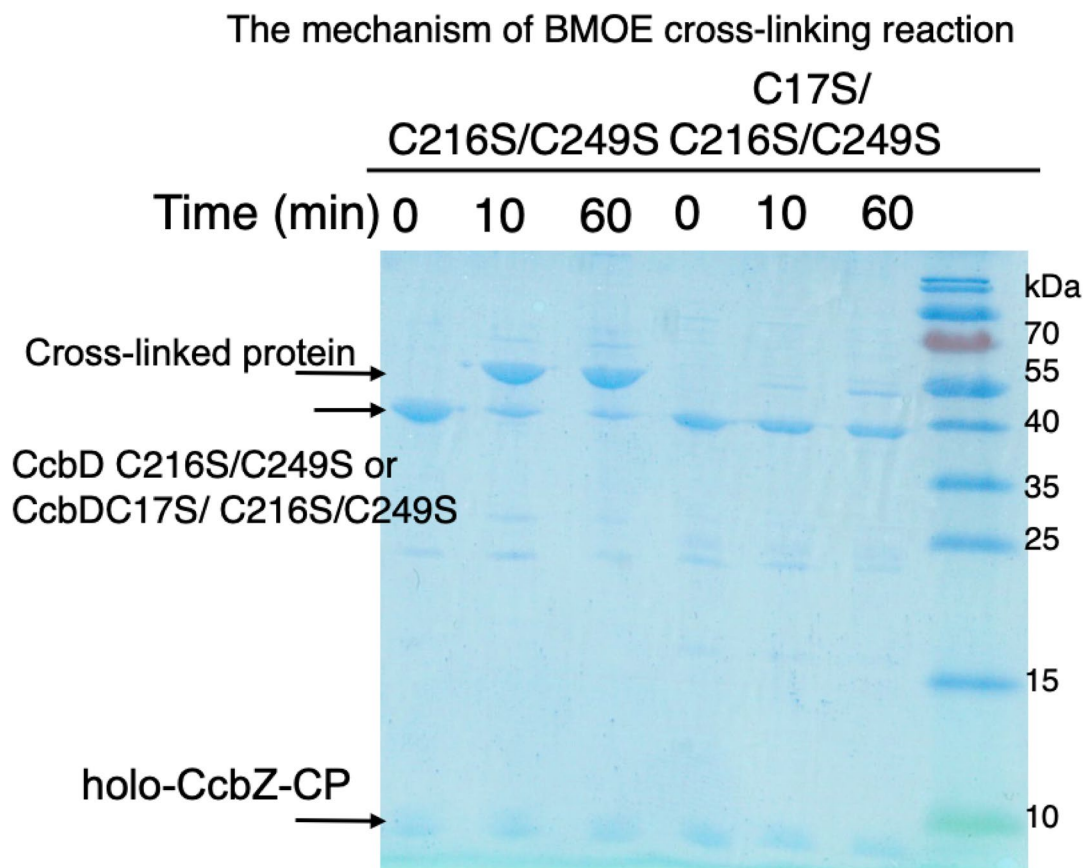


Extended Data Fig. 4 | In vitro enzyme reaction of CcbD with amine-containing substrates. (A) Structures of substrates and products. **(B)** LC-MS charts of the enzyme reactions of CcbD with various amine-containing products.

Blue, orange, and gray charts represent EIC = 116, EICs of amine-containing substrates, and EICs of product masses, respectively. Hilic column and C_{18} column were used to separate the enzyme reaction products.

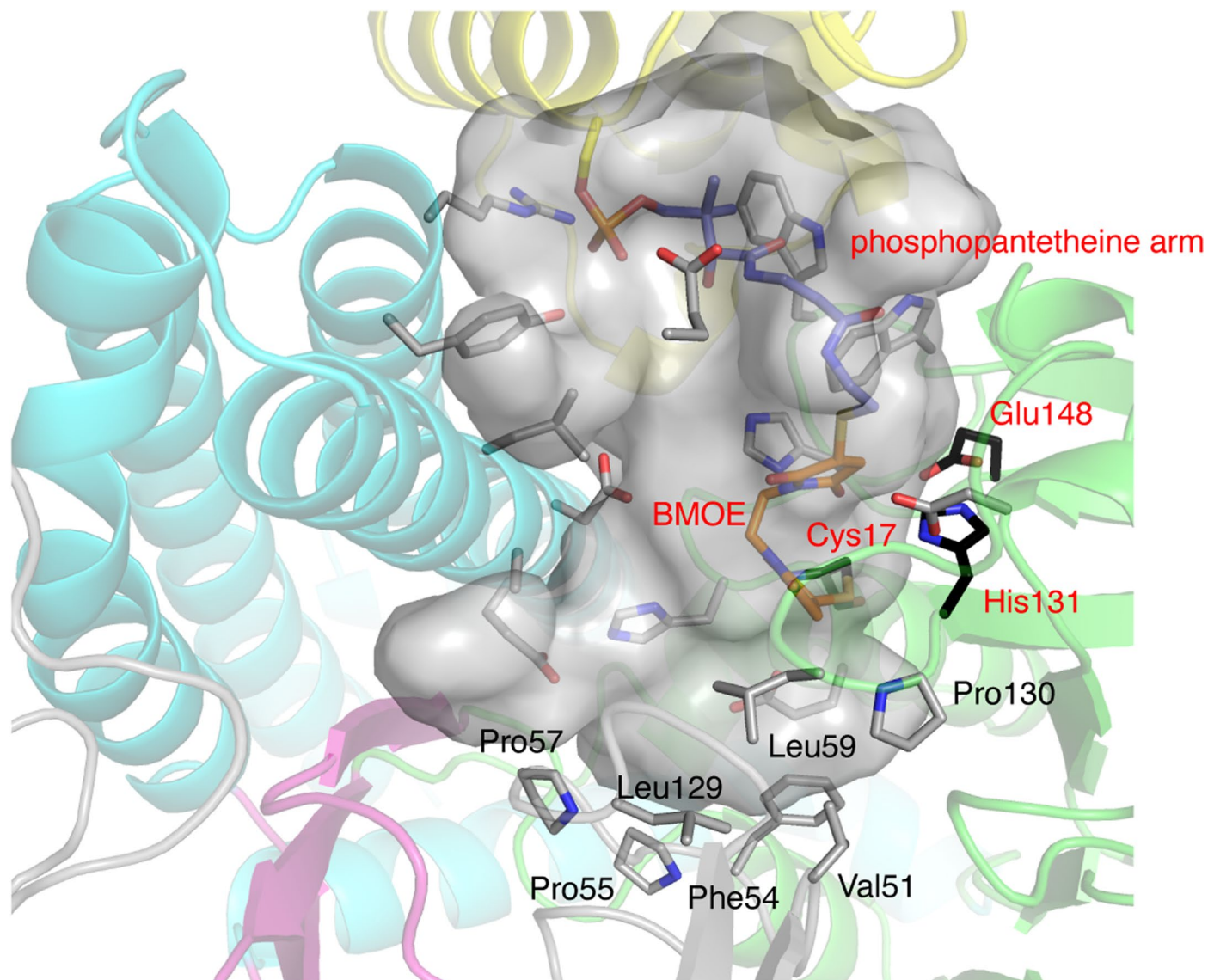


Extended Data Fig. 5 | Overall structure of CcbD. Comparison of overall structures among (A) CcbD, (B) putative C39-like peptidase protein from *Bacillus anthracis*, and (C) phytochelatin synthase-like NsPCS. (D) The active site of CcbD. (E) The dimer structure of NsPCS. (F) The active site of NsPCS.



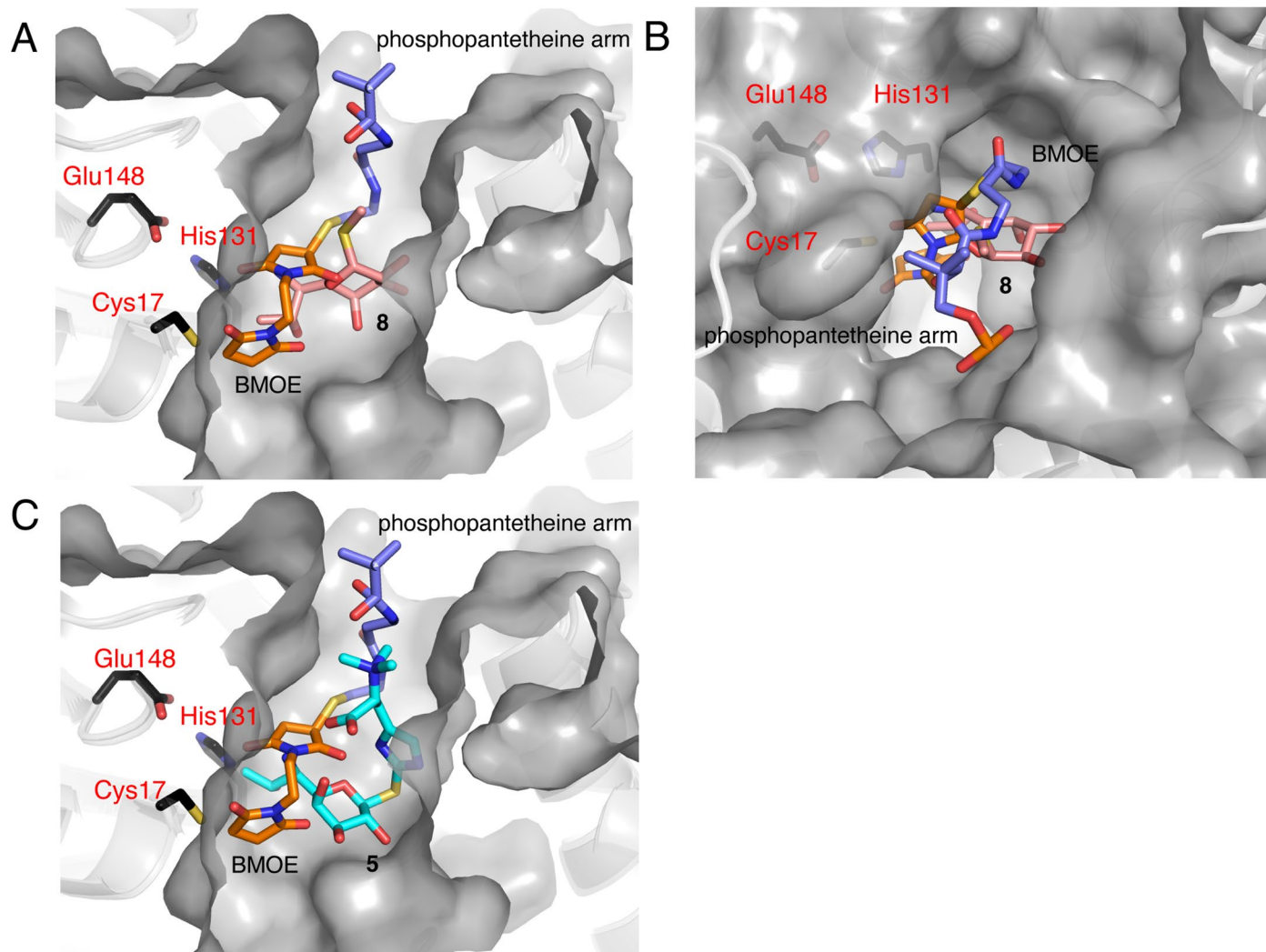
Extended Data Fig. 6 | Cross-linking reactions of CcbD C216/C249 and CcbD C17/C216/C249 variants. The incubation of the C216S/C249S variant and holo-CcbZ-CP in the presence of BMOE afforded the covalent complex (cross-linked protein). In contrast, no apparent cross-linking was observed between the C17S/

C216S/C249S variant and CcbZ-CP. PageRuler™ Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific Inc.) was used as marker. All experiments were repeated independently three times with similar results.

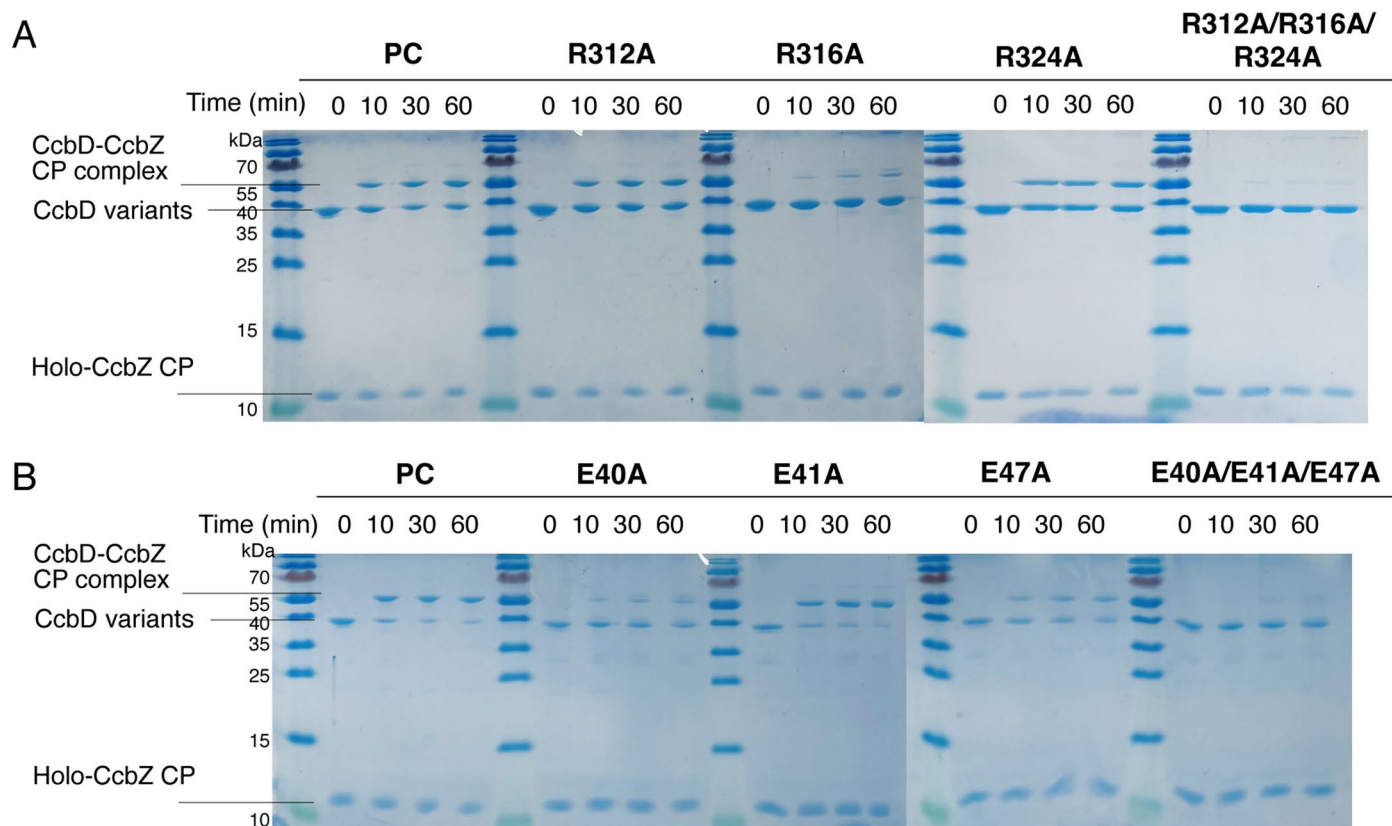


Extended Data Fig. 7 | Close-up views of the active sites of CcbD and the CcbZ-CP complex. A large space is observed deeper inside the cavity, beyond the maleimide binding site, which would be the binding site of alkyl-chain of PPL and large side chains of amino acids. The gray surface represents the active site

pocket of CcbD. Catalytic triad residues are colored black and highlighted red. Covalently binding BMOE and phosphopantetheine arm are shown as orange and purple, respectively.

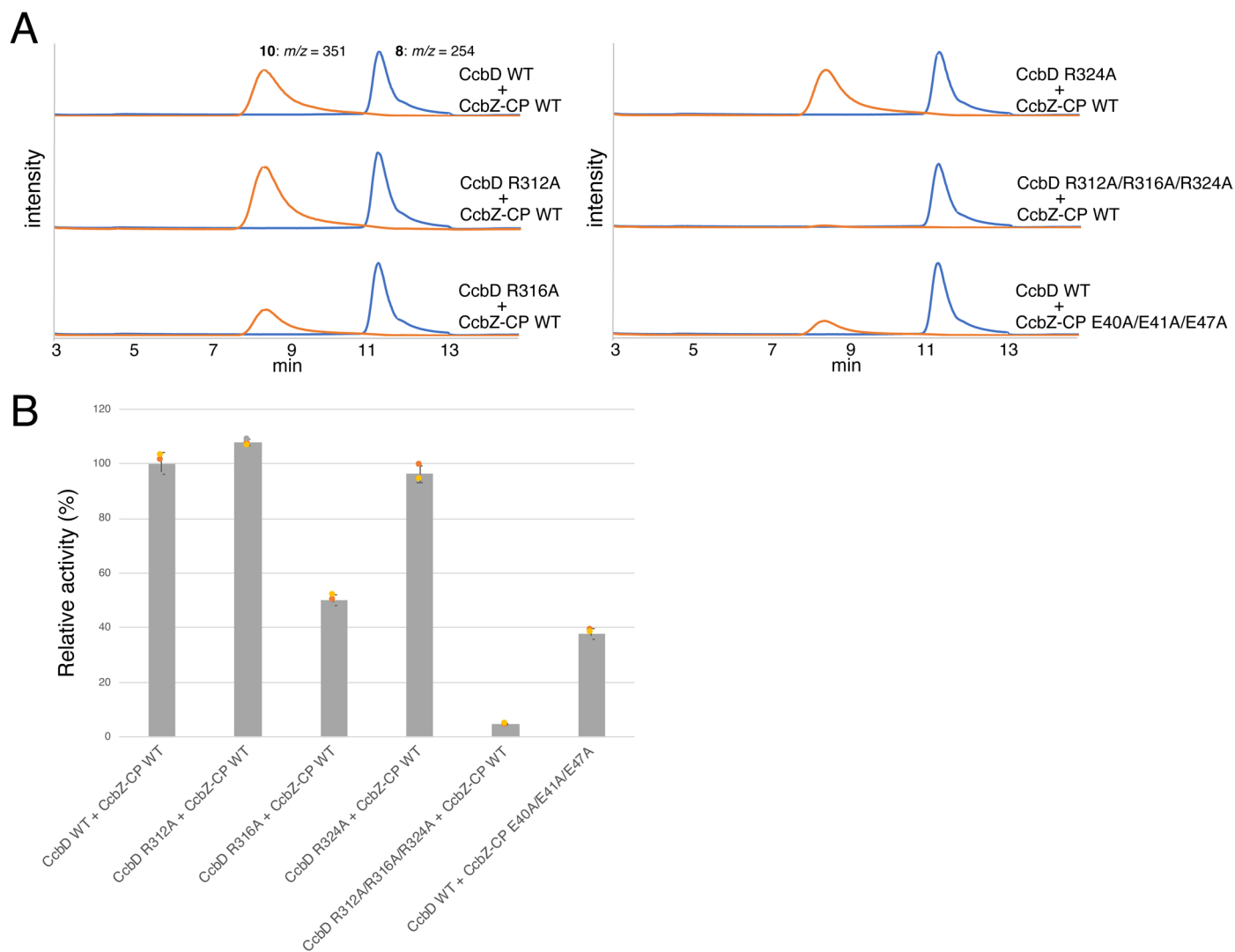


Extended Data Fig. 8 | Comparison of the binding sites of sugar substrates and phosphopantetheine moiety of CP. (A and B) Superimposed views of the complex structures of CcbD/**8** and CcbD/CcbZCP from different angles. **(C)** Docking model of **5** in the active site of CcbD/CcbZCP complex structure.



Extended Data Fig. 9 | Cross-linking reactions of the variants on the interface between CcbD and CcbZ-CP. (A) Cross-linking reactions of CcbD variants and wild type of CcbZ-CP. **(B)** Cross-linking reactions of wild type of CcbD and CcbZ-CP variants. PC is a positive control for the cross-linking reaction. PageRuler

Prestained Protein Ladder, 10 to 180 kDa (Thermo Fisher Scientific Inc.) was used as marker. All experiments were repeated independently three times with similar results.



Extended Data Fig. 10 | Mutagenesis experiments of the residues on the interface between CcbD and CcbZ-CP. (A) LC-MS charts of the enzyme reactions of CcbD and CcbZ-CP variants. (B) Relative activities of CcbD and CcbZ-CP variants. Blue and orange charts represent EIC = 254 and EIC = 351,

respectively. The bars are means of $n = 3$ independent experiments and the error bars indicate standard deviations (SD). All experiments were repeated independently three times with similar results.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input checked="" type="checkbox"/>	<input type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input checked="" type="checkbox"/>	<input type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted <i>Give P values as exact values whenever suitable.</i>
<input checked="" type="checkbox"/>	<input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input checked="" type="checkbox"/>	<input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The data supporting the findings of this study are available within the article and its Supplementary Information Files, source files or from the corresponding authors on request. The crystallographic data that support the findings of this study are available from the Protein Data Bank (<http://www.rcsb.org>). The coordinates and the structure factor amplitudes for the selenomethionine labeled CcbD, complex structures of CcbD with 8, and CcbD with CcbZ-CP were deposited under accession code 7YN1, 7YN2, and 7YN3 respectively.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender	<input type="text" value="n/a"/>
Population characteristics	<input type="text" value="n/a"/>
Recruitment	<input type="text" value="n/a"/>
Ethics oversight	<input type="text" value="na/"/>

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For all statistical experiments, including kinetics analysis and conversion rate values were determined with n=3 as it is common practice in the field, exemplified in Nature, 527, 539-547 (2015) and J. Am. Chem. Soc. 141, 9964-9979 (2019).
Data exclusions	No data are excluded.
Replication	All attempts are performed independently and successfully replicated more than three times. We stated the number of replicates for each experiment in the paper. Furthermore, the reproducibility of the assays was confirmed by including appropriate positive and negative controls.
Randomization	This is not relevant to our study because during the in vitro and structural analysis of target enzymes, the products, catalyzed reaction, and structure of enzymes were not known.
Blinding	Blinding is not relevant for this study because we use enzymatically synthesized substrates and mutant enzymes for in vitro reactions. During these analysis, the products, catalyzed reaction, and structure of enzymes were not known.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

Methods

n/a	Included in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging