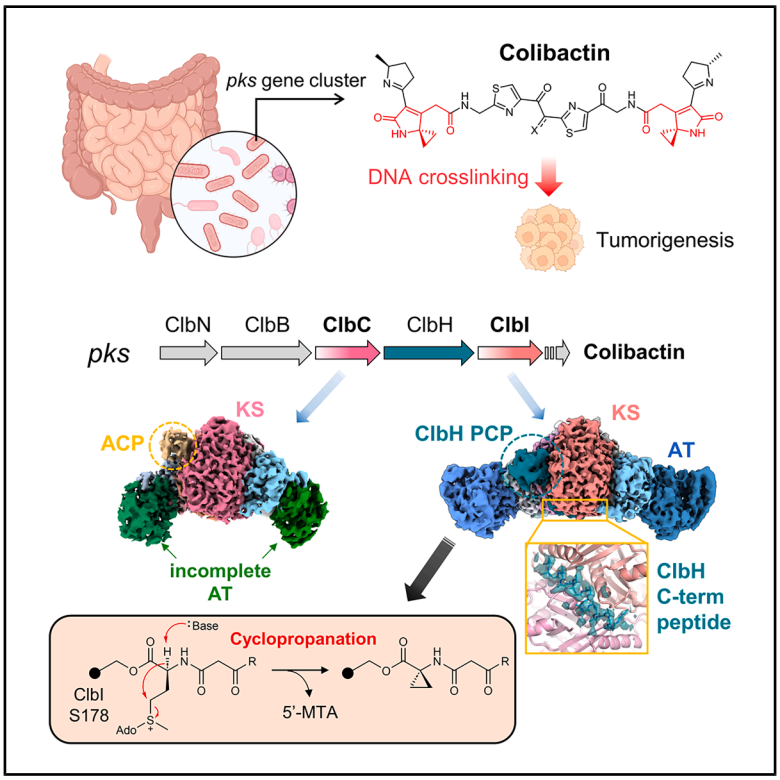
Article

Structure

Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin

Graphical abstract



Highlights

- ClbC structure reveals shared features with incomplete acyltransferase domains
- Carrier protein binding sites of ClbI were visualized using a substrate-mimic crosslinker
- ClbH uses its C-terminal peptide to recognize ClbI via a unique interaction

Authors

Minjae Kim, Jinwoo Kim, Gyu Sung Lee, ..., Seth B. Herzon, Chung Sub Kim, Jin Young Kang

Correspondence

chungsub.kim@skku.edu (C.S.K.), jykang59@kaist.ac.kr (J.Y.K.)

In brief

Colibactin is a human microbiome-derived genotoxin that promotes tumorigenesis. Kim, Kim, Lee present cryo-EM structures of the colibactin-biosynthetic polyketide synthases (PKSs), ClbC and ClbI. The structures reveal unique interactions between ClbH and ClbI and candidate residues in ClbI responsible for cyclopropanation, which is supported by complementary LC-MS analysis.

Kim et al., 2025, Structure *33*, 1–16
July 3, 2025
https://doi.org/10.1016/j.str.2025.04.017

CellPress

Structure

CellPress

Article

Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin

Minjae Kim,[1,7] Jinwoo Kim,[1,7] Gyu Sung Lee,[2,7] Paul Dominic B. Olinares,[3] Yougant Airan,[4] Jasmine L. Chow,[4] Jongseok Park,[1] Yujin Jeong,[1] Jiho Park,[1] Brian T. Chait,[3] Seth B. Herzon,[4,5] Chung Sub Kim,[2,6,*] and Jin Young Kang[1,8,*]

[1]Department of Chemistry, Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Republic of Korea
[2]Department of Biopharmaceutical Convergence, Sungkyunkwan University, Suwon, Republic of Korea
[3]Laboratory of Mass Spectrometry and Gaseous Ion Chemistry, The Rockefeller University, New York, NY, USA
[4]Department of Chemistry, Yale University, New Haven, CT 06520, USA
[5]Department of Pharmacology and Therapeutic Radiology, Yale School of Medicine, New Haven, CT 06520, USA
[6]School of Pharmacy, Sungkyunkwan University, Suwon, Republic of Korea
[7]These authors contributed equally
[8]Lead contact
*Correspondence: chungsub.kim@skku.edu (C.S.K.), jykang59@kaist.ac.kr (J.Y.K.)
https://doi.org/10.1016/j.str.2025.04.017

SUMMARY

Colibactin, a human microbiome-derived genotoxin, promotes colorectal cancer by damaging the host gut epithelial genomes. While colibactin is synthesized via a hybrid non-ribosomal peptide synthetase (NRPS)-polyketide synthase (PKS) pathway, known as *pks* or *clb*, the structural details of its biosynthetic enzymes remain limited, hindering our understanding of its biosynthesis and clinical application. In this study, we report the cryo-EM structures of two colibactin-producing PKS enzymes, ClbC and ClbI, captured in different reaction states using a substrate-mimic crosslinker. Our structural analysis revealed the binding sites of carrier protein (CP) domains of the ClbC and ClbI on their ketosynthase (KS) domains. Further, we identified a novel NRPS-PKS docking interaction between ClbI and its upstream enzyme, ClbH, mediated by the C-terminal peptide ClbH and the dimeric interface of ClbI, establishing a 1:2 stoichiometry. These findings advance our understanding of colibactin assembly line and provide broader insights into NRPS-PKS natural product biosynthesis mechanisms.

INTRODUCTION

Human microbiota produces secondary metabolites that modulate the host's immune responses, nutrient sensation, and metabolism, serving as a chemical treasury for potential therapeutic targets and drug candidates.[1] Colibactin, a genotoxin derived from the microbiota, crosslinks adenine bases of the human gut epithelial DNA using its cyclopropane group, causing cell-cycle arrest, genome instability, and tumorigenesis.[2–8] Its clinical significance in colorectal cancer, neonatal systemic infection, and inflammatory bowel diseases has been reported.[9–11] The colibactin biosynthetic gene cluster '*pks*' or '*clb*' contains 19 genes (*clbA* to *clbS*), encoding three non-ribosomal peptide synthetases (NRPSs), three polyketide synthases (PKSs), two hybrid NRPS-PKS enzymes, and other accessory proteins.[2] The *pks* gene cluster was also found in other non-human associated microbiomes such as a marine sponge, an olive tree, and honeybees.[12–14] A recent study suggested that the production of colibactin may benefit the *pks*-harboring bacteria by triggering the lytic pathway of neighboring prophages through the bacterial SOS response. This process may eliminate potential competitors, thereby enhancing the survival and proliferation of colibactin-producing bacteria.[15] Among the colibactin biosynthetic enzymes, the structures of ClbQ, ClbM, ClbS, ClbP, and ClbK have been reported.[16–21] Of these, only ClbK is directly involved in the biosynthetic reaction, while the others assist the biosynthesis or protect the bacteria from the toxicity of the molecule. However, the reported ClbK structure is incomplete and represents only a single state, limiting our understanding of colibactin biosynthesis.[20]

In this study, we investigated the two PKS enzymes in the colibactin biosynthetic pathway: ClbC and ClbI (Figure S1A). As minimal modular PKS enzymes, ClbC and ClbI consist of three domains – a ketosynthase (KS), an acyltransferase (AT), and an acyl carrier protein (ACP). In a minimal modular PKS system, the AT domain first recognizes a substrate, acyl-CoA, and transfers it to a holo-ACP (Transacylation) (Figure S1B).[22] The ACP then transfers the substrate to the conserved cysteine of the KS domain in the downstream PKS module (Translocation). The downstream ACP domain, loaded with another acyl group,

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

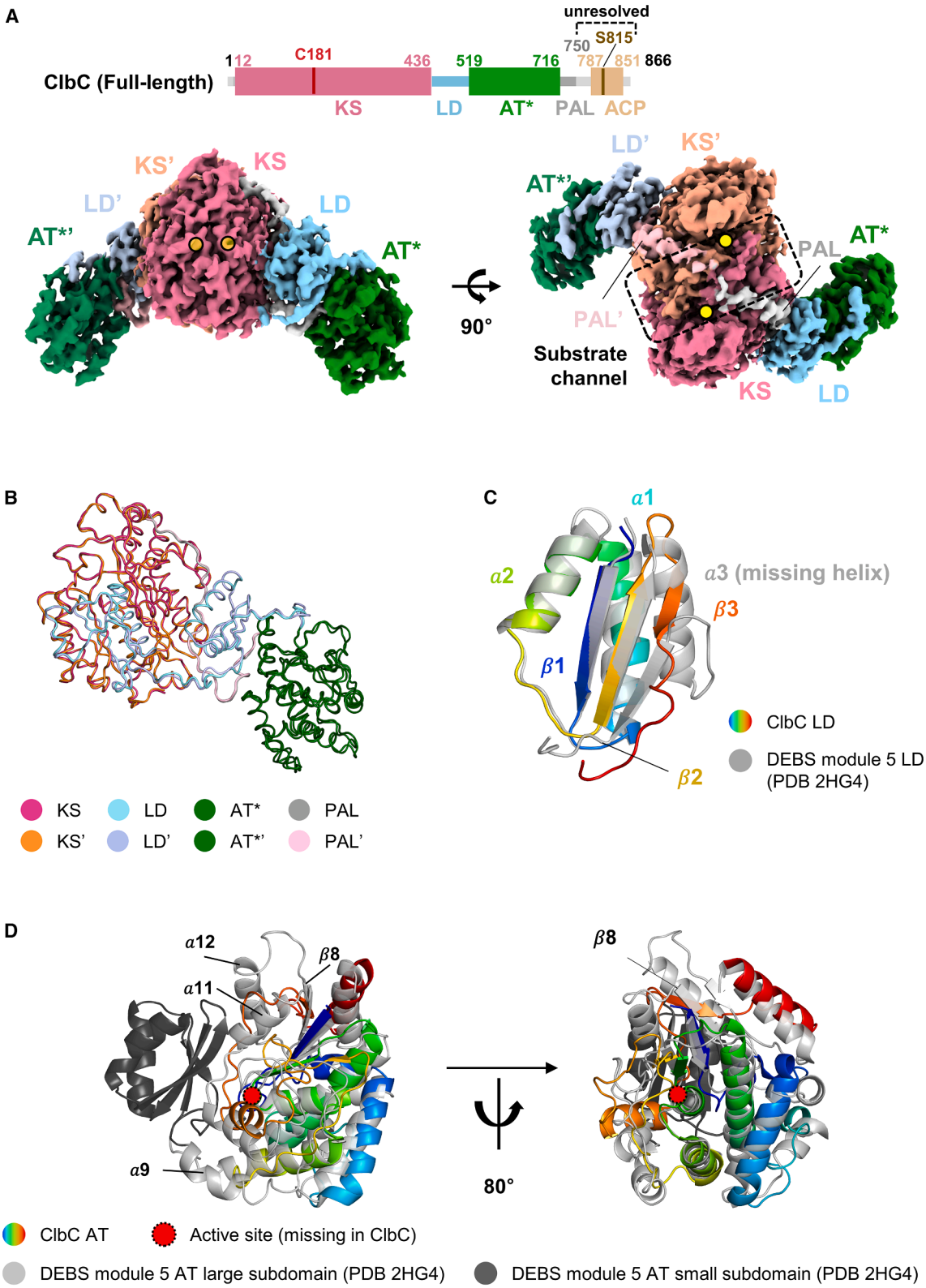


Figure 1. Cryo-EM structural analysis of apo-ClbC

(A) The domain structure and cryo-EM map of a full-length apo-ClbC, colored by domains. Abbreviations: KS – ketosynthase, LD – linker domain, AT – acyltransferase (* inactive), PAL – post-AT linker, ACP – acyl carrier protein. The map is shown in front (left) and top (right) views. Yellow circles indicate the active site

(legend continued on next page)

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

subsequently attacks the KS-tethered substrate through a decarboxylative Claisen condensation reaction (Elongation), extending the intermediate. The elongated intermediate is either transferred to the next downstream module or released by an additional thioesterase (TE) domain. Both ClbC and ClbI elongate a colibactin intermediate by a two-carbon unit using malonyl-CoA. Furthermore, ClbI generates colibactin's warhead, a cyclopropane moiety (Figure S1A).[4] Interestingly, ClbI contains a serine in the active site of its KS domain, contrary to other PKS enzymes with cysteine at this site.[23] This active site serine is suggested to play a crucial role in generating the cyclopropane group.[4] Notably, the AT domain of ClbC is inactive, yet ClbC successfully adds a two-carbon unit, presumably by utilizing AT domains of other proteins.[22]

Here, we determined cryo-electron microscopy (cryo-EM) structures of ClbC and ClbI in various states using a substrate-mimicking crosslinker. Our structural analyses of ClbC and ClbI revealed the binding sites of carrier protein (CP) domains for translocation and elongation reactions and, furthermore, a novel docking interaction between ClbI and ClbH, an upstream NRPS enzyme of ClbI. Additional liquid chromatography-mass spectrometry (LC-MS) analysis of colibactin precursors produced with ClbI mutants based on the structures further identified potential candidate residues that may be involved in the cyclopropanation of the colibactin precursor.

RESULTS

Apo-ClbC contains an inactive AT domain

We first determined the cryo-EM structure of the full-length ClbC in an apo state (without the phosphopantetheine group) at 3.1 Å resolution and built the coordinates using a homology model generated by SWISS-MODEL and a monomer model by AlphaFold2 (Figures 1A, S2, and S3; Table 1).[24–26] After protein purification, we confirmed the phosphopantetheinylation (ppantylation) activity using a BODIPY-CoA loading assay.[27] Native mass spectrometry (nMS) analysis also verified that ClbC is a homodimer, and the ppantylation reaction was completed (Figure S2C).[27,28] The cryo-EM structure of apo-ClbC revealed its homodimeric structure like typical type I PKS enzymes, having the KS-KS dimerization interface spanning a ~2,560 Å^2 interface (Figure S2D).[29,30] The ACP density was not observed, presumably due to its mobility. The dimer interface residues were highly conserved (Figure S2E),[31] and the active sites of both protomers are in a single 'substrate channel,' where the translocation and elongation reactions occur (Figure 1A).[32,33] The two ClbC protomers exhibit highly similar structures, having an r.m.s.d.(C_α) of 0.86 Å between the two protomers. However, applying C2 symmetry during the cryo-EM data analysis did not improve the map, implying the protomers have subtle structural differences (Figure 1B). 3D variability analysis of the apo-ClbC revealed a relatively larger swinging motion of linker domain (LD)-AT domains in both symmetric and asymmetric directions compared to the central KS domains, explaining the lower local resolution of LD-AT regions (Video S1 and Figure S3D).

As expected from its sequence, ClbC has a partial AT domain and an incomplete LD. Compared to a typical type I PKS LD that contains a 3α3β fold, ClbC LD lacks the last α-helix (Figure 1C).[29] The canonical AT domain comprises a large subdomain with an α/β hydrolase fold and a small subdomain with a ferredoxin-like fold inserted within the large subdomain.[34] However, the ClbC AT lacks the entire small subdomain, and its large subdomain of α/β hydrolase-fold is missing three α helices and one β strand (Figure 1D). Furthermore, two canonical active site motifs, the GHSxG motif for the acyl-*O*-ester formation and the HAFH motif for malonyl-CoA recognition, were absent in the ClbC AT, confirming its inactivity.[3,14,27] While ClbC is known to utilize other ATs, presumably from ClbB, ClbI, or FabD, to incorporate a two-carbon unit, how ClbC interacts with these enzymes and the role of this inactive AT remain unknown.[27]

A substrate-mimic crosslinker revealed ClbC ACP binding on ClbC KS

To capture the elongation state where the malonyl-loaded ClbC ACP binds to the ClbC KS, we utilized a malonyl-loaded ppant-mimic crosslinker harboring a *trans*-β-chloroacrylamide moiety (Figure 2A).[35] For crosslinking, the crosslinker and Coa enzymes were added to the full-length ClbC and incubated at room temperature for 30 min. During incubation, the crosslinker was attached to the active serine (S815) of the ClbC ACP and subsequently reacted with the active cysteine (C181) of the ClbC KS domain, forming an ACP-bound ClbC complex (see STAR Methods section for details) (Figure S4A). To verify the specificity of the crosslinking, various ACP/PCP domains loaded with the same crosslinker were incubated with ClbCΔACP (ClbC lacking its ACP domain). As a result, crosslinking was only observed with its original CP partners, ClbC ACP and ClbB ACP (Figure S4B).

A 3.9 Å cryo-EM map of the ACP-bound ClbC complex revealed that a single ACP is bound to the upper region of the ClbC substrate channel opening (Figures 2B and S5). Due to the low local resolution of the ClbC ACP, we generated the ClbC ACP model using AlphaFold2 and fit it into the density map with some modifications using the ppantylation helix (helix II) as a reference (Figure 2C). Mutants with alanine substitutions at the ACP-KS interface showed reduced crosslinking, particularly in the D144A/Q221A and D144A/M224A, indicating the significance of these residues (Figure S4C). We were unable to determine which protomer the ACP belongs to because it is

cysteine residues in the KS domains. They are semi-transparent in front view due to their positioning behind the protein. The black dotted box (top view) highlights the substrate channel. The unresolved region of the protein is marked in the domain scheme. (See also Figures S2 and S3).

(B) The Cα traces of the apo-ClbC protomers are superimposed by the KS domains. Each domain is colored as indicated. The r.m.s.d (Cα) of the two protomers is 0.86 Å.

(C) The LD domains of the apo-ClbC and the DEBS module 5 (PDB 2HG4) are superimposed. The apo-ClbC LD and the DEBS module 5 LD are colored in rainbow (N-terminal: blue, C-terminal: red) and gray, respectively.

(D) The AT domains of the apo-ClbC and the DEBS module 5 (PDB 2HG4) are superimposed. The apo-ClbC AT is colored in rainbow (N-terminal: blue, C-terminal: red), and the large and the small subdomains of the DEBS module 5 AT are colored in gray and dark gray, respectively. A dotted red circle marks the active site of the AT domain. (See also Figure S13).

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

Table 1. Cryo-EM data acquisition and refinement parameters

Sample	apo-ClbC	ACP-bound ClbC	apo-ClbI	ACP-bound ClbI	ClbH-bound ClbI
EMDB	EMB-38222	EMB-38405	EMB-38406	EMB-38410	EMB-38411
PDB	8XBL	8XJT	8XJU	8XJY	8XJZ
Data collection and processing					
Data collection Facility	NCCAT	IMP	NCCAT	NCCAT	IMP
Microscope	Titan Krios	Titan Krios G4	Titan Krios	Titan Krios	Titan Krios G4
Voltage (kV)	300	300	300	300	300
Detector	Gatan K3	BioQuantum K3	Gatan K3	Gatan K3	BioQuantum K3
Electron exposure ($e^-/Å^2$)	58.85	50	58.85	50	50
Defocus range (μm)	−0.8–−2.0	−0.8–−2.4	−0.8–−2.0	−0.8–−2.5	−0.8–−2.4
Data collection mode	counting	counting	super-resolution	super-resolution	super-resolution
Pixel size (Å)	0.83	1.090	0.825	1.067	1.0902
Symmetry imposed	C1	C1	C1	C1	C1
Number of total movies	12,611	13,070	11,100	15,349	7,283
Initial particle images (no.)	4,359,102	6,827,074	6,576,361	11,180,913	3,785,870
Final particle images (no.)	295,478	301,978	1,307,444	362,401	246,313
Map resolution (Å) - FSC threshold 0.143	3.06	3.88	2.88	3.29	3.67
Refinement					
Initial model used	Homology/AlphaFold	apo-ClbC	Homology/AlphaFold	apo-ClbI	apo-ClbI
Map sharpening B factor ($Å^2$)	−88.09	−175.1	−108.8	−137.9	−111.5
Model composition					
Non-hydrogen atoms	11,509	11,954	12,957	14215	13,641
Protein residues	1,497	1,555	1,711	1,879	1,805
Nucleic acid residues	0	0	0	0	0
Ligands	0	0	0	0	0
B factors ($Å^2$)					
Protein	55.67	57.01	40.04	50.03	41.80
R.m.s. deviations					
Bond lengths (Å)	0.011	0.014	0.010	0.006	0.011
Bond angles (°)	1.138	1.284	1.344	1.060	1.195
Validation					
MolProbity score*	1.85	1.77	1.73	1.66	1.76
Clashscore	7.37	3.91	5.80	4.51	3.92
Poor rotamers (%)	0	0	0	0	0
Ramachandran plot					
Favored (%)	93.03	88.43	93.78	93.34	99.94
Disallowed (%)	0	0.06	0	0.05	0

positioned at a distance that allows either protomer to connect via a 29-aa peptide linking the C-terminus of ClbC AT (G748) to the N-terminus of ClbC ACP (F777).

When the structures of ACP-bound ClbC and apo-ClbC complexes were superimposed by their KS domains, a slight expansion of the overall dimer was observed in the ACP-bound ClbC complex (Figure 2D; Table S1). This expansion occurred in both protomers, suggesting that the conformational change induced by an ACP binding on one protomer is propagated to the other, either directly through the contact between the bound ACP and both protomers or indirectly via the dimeric interface. Furthermore, this ACP binding-induced expansion enlarges the volume of the active site pocket in the substrate channel from 1192 $Å^3$ (apo) to 1723 $Å^3$ (ACP-bound), presumably facilitating the inclusion of the precursor tethered to the bound ACP.[36]

ACP binding induces a loop-folding of ClbI KS

Next, we determined the cryo-EM structures of apo- and cross-linked ClbI complexes using the same substrate-mimic crosslink probe. The 2.9 Å-map of the full-length apo-ClbI revealed a typical dimeric structure characteristic of type I PKS as expected (Figures 3A, S6, and S7; Table 1). The ACP domain and a loop (135–153 in chain A, 135–151 in chain B) located at the opening of the substrate channel were not visible due to their high

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

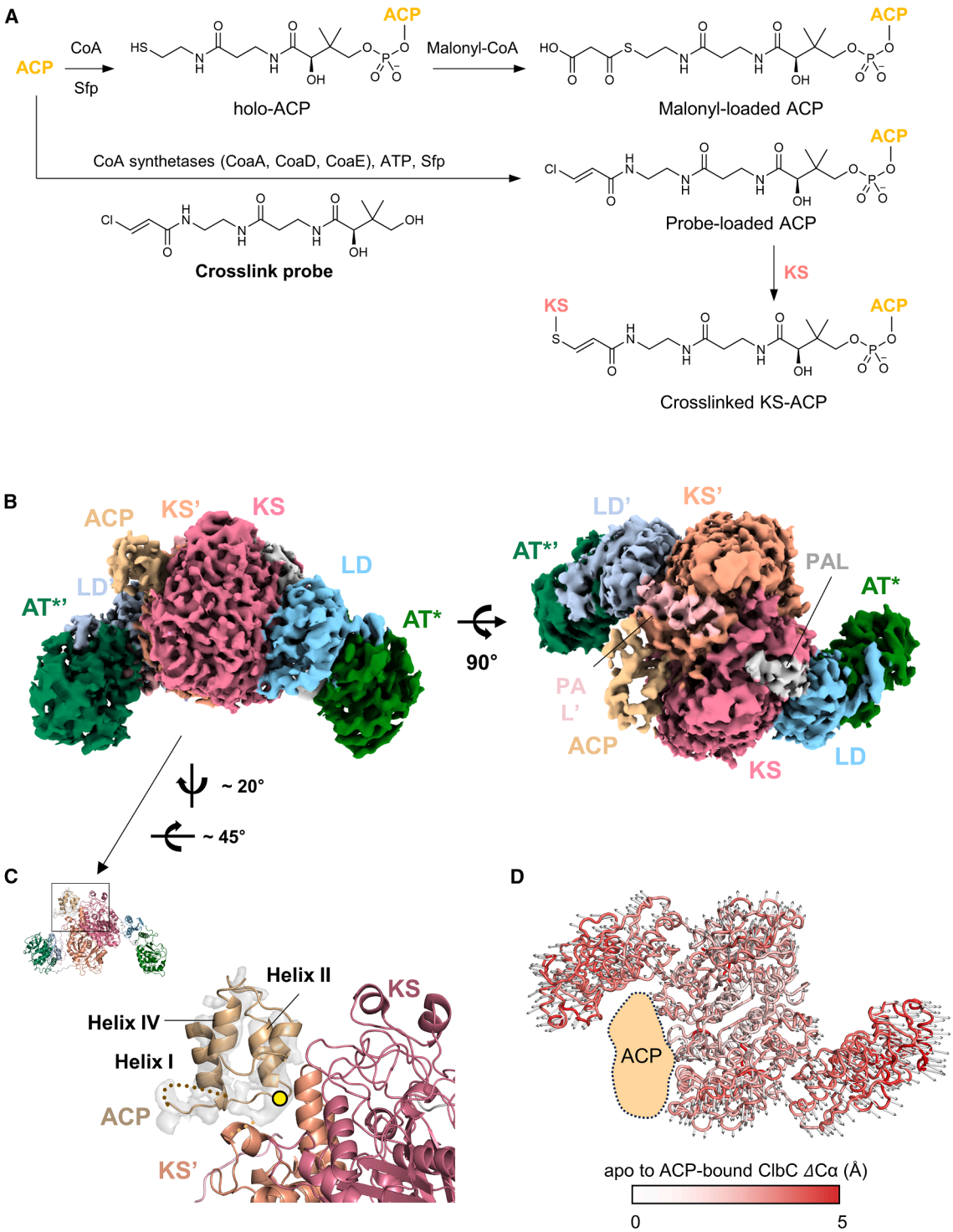


Figure 2. Cryo-EM structure of ACP-bound ClbC

(A) Chemical structures of the ppant arm in holo-ACP, the malonyl-loaded ppant arm in the malonyl-loaded ACP, a ppant-mimic crosslinker with a *trans*-β--chloroacrylamide group that we use in this study,[35] and its crosslinked structure between KS and ACP domains.

(B) The local-filtered cryo-EM map of the ACP-bound ClbC. Domains are colored as indicated. (See also Figures S4 and S5).

(legend continued on next page)

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

mobility. Similar to the apo-ClbC complex, the residues at the dimeric interface were more conserved than those not involved in the interface (Figures S6C and S6D).[30,31] 3D variability analysis of the apo-ClbI revealed more rigid-body-like movements than that of apo-ClbC, presumably due to the more robust connection of LD and AT to the KS domain. In addition, a symmetric "breathing motion" between the two protomers was observed, which may expand the opening and inner volume of the substrate channel (Video S2).

Then, we generated a crosslinked complex between ClbI ACP and ClbI(S178C)ΔACP (hereafter referred to as the 'ACP-bound ClbI complex') by using the substrate-mimic crosslinker to capture ClbI in its elongation state (Figures 3B, S8A, and S8B). ClbI(S178C)ΔACP lacks the ACP and contains a single mutation, S178C, necessary for crosslinking with a thiol-reactive probe. Crosslinking assays with various CPs confirmed the specificity of the crosslinking reaction (Figure S8C). Surprisingly, the 3.3 Å-resolution map of the ACP-bound ClbI complex revealed two ACPs bound to both protomers of the ClbI dimer (Figures 3B and S9; Table 1). This was unexpected because previous studies suggested that only one protomer of a PKS dimer is involved in the reaction at a given time.[33] To investigate whether this conformation was an artifact of using separate ACP and KS-AT domains, we collected an additional cryo-EM dataset of a full-length, crosslinked ClbI complex, but it also exhibited two ACP-bound ClbI dimers, suggesting that the dual ACP binding is not an artifact from using separate KS-AT and ACP domains (Figure S8E). However, it remains possible that the crosslinking of two ACPs to a ClbI dimer results from the accumulated crosslinking events rather than reflecting the native reaction process, considering the low statistical probability of simultaneous binding of two ACPs under physiological conditions.

In the ACP-bound ClbI complex, the ACP binding sites on the KS domains were highly similar between two protomers, displaying a 2-fold pseudosymmetry (r.m.s.d.(Cα) = 0.45 Å), indicating the ACP binding specificity (Figure S8F). However, the locations and orientations of the bound ACPs differed between ClbC and ClbI, underscoring the diversity of ACP binding modes. Notably, while the extended length of the crosslinker is approximately 20 Å, the distance between the ClbI KS S178 oxygen and the ACP S952 oxygen was about 13 Å, suggesting that the ppant group might be sequestered in the ACP cavity, as observed in some cases,[32–34] or significantly bent within the substrate channel (Figure 3C).[37–39] We also observed elongated density at the KS active site S178 and the ACP active site S952 (Figure S9H). While this density supports the presence of substrate-mimic crosslinking, it was not of sufficient quality for accurate modeling of the crosslinker.

Interestingly, we found that a loop at the substrate channel opening, which was disordered in the apo-ClbI, folded into two helices upon ACP binding (Figure 3D). This conformational change increased the ClbI-ClbI dimeric interface area from ~1,760 Å^2 (apo-ClbI) to ~2,240 Å^2. Furthermore, the folded helices provided an additional binding surface of ~490 Å^2 between the ACP and the non-reacting ClbI KS, which is larger than the interface of ~310 Å^2 between the ACP and the reacting ClbI KS. This suggests that ACP binding to the substrate channel is significantly facilitated by the other protomer that does not participate in the reaction, highlighting the role of the PKS dimeric structure. This ACP binding-induced loop folding was also observed in the cryo-EM structure of the holo-Lsd14 in the condensation state.[33]

ClbH-bound ClbI complex revealed a novel NRPS-PKS docking interaction

ClbH receives the colibactin intermediate from ClbC, adds the unusual amino acid substrate *S*-adenosyl-L-methionine (SAM) to it, and then transfers the elongated intermediate to ClbI for cyclopropane formation (Figure S1A). To capture the translocation state of ClbI, where the upstream ClbH PCP (peptidyl carrier protein) transfers the intermediate to the downstream ClbI, we prepared a full-length, crosslinker-loaded ClbH and formed the ClbH-crosslinked ClbI complex (hereafter referred to as 'ClbH-bound ClbI') using the same substrate-mimic crosslinker (Figures 4A, S10, and S11; Table 1).

A 3.7 Å-resolution map of the ClbH-bound ClbI complex revealed only a single ClbH PCP domain bound to the lower side of the ClbI substrate channel, with no other ClbH domains visible. This suggests that other ClbH domains remain mobile (Figure 4B). NRPS PCPs are homologous with PKS ACPs, both containing four helices and being modified by ppantylation.[40] The location of the ClbH PCP binding is unique as most CPs in known structures bind to the upper region of the KS, except for ACP1 in Mtb Pks13, which also binds to the lateral side of Pks13.[41] The ClbH PCP forms a ~400 Å^2 interface with the reacting KS but has no interaction with the non-reacting KS, in contrast to the ClbI ACP (Figure 3B). The mobile loop folded in the ACP-bound ClbI complex was not visible in the ClbH-bound ClbI complex as in the apo-ClbI, presumably because the bound PCP and the loop do not interact with each other. Similar to the ACP-bound ClbI complex, the side chains of ClbH PCP S1524 and ClbI KS S178 exhibited extended densities, likely originating from the crosslinker connecting the two residues (Figure S12A). The distance between the ClbH S1524 side chain oxygen and the ClbI S178 side chain oxygen was ~18 Å, comparable to the length of the crosslink probe contrary to the sequestered ClbI ACP (Figure 4D). In addition, the ClbH-bound ClbI exhibited an expanded conformation compared to the apo-ClbI as observed in ACP-bound ClbC (Figure 4E; Table S1).

Notably, we found an unexpected helical density at the bottom side of the ClbI dimeric interface (Figure 4C). Given its proximity to the modeled ClbH PCP, the unidentified helical density

(C) The zoomed ACP region of the ACP-bound ClbC. The local-filtered map of the ACP region, shown in gray, and the model for the ACP and the neighboring region are drawn. A dotted line shows the unmodeled loop between helix I and helix II. A yellow circle marks the active serine of the ACP. The orientation of the figure compared to the Figure 2B is shown by a black box in the top left corner.
(D) Structural changes in ACP-bound ClbC compared to apo-ClbC. Apo-ClbC is shown in loop format and colored by Cα distance differences (ACP-bound ClbC vs. apo-ClbC) after KS dimer superposition. Gray arrows (scaled by a factor of 2) indicate the direction and magnitude for the Cα(ACP-bound ClbC)-Cα(apo-ClbC) shifts. The average ΔCα is 2.2 Å (±1.0 Å SD), with a range of 0.12–6.0 Å. (See also Table S1).

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

CellPress

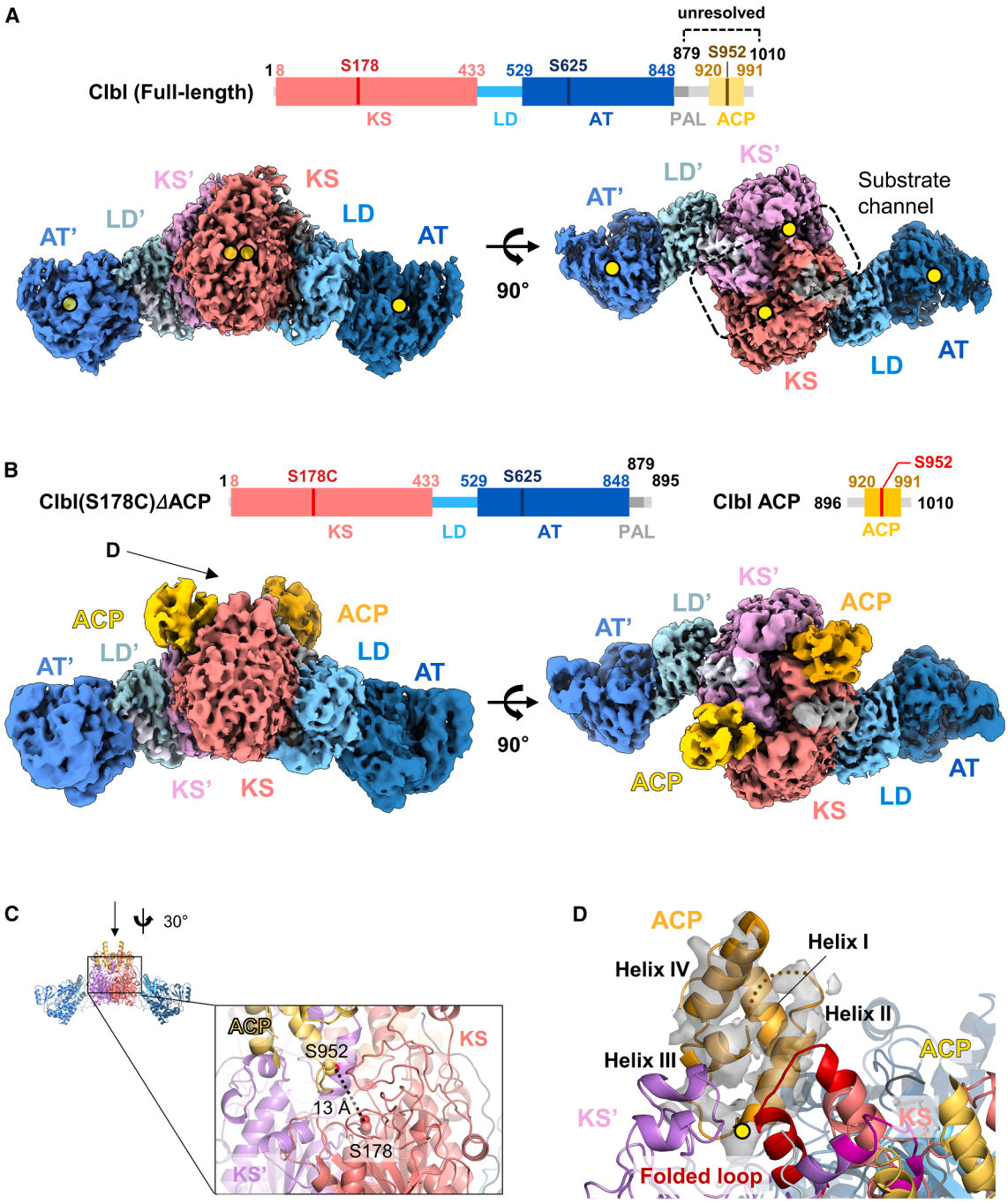


Figure 3. Cryo-EM structures of apo-ClbI and ACP-bound ClbI

(A) Domain structure of full-length ClbI (top) and the local-filtered cryo-EM map of the apo-ClbI (bottom), colored by domains and shown in front (left) and top (right) views. Yellow circles indicate the locations of the active site serines in the KS and AT domains; three are semi-transparent in the front view as they are behind the protein. (See also Figures S6 and S7).

(B) The ClbI(S178C)ΔACP and ClbI ACP constructs used in the assembly of the ACP-bound ClbI complex are shown above the local-filtered cryo-EM map of the ACP-bound ClbI. The map is colored by domains and shown in front (left) and top (right) views. (See also Figures S8 and S9) A black arrow (labeled as D) indicates the view of Figure 3D.

(C) ACP-bound ClbI, in front view of Figure 3B, is rotated by ~30° along a vertical axis and zoomed in to show the distance between the ACP active serine S952 oxygen and the KS active serine S178 oxygen, which is 13 Å in both protomers.

(D) The local-filtered map of the ACP region superimposed with the models for the ACP and neighboring region in the ACP-bound ClbI, as in Figure 3B. The unmodeled loop between helix I and helix II is shown as a dotted line. The active site serine of the ACP is marked with a yellow circle. The loop folded upon ACP binding is colored red and labeled. The folded loop in the other protomer is colored in magenta but not labeled for clarity.

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

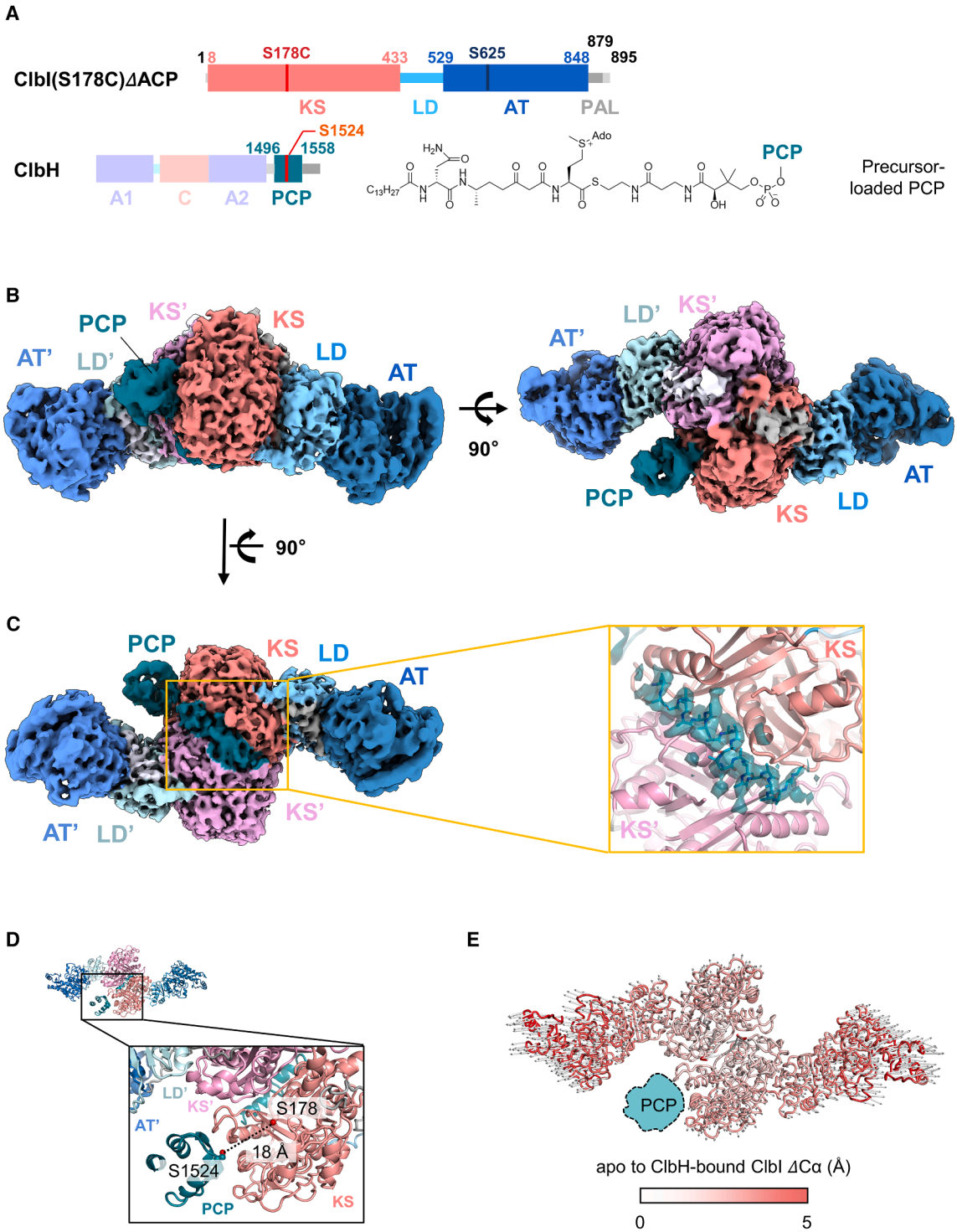


Figure 4. Cryo-EM structure of ClbH-bound ClbI

(A) The ClbI and ClbH constructs used in the ClbH-bound ClbI complex formation and the chemical structure of the precursor-loaded ClbH PCP ppant group are drawn. Abbreviation: A1 – adenylation 1, C – condensation, A2 – adenylation 2, PCP – peptidyl carrier protein.

(B) The local-filtered cryo-EM map of the ClbH-bound ClbI is colored by domains and shown in front (left) and top (right) views. (See also Figures S10, S11, and S12A).

(C) From the bottom view of the ClbH-bound ClbI, a helical density, colored in teal like ClbH PCP and located at the dimeric interface, is zoomed and redrawn in a surface format and superimposed by the alanine α-helix shown in a stick format. (See also Figure S12B).

(legend continued on next page)

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

likely originates from the unmodelled C-terminal peptide of ClbH (P1563-H1598). In some NRPS systems, the C-terminal region of the PCP contains a COM (COMmunication or COmmunication-Mediating) domain, which facilitates the recognition of the downstream NRPS module.[42] Based on this, we hypothesized that the observed density similarly arises from the C-terminal region of ClbH PCP and plays a role in recognizing the downstream enzyme, ClbI. To test this hypothesis, we generated ClbH mutants lacking the C-terminal peptide and assessed crosslinking between ClbI(S178C)ΔACP and the mutants (Figure 5A). Interestingly, the ClbH mutants missing the C-terminal peptide were unable to crosslink with ClbI, suggesting that the C-terminal peptide is essential for the ClbH-ClbI interaction. To further validate that the density belongs to the C-terminal peptide of ClbH, we prepared a separate ClbH PCP domain containing the entire C-terminal peptide and formed a crosslinked complex with ClbI(S178C)ΔACP for cryo-EM structure determination. The resulting cryo-EM map of the complex also exhibited the same helical density in the same location, supporting the idea that this density originates from the C-terminal region of the ClbH PCP (Figure S12B). Furthermore, mass photometry analysis confirmed that the complex consists of a ClbH monomer bound to a ClbI dimer, confirming the unexpected helical density originates from a single ClbH molecule rather than two (Figure 5B).

ClbI secures the upstream precursor to execute cyclopropanation

The structures of ClbI complexes in various states revealed that the upstream PCP binding site does not overlap with the downstream ACP binding site on ClbI KS (Figure S12C). Notably, the presence of a serine residue at the ClbI KS active site, instead of the typical cysteine, may impose a higher energy barrier for precursor translocation. Thus, we hypothesized that upon the simultaneous binding of ClbH PCP and malonyl-loaded ClbI ACP, the precursor may be directly transferred from the upstream ClbH PCP to the downstream ClbI ACP within the ClbI KS substrate channel without covalent translocation to the KS active site serine.

To test this hypothesis, we performed an *in vitro* colibactin precursor biosynthesis reaction and analyzed the reaction products using LC-MS (Figure 6A). To investigate the translocation, cyclopropanation, and elongation of the colibactin precursor, we prepared three reaction mixtures. In the first reaction (**Reaction 1**), the reaction mixture (labeled as **R1**) contained ClbN, ClbB, ClbC, ClbH, and their respective substrates, myristoyl-CoA, L-Asn, L-Ala, malonyl-CoA, and SAM, to produce the SAM-incorporated ClbH precursor (**1,** *m/z* 864, Figure 6B). In the second reaction (**Reaction 2**), the mixture included all the components of Reaction 1, with the addition of ClbIΔACP. If the ClbI follows the canonical pathway, the ClbH precursor should be covalently transferred to the KS active site of ClbI in Reaction 2. However, if direct precursor translocation from ClbH PCP to malonyl-loaded ClbI ACP occurs as hypothesized, precursor transfer would not be observed in Reaction 2, as ClbI in this reaction lacks the ACP domain. The third reaction mixture (**Reaction 3**) contained all the components of Reaction 1, along with full-length ClbI. Regardless of the precursor translocation mechanism, this reaction should produce a cyclopropanated and elongated colibactin precursor (hereafter referred to as ClbI precursor, **4** in Figure 6B), confirming the validity of our assay.

Following overnight incubation at room temperature, the reaction mixtures of **Reaction 2** and **Reaction 3** were applied to Ni-NTA beads to separate the histidine-tagged Clb proteins from ClbI, which had its histidine tag cleaved prior to the reaction. The ClbI fractions obtained from this separation were designated as **R2-I** and **R3-I**, while the eluates from the Ni-NTA beads, containing histidine-tagged Clb proteins, were labeled as **R2-H** and **R3-H** to emphasize their inclusion of ClbH. All the fractions were extracted using ice-cold methanol. Additionally, the remaining pellet from each fraction was hydrolyzed by heat (70°C) in a basic condition, then neutralized and re-extracted to recover all hydrolyzed and protein-tethered products.

LC-MS analysis of **R1** confirmed the activities of the purified Clb proteins by detecting the SAM-containing ClbH precursor **1**, confirming the activity of the purified Clb proteins (Figure 6B). In **Reaction 2**, the ClbH precursor **1** was detected in **R2-I**, indicating that it was covalently transferred to ClbI KS. Furthermore, LC-MS analysis of **R2-I** also revealed the presence of the cyclopropane-containing precursor (**3**, *m/z* 567), while neither **1** nor **3** was detected in **R2-H.** This result suggests that the ClbH precursor is efficiently transferred to ClbI and subsequently undergoes cyclopropanation, overcoming the higher energy barrier of translocation. Lastly, both the cyclopropane-containing precursor **3** and the spirocyclic precursor **5** (*m/z* 547, Figure 6B), which derived the two carbon-elongated cyclopropane-containing precursor **4**, were detected in **R3-I**, confirming the cyclopropanation and elongation activity of ClbI. In summary, LC-MS analysis supports the translocation of the ClbH precursor to ClbI KS, followed by cyclopropanation within ClbI before elongation, which contrasts with our initial hypothesis. This outcome aligns with the prior report that the ClbI S178A mutation abolishes cyclopropane formation of the colibactin precursor, as the mutation disrupts the translocation process, which is a prerequisite for cyclopropanation.[4]

Utilizing this method, we further investigated the function of the ClbH C-terminal peptide in the colibactin precursor synthesis. In the experiment, we again conducted the one-pot reaction with Clb proteins but used either WT ClbH or Mut1 ClbH lacking the C-terminal 12-aa peptide. To prevent precursor elongation, we used ClbIΔACP. The reaction mixtures were again separated into ClbH and ClbI fractions, and the precursors were detected by LC-MS (Figure 6C). In the WT ClbH-containing reaction mixture, the majority of the ClbH precursor **2** (*m/z* 615, Figure 6B), a de-adenosylated derivative of ClbH precursor **1**,

(D) From the top view of the ClbH-bound ClbI structure in Figure 4B, shown in the left top corner, the boxed region is zoomed in to show the distance between the ACP active site S1524 and the KS active site S178. The distance between S1524 oxygen and S178 oxygen is 18 Å, comparable with the distance in a free ppant moiety of ~20 Å.

(E) Structural changes in ClbH-bound ClbI compared to apo-ClbI. Apo-ClbI is shown in a loop format and colored by Cα distance differences (ClbH-bound ClbI vs. apo-ClbI) after KS dimer superposition. Gray arrows (scaled by a factor of 2) indicate the direction and distance of the Cα(ClbH-bound ClbI)-Cα(apo-ClbI) shifts. The average ΔCα is 2.3 Å (±1.2 Å SD) with a range of 0.20–8.5 Å. (See also Table S1).

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

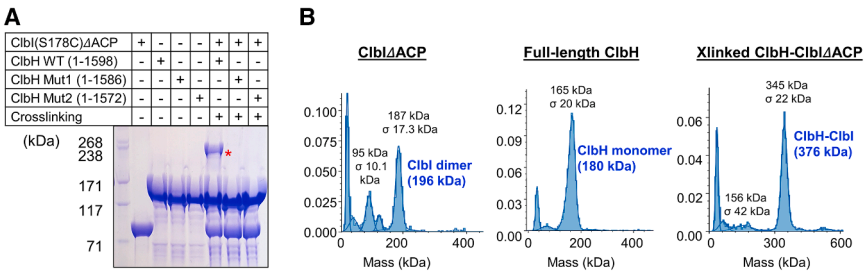


Figure 5. Characterization of ClbH-bound ClbI

(A) SDS-PAGE analysis of crosslinking between ClbI(S178C)ΔACP and wild-type ClbH or ClbH mutants lacking C-terminal peptide. Only wild-type ClbH crosslinked with ClbI, while ClbH mutants lacking either 12-aa (Mut1) or 26-aa (Mut2) C-terminal peptide failed. The crosslinked complex is marked with a red asterisk (*).

(B) Mass photometry analysis of ClbIΔACP, full-length ClbH, and their crosslinked complexes, revealing the stoichiometry between ClbIΔACP and ClbH in the crosslinked complex is 2:1.

was detected in the ClbI fraction, indicating efficient precursor translocation from ClbH to ClbI. However, with the ClbH mutant, more than two-thirds of the precursor remained in the ClbH fraction. Additionally, no further processed precursor was detected in any of the samples, confirming that the observed results reflect the precursor transfer status. These results demonstrate the critical role of the ClbH C-terminal peptide in the ClbH precursor transfer.

The formation of the cyclopropane group in the colibactin precursor is likely to involve α-hydrogen abstraction followed by the release of 5′-methylthioadnosine (MTA) (Figure 6D, top).[43] To understand the molecular mechanism of cyclopropanation based on the structure, we selected residues that are situated on the inner surface of the substrate channel in ClbI and may extract α-hydrogen for cyclopropanation. These residues were substituted with alanines, and the mutants were introduced to *in vitro* colibactin biosynthesis for the LC-MS analysis as in previous experiments (Figure 6A). Interestingly, the spirocyclic precursor species (**5**, *m/z* 547) detected in earlier experiments was absent in this analysis (Figures 6B and 6D). Instead, the ClbI product intermediate (**4**, *m/z* 609), which is spontaneously transformed to **5** by cyclodehydration and decarboxylation, was detected in reaction mixtures of the wild-type, M125A, S177A, T283A, T316A, and T318A mutants, suggesting these residues are not critical for the reaction (Figures 6D and 6E). Meanwhile, S178A, H314A, H353A, D355A, S417A, and M420A mutants failed to produce precursor **4**, implying their roles in translocation, elongation, and/or cyclopropanation. Among these, S178 is the active serine that tethers the upstream intermediate, while H314 and H353 are known to participate in translocation and elongation.[44] M420 is a gating residue within the gate loop crucial for enzyme activity.[45] Meanwhile, the precise roles of D355 and S417 remain unclear, suggesting their potential involvement in cyclopropanation. Notably, both residues exhibit similar conservation scores in a multiple sequence analysis (MSA) of bacterial PKSs (Figure S12D). Next, we generated precursor models using automated covalent docking using Autodock4.[46] To facilitate computation by reducing the precursor's rotameric states, we substituted the myristoyl (C14) chain with a nonanoyl (C9) chain and then analyzed the top ten docking models to determine whether any positioned the α-proton and the cyclopropane-forming carbon near D355 or S417, as suggested by our LC-MS analysis. Among these models, we identified one model in which the α-proton is positioned near the carboxylate side chain of D355 (Figure 6E). In one of the rotameric positions of D355, the distance between the α-proton and the carboxylic oxygen decreases to 1.1 Å, supporting the potential role of D355 in α-proton abstraction. However, further experimental validation would be required to establish the specific residues involved in cyclopropanation.

DISCUSSION

AT-less PKSs, also known as *trans*-AT PKSs, are broadly found in natural product biosynthetic gene clusters and utilize more diverse modules than *cis*-AT PKSs.[47] A recent study suggested that nearly one-quarter of the bacterial modular PKSs are attributed to the *trans*-AT category.[48] Sequence analyses of AT domains distinctly classify *cis*-AT and *trans*-AT groups, supporting the existence of distinct evolutionary lineages between the two classes.[48] Meanwhile, the classification of PKSs that entirely lack AT domains versus those possessing partial or degraded AT domains remains ambiguous within databases such as SBSPKS and Uniprot.[49,50] Thus far, only three partial-AT PKS structures, OzmQ, ClbK, and ClbC, have been determined, which limits our understanding of the molecular mechanisms of partial-AT PKSs (Table S2).[20,51]

Among the three structures, only ClbC LD has a variation (having a β strand instead of α helix) while OzmQ and ClbK maintain intact LD folds (Figures S13A and S13B). Meanwhile, their partial-AT domains shared two common features: First, all these partial-AT PKSs lack the ferredoxin-like small subdomain, resulting in the absence of the active site pocket formed between the small and large subdomains (Figures S13C–S13F). Second, the active site motifs, such as the acylation site and substrate recognition site, are also altered in these partial-AT PKSs (Figure S14A). Third, the structures were varied near the boundaries of the missing small subdomain location and relatively conserved in the N-terminal and C-terminal regions of the partial AT domain, probably due to their connection to the neighboring domains, such as LD and ACP. Meanwhile, the structural differences of the partial AT domains would diversify the overall architecture of the partial-AT-containing PKSs, which may regulate the communication between the partial AT-containing PKS and the *trans*-AT partners.

To examine whether the characteristics we observed are consistent across other partial-AT domains, we queried the SBSPKS v2 database for *trans*-AT sequences with a long linker between the KS and ACP domains and predicted their structures using AlphaFold3.[49,52] The analysis identified three additional partial-AT-containing modules: the first module of ZmaA (zwittermicin biosynthetic gene cluster), DebB (FK288 biosynthetic gene cluster), and ClbO (colibactin biosynthetic gene cluster).

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

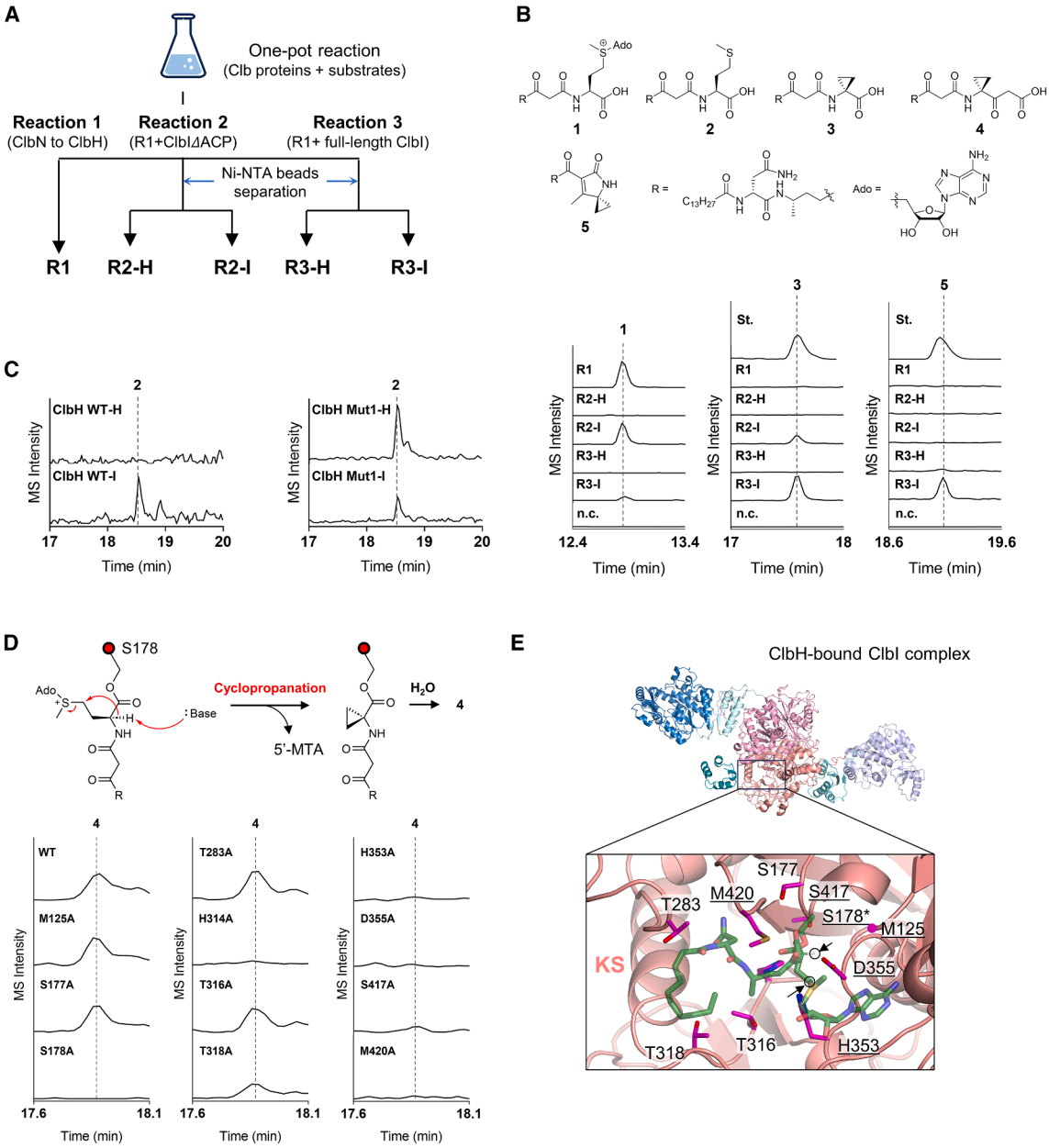


Figure 6. LC-MS analysis of colibactin intermediates

(A) Scheme for *in vitro* colibactin precursor biosynthesis. ClbN, ClbB, ClbC, and ClbH were incubated in with their substrates alone (**Reaction 1**), with ClbIΔACP (**Reaction 2**), or full-length ClbI (**Reaction 3**). After the reaction, ClbI was separated from other His-tagged Clb proteins, and the fractions (**R1**, **R2-H**, **R2-I**, **R3-H**, and **R3-I)** were analyzed by LC-MS.

(B) Chemical structures of five colibactin precursors (**1**–**5**) and Extract Ion Chromatograms [EICs, *m/z* 864 (**1**), 567 (**3**), and 547 (**5**)] from five fractions (**R1**, **R2-H**, **R2-I**, **R3-H**, and **R3-I**).

(C) The EICs of **2** (*m/z* 615) from the *in vitro* colibactin biosynthesis reaction mixtures using WT and Mut1 ClbH. The reaction mixtures were separated into ClbI and ClbH fractions, resulting in four fractions (**ClbH WT-H**, **ClbH WT-I**, **ClbH Mut1-H**, and **ClbH Mut1-I**).[43]

(D) (Top) Proposed mechanism of cyclopropanation.[43] (Bottom) The EIC peaks of colibactin precursor **4** (*m/z* 609) detected from the reaction mixture containing ClbI WT or single mutants.

(E) ClbH precursor modeled in the ClbI substrate channel (shown at 30% transparency to indicate it is from docking experiment, not from cryo-EM map). The α-proton and the carbon involved in the cyclopropanation (Figure 6D, top) are circled with arrows. Residues tested in Figure 6D are shown in stick format, except for M125, shown by a magenta circle at its Cα. The active serine (S178*) was omitted as it is covalently linked to the precursor. Mutations that impaired precursor production are underlined.

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

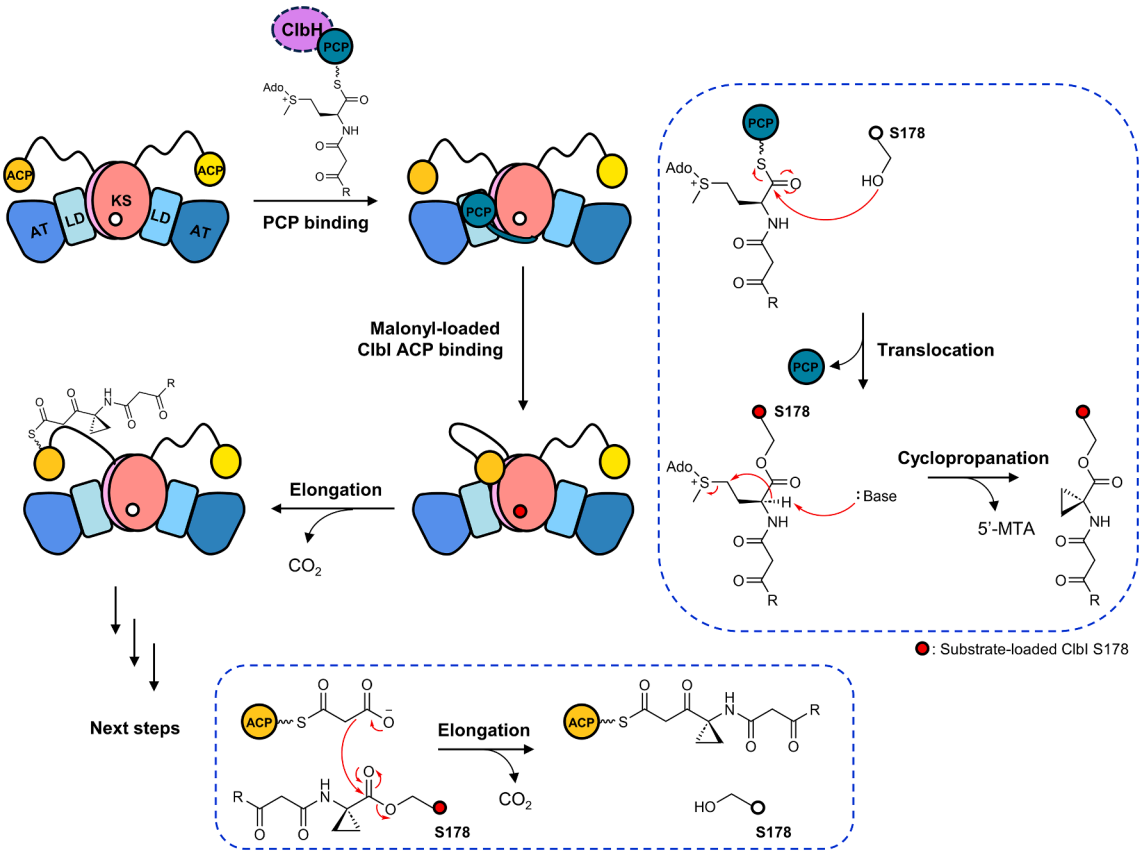


Figure 7. Proposed molecular mechanism of ClbI

First, ClbI recognizes ClbH PCP through a docking interaction between the C-terminal helix of ClbH and the dimeric groove of ClbI, allowing ClbH PCP to bind to the ClbI KS domain. Then, the colibactin precursor tethered to ClbH PCP is translocated to the active site serine (S178) of the ClbI KS and is modified to form a cyclopropane group within the substrate channel. Following this, the malonyl-loaded ClbI ACP binds to the ClbI KS domain, transferring the cyclopropane-containing precursor and incorporating a two-carbon unit. (See also Figure S12C).

Interestingly, the predicted structures of these modules shared the characteristics observed in the three existing partial-AT structures (Figure S14B). Notably, potential partial-AT sequences were only found from hybrid NRPS-PKS, not *trans*-AT PKS clusters, supporting the idea that partial AT domains may arise through evolutionary degradation, particularly in systems combining type I PKS and NRPS modules.[51]

In the NRPS, PKS, and their hybrid biosynthetic pathways, a correct precursor transfer between upstream and downstream enzymes is crucial for precise assembly. For the task, the 6-deoxyerythronolide B synthase (DEBS) system utilizes 'docking domain (DD)' elements.[53,54] DDs are located at the C-terminus of the upstream PKS (^{C}DD) and the N-terminus of the downstream PKS (^{N}DD), and together form a four-helix bundle, suggesting a potential to engineer the protein-protein interaction between PKS enzymes.[54] In the NRPS system, COM domains direct enzyme-enzyme communication.[42,55] In the SrfA-C enzyme that produces surfactin, an α-helix of the COM^{D} (a donor COM) is bound to the hand motif of COM^{A} (an acceptor COM), which comprises one α-helix and three β strands.[55] The two β strands of the COM^{A} are located in the catalytic domain of the downstream enzyme, making the COM^{A} dependent on the catalytic activity. In the tomaymycin biosynthetic pathway, the N-terminal domain of TomB works as an adaptor with its four-helix bundle structure and recruits the upstream PCP domain of TomA.[56] In some hybrid PKS-NRPS systems (e.g., the rhabdopeptide and enacyloxin assembly pathways), the C-terminal short linear motif (SLiM) of the upstream module interacts with the downstream N-terminal β-hairpin motif (βhD) at the PCP/C domain interface.[57,58]

From the ClbH-bound ClbI complex structure, we found a novel docking interaction where the C-terminal peptide of the upstream ClbH PCP binds to the groove of the downstream ClbI KS dimer, achieving 1:2 stoichiometry. This contrasts with the 2:2 and 1:1 stoichiometries previously seen in DD and COM domain interactions, respectively. C-terminal sequences of other Clb proteins are weakly conserved 20–30 aa peptides following their CP domains (Figure S15A). AlphaFold predictions showed similar helical structures in the C-terminal peptides of ClbB ACP, ClbK PCP, and ClbJ PCP2 (Figure S15B). Furthermore, we identified protein sequences similar to ClbH PCP using BLAST and found that most of them maintain C-terminal peptides forming a helix following the CP domains (Figure S15C), though the sequence conservation of this helix is lower than that of the CP regions, suggesting that enzyme-enzyme docking interactions are universal yet highly specific to each enzyme pair.

ClbI catalyzes two pivotal reactions: a decarboxylative Claisen condensation for adding a two-carbon unit into the colibactin precursor and a cyclopropanation reaction via sulfonium-based nucleophilic substitution. Based on our structural and biochemical analyses alongside prior literature, we propose a mechanistic model for these reactions as follows (Figure 7)[4,44]: Initially, ClbH recognizes ClbI through a specific interaction between the C-terminal helix of the ClbH and the dimeric surface of the ClbI, and the ClbH PCP subsequently binds to the lateral side of the ClbI substrate channel opening. The SAM-containing precursor on ClbH PCP is then transferred to the active serine (S178) of the ClbI KS domain (Translocation). Within the substrate channel, cyclopropanation occurs, facilitated by an unidentified base that abstracts a proton from the α-carbon of the precursor, generating 5′-MTA (5′-methylthioadenosine) as a byproduct (Cyclopropanation). Finally, the malonyl-loaded ClbI ACP binds to the substrate channel, undergoes decarboxylation and subsequently engages with the cyclopropane-containing precursor, extending it by a two-carbon unit.[44]

In summary, we determined the cryo-EM structures of the two colibactin biosynthetic PKS enzymes, ClbC and ClbI, in the apo- and substrate-bound states. From the structures, we visualized an incomplete AT domain of ClbC, identified CP binding sites, and discovered a novel NRPS-PKS docking interaction between ClbH and ClbI. Furthermore, we suggested a working model of ClbI for the cyclopropane generation and elongation. Our discoveries will deepen our understanding of colibactin biosynthesis and provide a structural foundation for microbiome-associated therapeutics and further chemical biosynthesis engineering.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jin Young Kang (jykang59@kaist.ac.kr).

Materials availability

All unique/stable reagents generated in this study are available from the lead contact without restriction.

Data and code availability

- Cryo-EM data have been deposited in the Electron Microscopy DataBank (EMDB) under accession numbers EMD-38222 (apo-ClbC), EMD-38405 (ACP-bound ClbC), EMD-38406 (apo-ClbI ClbI), EMD-38410 (ACP-bound ClbI), and EMD-38410 (ClbH-bound ClbI). Coordinates have been deposited in the RCSB Protein DataBank under accession codes 8XBL (apo-ClbC), 8XJT (ACP-bound ClbC), 8XJU (apo-ClbI), 8XJY (ACP-bound ClbI) and 8XJZ (ClbH-bound ClbI). They are publicly available as of the date of publication. All other data are available in the main text or the supplemental information.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This work was funded by POSCO TJ Science Fellowship (J.Y.K), National Research Foundation of Korea (2020R1F1A107746211 to J.Y.K., 2021R1C1C100656014 to J.Y.K., 2021R1C1C1011045 to C.S.K., and 2022R1A6A1A03054419 to C.S.K), BK21 FOUR Project (C.S.K), and US National Institute of Health grant (P41 GM109824 and GM103314 to B.T.C.). We thank Dr. Jin-Seok Choi and Dr. Seong-Gyu Lee at the KAIST Analysis Center for Research Advancement (KARA) for help with cryo-EM grid screening, Dr. Jiwon Kim and Sujeong Kim at the Institute of Membrane Protein (IMP) for help with cryo-EM data collection, and Dr. Ed Eng, Elina Kopylov, Charlie Dubbeldam, and Carolina Hernandez at the New York Structural Biology Center (NYSBC) for help with cryo-EM data collection. We also thank Global Science experimental Data hub Center (GSDC) at the Korea Institute of Science and Technology Information (KISTI) and the IBS Research Solution Center for computing resources and technical support. We thank Geunyoung Cho at the IMP for helping with the mass photometry experiment.

AUTHOR CONTRIBUTIONS

Conceptualization, M.K., J.K., and J.Y.K.; methodology, M.K., J.K., and J.Y.K.; investigation, M.K., J.K., G.S.L., P.D.B.O., Y.A., J.L.C., J.P., Y.J., and J.P.; visualization, M.K., J.K., G.S.L., P.D.B.O., C.S.K., and J.Y.K.; writing—original draft, M.K., J.K., G.S.L., P.D.B.O., C.S.K., and J.Y.K.; writing—review and editing, M.K., J.K., P.D.B.O., C.S.K., and J.Y.K.; funding acquisition, B.T.C., C.S.K., and J.Y.K.; supervision, S.B.H., C.S.K., and J.Y.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Expression and purification of proteins
  - The phosphopantetheinylation (ppantylation) of the Clb proteins
  - Synthesis of the BODIPY-CoA
  - BODIPY-CoA loading assay of the Clb proteins
  - Preparation of crosslinked Clb complexes for cryo-EM
  - Cryo-EM sample preparation
  - Cryo-EM data acquisition and processing
  - Model building, refinement, and validation
  - Native mass spectrometry (nMS) analysis
  - LC-MS sample preparation and analysis
  - Mass photometry experiment
  - Multiple sequence alignment of Clb proteins and ClbH-PCP-like proteins
  - Molecular docking
  - Crosslinker modeling
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.str.2025.04.017.

Received: January 13, 2025
Revised: March 14, 2025
Accepted: April 22, 2025
Published: May 16, 2025

REFERENCES

1. Donia, M.S., and Fischbach, M.A. (2015). Small molecules from the human microbiota. Science *349*, 1254766. https://doi.org/10.1126/science.1254766.
2. Nougayrede, J.-P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006). Escherichia coli Induces DNA Double-Strand Breaks in

Eukaryotic Cells. Science *313*, 848–851. https://doi.org/10.1126/science.1127059.

3. Vizcaino, M.I., and Crawford, J.M. (2015). The colibactin warhead cross-links DNA. Nat. Chem. *7*, 411–417. https://doi.org/10.1038/nchem.2221.
4. Zha, L., Jiang, Y., Henke, M.T., Wilson, M.R., Wang, J.X., Kelleher, N.L., and Balskus, E.P. (2017). Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring. Nat. Chem. Biol. *13*, 1063–1065. https://doi.org/10.1038/nchembio.2448.
5. Xue, M., Kim, C.S., Healy, A.R., Wernke, K.M., Wang, Z., Frischling, M.C., Shine, E.E., Wang, W., Herzon, S.B., and Crawford, J.M. (2019). Structure elucidation of colibactin and its DNA cross-links. Science *365*, eaax2685. https://doi.org/10.1126/science.aax2685.
6. Wilson, M.R., Jiang, Y., Villalta, P.W., Stornetta, A., Boudreau, P.D., Carrá, A., Brennan, C.A., Chun, E., Ngo, L., Samson, L.D., et al. (2019). The human gut bacterial genotoxin colibactin alkylates DNA. Science *363*, eaar7785. https://doi.org/10.1126/science.aar7785.
7. Pleguezuelos-Manzano, C., Puschhof, J., Rosendahl Huber, A., van Hoeck, A., Wood, H.M., Nomburg, J., Gurjao, C., Manders, F., Dalmasso, G., Stege, P.B., et al. (2020). Mutational signature in colorectal cancer caused by genotoxic pks+ E. coli. Nature *580*, 269–273. https://doi.org/10.1038/s41586-020-2080-8.
8. Dziubańska-Kusibab, P.J., Berger, H., Battistini, F., Bouwman, B.A.M., Iftekhar, A., Katainen, R., Cajuso, T., Crosetto, N., Orozco, M., Aaltonen, L.A., and Meyer, T.F. (2020). Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. Nat. Med. *26*, 1063–1069. https://doi.org/10.1038/s41591-020-0908-2.
9. Buc, E., Dubois, D., Sauvanet, P., Raisch, J., Delmas, J., Darfeuille-Michaud, A., Pezet, D., and Bonnet, R. (2013). High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer. PLoS One *8*, e56964. https://doi.org/10.1371/journal.pone.0056964.
10. McCarthy, A.J., Martin, P., Cloup, E., Stabler, R.A., Oswald, E., and Taylor, P.W. (2015). The genotoxin colibactin is a determinant of virulence in Escherichia coli K1 experimental neonatal systemic infection. Infect. Immun. *83*, 3704–3711. https://doi.org/10.1128/IAI.00716-15.
11. Dubinsky, V., Dotan, I., and Gophna, U. (2020). Carriage of Colibactin-producing Bacteria and Colorectal Cancer Risk. Trends Microbiol. *28*, 874–876. https://doi.org/10.1016/j.tim.2020.05.015.
12. Moretti, C., Hosni, T., Vandemeulebroecke, K., Brady, C., De Vos, P., Buonaurio, R., and Cleenwerck, I. (2011). Erwinia oleae sp. nov., isolated from olive knots caused by Pseudomonas savastanoi pv. savastanoi. Int. J. Syst. Evol. Microbiol. *61*, 2745–2752. https://doi.org/10.1099/ijs.0.026336-0.
13. Bondarev, V., Richter, M., Romano, S., Piel, J., Schwedt, A., and Schulz-Vogt, H.N. (2013). The genus *Pseudovibrio* contains metabolically versatile bacteria adapted for symbiosis. Environ. Microbiol. *15*, 2095–2113. https://doi.org/10.1111/1462-2920.12123.
14. Engel, P., Vizcaino, M.I., and Crawford, J.M. (2015). Gut Symbionts from Distinct Hosts Exhibit Genotoxic Activity via Divergent Colibactin Biosynthesis Pathways. Appl. Environ. Microbiol. *81*, 1502–1512. https://doi.org/10.1128/AEM.03283-14.
15. Silpe, J.E., Wong, J.W.H., Owen, S.V., Baym, M., and Balskus, E.P. (2022). The bacterial toxin colibactin triggers prophage induction. Nature *603*, 315–320. https://doi.org/10.1038/s41586-022-04444-3.
16. Mousa, J.J., Yang, Y., Tomkovich, S., Shima, A., Newsome, R.C., Tripathi, P., Oswald, E., Bruner, S.D., and Jobin, C. (2016). MATE transport of the E. coli-derived genotoxin colibactin. Nat. Microbiol. *1*, 15009. https://doi.org/10.1038/nmicrobiol.2015.9.
17. Guntaka, N.S., Healy, A.R., Crawford, J.M., Herzon, S.B., and Bruner, S.D. (2017). Structure and Functional Analysis of ClbQ, an Unusual Intermediate-Releasing Thioesterase from the Colibactin Biosynthetic Pathway. ACS Chem. Biol. *12*, 2598–2608. https://doi.org/10.1021/acschembio.7b00479.
18. Tripathi, P., Shine, E.E., Healy, A.R., Kim, C.S., Herzon, S.B., Bruner, S.D., and Crawford, J.M. (2017). ClbS Is a Cyclopropane Hydrolase That Confers Colibactin Resistance. J. Am. Chem. Soc. *139*, 17719–17722. https://doi.org/10.1021/jacs.7b09971.
19. Cougnoux, A., Gibold, L., Robin, F., Dubois, D., Pradel, N., Darfeuille-Michaud, A., Dalmasso, G., Delmas, J., and Bonnet, R. (2012). Analysis of structure-function relationships in the colibactin-maturating enzyme ClbP. J. Mol. Biol. *424*, 203–214. https://doi.org/10.1016/j.jmb.2012.09.017.
20. Bonhomme, S., Contreras-Martel, C., Dessen, A., and Macheboeuf, P. (2023). Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin. Structure *31*, 700–712.e4. https://doi.org/10.1016/j.str.2023.03.012.
21. Velilla, J.A., Volpe, M.R., Kenney, G.E., Walsh, R.M., Balskus, E.P., and Gaudet, R. (2023). Structural basis of colibactin activation by the ClbP peptidase. Nat. Chem. Biol. *19*, 151–158. https://doi.org/10.1038/s41589-022-01142-z.
22. Robbins, T., Liu, Y.-C., Cane, D.E., and Khosla, C. (2016). Structure and mechanism of assembly line polyketide synthases. Curr. Opin. Struct. Biol. *41*, 10–18. https://doi.org/10.1016/j.sbi.2016.05.009.
23. Chen, Y., Kelly, E.E., Masluk, R.P., Nelson, C.L., Cantu, D.C., and Reilly, P.J. (2011). Structural classification and properties of ketoacyl synthases. Protein Sci. *20*, 1659–1667. https://doi.org/10.1002/pro.712.
24. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2.
25. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. Nat. Methods *19*, 679–682. https://doi.org/10.1038/s41592-022-01488-1.
26. Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., et al. (2018). SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. *46*, W296–W303. https://doi.org/10.1093/nar/gky427.
27. Zha, L., Wilson, M.R., Brotherton, C.A., and Balskus, E.P. (2016). Characterization of Polyketide Synthase Machinery from the pks Island Facilitates Isolation of a Candidate Precolibactin. ACS Chem. Biol. *11*, 1287–1295. https://doi.org/10.1021/acschembio.6b00014.
28. La Clair, J.J., Foley, T.L., Schegg, T.R., Regan, C.M., and Burkart, M.D. (2004). Manipulation of Carrier Proteins in Antibiotic Biosynthesis. Chem. Biol. *11*, 195–201. https://doi.org/10.1016/j.chembiol.2004.02.010.
29. Tang, Y., Kim, C.-Y., Mathews, I.I., Cane, D.E., and Khosla, C. (2006). The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. Proc. Natl. Acad. Sci. USA *103*, 11124–11129. https://doi.org/10.1073/pnas.0601924103.
30. Krissinel, E., and Henrick, K. (2007). Inference of Macromolecular Assemblies from Crystalline State. J. Mol. Biol. *372*, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.
31. Yariv, B., Yariv, E., Kessel, A., Masrati, G., Chorin, A.B., Martz, E., Mayrose, I., Pupko, T., and Ben-Tal, N. (2023). Using evolutionary data to make sense of macromolecules with a "face-lifted" ConSurf. Protein Sci. *32*, e4582. https://doi.org/10.1002/pro.4582.
32. Pan, H., Tsai, S.c., Meadows, E.S., Miercke, L.J.W., Keatinge-Clay, A.T., O'Connell, J., Khosla, C., and Stroud, R.M. (2002). Crystal Structure of the Priming β-Ketosynthase from the R1128 Polyketide Biosynthetic Pathway. Structure *10*, 1559–1568. https://doi.org/10.1016/S0969-2126(02)00889-4.
33. Bagde, S.R., Mathews, I.I., Fromme, J.C., and Kim, C.-Y. (2021). Modular polyketide synthase contains two reaction chambers that operate asynchronously. Science *374*, 723–729. https://doi.org/10.1126/science.abi8532.

34. Liew, C.W., Nilsson, M., Chen, M.W., Sun, H., Cornvik, T., Liang, Z.-X., and Lescar, J. (2012). Crystal Structure of the Acyltransferase Domain of the Iterative Polyketide Synthase in Enediyne Biosynthesis. J. Biol. Chem. *287*, 23203–23215. https://doi.org/10.1074/jbc.M112.362210.

35. Worthington, A.S., Rivera, H., Torpey, J.W., Alexander, M.D., and Burkart, M.D. (2006). Mechanism-Based Protein Cross-Linking Probes To Investigate Carrier Protein-Mediated Biosynthesis. ACS Chem. Biol. *1*, 687–691. https://doi.org/10.1021/cb6003965.

36. Ye, B., Tian, W., Wang, B., and Liang, J. (2024). CASTpFold: Computed Atlas of Surface Topography of the universe of protein Folds. Nucleic Acids Res. *52*, W194–W199. https://doi.org/10.1093/nar/gkae415.

37. Sulpizio, A., Crawford, C.E.W., Koweek, R.S., and Charkoudian, L.K. (2021). Probing the structure and function of acyl carrier proteins to unlock the strategic redesign of type II polyketide biosynthetic pathways. J. Biol. Chem. *296*, 100328. https://doi.org/10.1016/J.JBC.2021.100328.

38. Evans, S.E., Williams, C., Arthur, C.J., Płoskoń, E., Wattana-amorn, P., Cox, R.J., Crosby, J., Willis, C.L., Simpson, T.J., and Crump, M.P. (2009). Probing the Interactions of Early Polyketide Intermediates with the Actinorhodin ACP from S. coelicolor A3(2). J. Mol. Biol. *389*, 511–528. https://doi.org/10.1016/j.jmb.2009.03.072.

39. Farmer, R., Thomas, C.M., and Winn, P.J. (2019). Structure, function and dynamics in acyl carrier proteins. PLoS One *14*, e0219435. https://doi.org/10.1371/journal.pone.0219435.

40. Mofid, M.R., Finking, R., and Marahiel, M.A. (2002). Recognition of Hybrid Peptidyl Carrier Proteins/Acyl Carrier Proteins in Nonribosomal Peptide Synthetase Modules by the 4′-Phophopantetheinyl Transferases AcpS and Sfp. J. Biol. Chem. *277*, 17023–17031. https://doi.org/10.1074/jbc.M200120200.

41. Kim, S.K., Dickinson, M.S., Finer-Moore, J., Guan, Z., Kaake, R.M., Echeverria, I., Chen, J., Pulido, E.H., Sali, A., Krogan, N.J., et al. (2023). Structure and dynamics of the essential endogenous mycobacterial polyketide synthase Pks13. Nat. Struct. Mol. Biol. *30*, 296–308. https://doi.org/10.1038/s41594-022-00918-0.

42. Fage, C.D., Kosol, S., Jenner, M., Öster, C., Gallo, A., Kaniusaite, M., Steinbach, R., Staniforth, M., Stavros, V.G., Marahiel, M.A., et al. (2021). Communication Breakdown: Dissecting the COM Interfaces between the Subunits of Nonribosomal Peptide Synthetases. ACS Catal. *11*, 10802–10813. https://doi.org/10.1021/acscatal.1c02113.

43. Ma, S., Mandalapu, D., Wang, S., and Zhang, Q. (2022). Biosynthesis of cyclopropane in natural products. Nat. Prod. Rep. *39*, 926–945. https://doi.org/10.1039/d1np00065a.

44. Robbins, T., Kapilivsky, J., Cane, D.E., and Khosla, C. (2016). Roles of Conserved Active Site Residues in the Ketosynthase Domain of an Assembly Line Polyketide Synthase. Biochemistry *55*, 4476–4484. https://doi.org/10.1021/acs.biochem.6b00639.

45. Mindrebo, J.T., Chen, A., Kim, W.E., Re, R.N., Davis, T.D., Noel, J.P., and Burkart, M.D. (2021). Structure and Mechanistic Analyses of the Gating Mechanism of Elongating Ketosynthases. ACS Catal. *11*, 6787–6799. https://doi.org/10.1021/acscatal.1c00745.

46. Bianco, G., Forli, S., Goodsell, D.S., and Olson, A.J. (2016). Covalent docking using autodock: Two-point attractor and flexible side chain methods. Protein Sci. *25*, 295–301. https://doi.org/10.1002/pro.2733.

47. Helfrich, E.J.N., and Piel, J. (2016). Biosynthesis of polyketides by trans-AT polyketide synthases. Nat. Prod. Rep. *33*, 231–316. https://doi.org/10.1039/C5NP00125K.

48. O'Brien, R.V., Davis, R.W., Khosla, C., and Hillenmeyer, M.E. (2014). Computational identification and analysis of orphan assembly-line polyketide synthases. J. Antibiot. *67*, 89–97. https://doi.org/10.1038/ja.2013.125.

49. Khater, S., Gupta, M., Agrawal, P., Sain, N., Prava, J., Gupta, P., Grover, M., Kumar, N., and Mohanty, D. (2017). SBSPKSv2: Structure-based sequence analysis of polyketide synthases and non-ribosomal peptide synthetases. Nucleic Acids Res. *45*, W72–W79. https://doi.org/10.1093/nar/gkx344.

50. Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E.H., Britto, R., Bye-A-Jee, H., Cukura, A., et al. (2023). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res. *51*, D523–D531. https://doi.org/10.1093/nar/gkac1052.

51. Lohman, J.R., Ma, M., Osipiuk, J., Nocek, B., Kim, Y., Chang, C., Cuff, M., Mack, J., Bigelow, L., Li, H., et al. (2015). Structural and evolutionary relationships of "AT-less" type I polyketide synthase ketosynthases. Proc. Natl. Acad. Sci. USA *112*, 12693–12698. https://doi.org/10.1073/pnas.1515460112.

52. Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Ronneberger, O., Willmore, L., Ballard, A.J., Bambrick, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature *630*, 493–500. https://doi.org/10.1038/s41586-024-07487-w.

53. Smith, H.G., Beech, M.J., Lewandowski, J.R., Challis, G.L., and Jenner, M. (2021). Docking domain-mediated subunit interactions in natural product megasynth(et)ases. J. Ind. Microbiol. Biotechnol. *48*, kuab018. https://doi.org/10.1093/jimb/kuab018.

54. Broadhurst, R.W., Nietlispach, D., Wheatcroft, M.P., Leadlay, P.F., and Weissman, K.J. (2003). The Structure of Docking Domains in Modular Polyketide Synthases ple, between the C terminus of DEBS 1 and the N termi-nus of DEBS 2) (Figure 1A). Recognition that the N termini of PKS multienzymes. Chem. Biol. *10*, 723–731. https://doi.org/10.1016/s1074-5521(03)00156-x.

55. Tanovic, A., Samel, S.A., Essen, L.-O., and Marahiel, M.A. (2008). Crystal Structure of the Termination Module of a Nonribosomal Peptide Synthetase. Science *321*, 659–663. https://doi.org/10.1126/science.1159850.

56. Karanth, M.N., Kirkpatrick, J.P., Krausze, J., Schmelz, S., Scrima, A., and Carlomagno, T. (2024). The specificity of intermodular recognition in a prototypical nonribosomal peptide synthetase depends on an adaptor domain. Sci. Adv. *10*, eadm9404. https://doi.org/10.1126/sciadv.adm9404.

57. Kosol, S., Gallo, A., Griffiths, D., Valentic, T.R., Masschelein, J., Jenner, M., de los Santos, E.L.C., Manzi, L., Sydor, P.K., Rea, D., et al. (2019). Structural basis for chain release from the enacyloxin polyketide synthase. Nat. Chem. *11*, 913–923. https://doi.org/10.1038/s41557-019-0335-5.

58. Hacker, C., Cai, X., Kegler, C., Zhao, L., Weickhmann, A.K., Wurm, J.P., Bode, H.B., and Wöhnert, J. (2018). Structure-based redesign of docking domain interactions modulates the product spectrum of a rhabdopeptide-synthesizing NRPS. Nat. Commun. *9*, 4366. https://doi.org/10.1038/s41467-018-06712-1.

59. Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530. https://doi.org/10.1016/J.JSB.2012.09.006.

60. Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296. https://doi.org/10.1038/nmeth.4169.

61. Cardone, G., Heymann, J.B., and Steven, A.C. (2013). One number does not fit all: Mapping local variations in resolution in cryo-EM reconstructions. J. Struct. Biol. *184*, 226–236. https://doi.org/10.1016/J.JSB.2013.08.002.

62. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612. https://doi.org/10.1002/JCC.20084.

63. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.H., and Ferrin, T.E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. Protein Sci. *30*, 70–82. https://doi.org/10.1002/pro.3943.

64. Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213–221. https://doi.org/10.1107/S0907444909052925.

65. Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. *60*, 2126–2132. https://doi.org/10.1107/S0907444904019158.

66. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics *25*, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033.

67. Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al. (2018). MolProbity: More and better reference data for improved all-atom structure validation. Protein Sci. *27*, 293–315. https://doi.org/10.1002/pro.3330.

68. Olinares, P.D.B., and Chait, B.T. (2020). Native Mass Spectrometry Analysis of Affinity-Captured Endogenous Yeast RNA Exosome Complexes. Methods Mol. Biol. *2062*, 357–382. https://doi.org/10.1007/978-1-4939-9822-7_17.

69. Reid, D.J., Diesing, J.M., Miller, M.A., Perry, S.M., Wales, J.A., Montfort, W. R., and Marty, M.T. (2019). MetaUniDec: High-Throughput Deconvolution of Native Mass Spectra. J. Am. Soc. Mass Spectrom. *30*, 118–127. https://doi.org/10.1007/S13361-018-1951-9/SUPPL_FILE/JS8B05914_SI_001.PDF.

70. Marty, M.T., Baldwin, A.J., Marklund, E.G., Hochberg, G.K.A., Benesch, J.L. P., and Robinson, C.V. (2015). Bayesian deconvolution of mass and ion mobility spectra: From binary interactions to polydisperse ensembles. Anal. Chem. *87*, 4370–4376. https://doi.org/10.1021/acs.analchem.5b00140.

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
CC118 competent *E. coli* cells	In-house	N/A
BL21(DE3) competent *E. coli* cells	Enzynomics	Cat# CP111
Chemicals, peptides, and recombinant proteins		
Protease Inhibitor Cocktail	Sigma Aldrich	Cat# 11836153001
BODIPY FL Maleimide	Thermo Fisher Scientific	Cat# B10250
crosslink probe containing a *trans*-3-chloroacrylamide moiety	WuXi AppTec	N/A
Coenzyme A sodium salt hydrate	Sigma Aldrich	Cat# C3144
Malonyl coenzyme A lithium salt	Sigma Aldrich	Cat# M4263
Myristoyl coenzyme A lithium salt	Sigma Aldrich	Cat# M4414
S-(5'-Adenosyl)-L-methionine chloride dihydrochloride	Sigma Aldrich	Cat# A7007
Adenosine 5-triphosphate disodium salt	Sigma Aldrich	Cat# A6419
L-Alanine	Sigma Aldrich	Cat# 05129
L-Asparagine	Sigma Aldrich	Cat# A0884
Critical commercial assays		
Phusion high-fidelity DNA polymerase	Thermo Fisher Scientific	Cat# F530L
Deposited data		
apo-ClbC structure	This study	EMD-38222 (PDB 8XBL)
ACP-bound ClbC structure	This study	EMD-38405 (PDB 8XJT)
apo-ClbI structure	This study	EMD-38406 (PDB 8XJU)
ACP-bound ClbI structure	This study	EMD-38410 (PDB 8XJY)
ClbH-bound ClbI structure	This study	EMD-38410 (PDB 8XJZ)
Recombinant DNA		
pET28a-ClbN	Addgene	Cat# 51497
pET28a-ClbB	Addgene	Cat# 114155
pET28a-ClbC	Addgene	Cat# 114145
pET28a-ClbH	Addgene	Cat# 114156
pET29b-ClbI	Addgene	Cat# 114151
pET29b-Sfp	Addgene	Cat# 75015
pET28a-coaA	Addgene	Cat# 50386
pET28a-coaD	Addgene	Cat# 50388
pET28a-coaE	Addgene	Cat# 50390
Software and algorithms		
RELION 3.1	Scheres et al.[59]	https://relion.readthedocs.io/
cryoSPARC v3	Punjani et al.[60]	https://cryosparc.com
Bsoft package (version 2.0.5)	Cardone et al.[61]	https://cbiit.github.io/Bsoft/
SWISS-MODEL	Waterhouse et al.[26]	https://swissmodel.expasy.org/
AlphaFold2	Jumper et al.[52]	https://colab.research.google.com/github/sokrypton/ColabFold/blob/v1.2.0/AlphaFold2.ipynb
UCSF Chimera (version 1.11.12)	Pettersen et al.[62]	https://www.cgl.ucsf.edu/chimera/
UCSF Chimera X (version 1.4)	Pettersen et al.[63]	https://www.cgl.ucsf.edu/chimerax/
PHENIX (version 1.18.2)	Adams et al.[64]	https://www.phenix-online.org/

(*Continued on next page*)

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Coot	Emsley et al.[65]	https://www2.mrc-lmb.cam.ac.uk/personal/pemsley/coot/
PyMOL (version 2.4.0)	Pettersen et al.[63]	https://pymol.org/
Clustal Omega	Waterhouse et al.[66]	https://www.ebi.ac.uk/jdispatcher/
Jalview	Waterhouse et al.[66]	https://www.jalview.org/
Other		
Ni-NTA agarose	Bioplus	Cat# 30210
HiTrap IMAC HP	Cytiva	Cat# 17092005
HiTrapQ HP column	Cytiva	Cat# 17115401
HiLoad 16/600 Superdex 200 pg column	Cytiva	Cat# 28989335
Superdex 200 Increase 10/300 GL column	Cytiva	Cat# 28990944
7K Zeba spin desalting column	Thermo Fisher Scientific	Cat# 89890

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

All proteins were produced in *E. coli* strain BL21(DE3) (Enzynomics). Cells were grown in LB broth media with appropriate antibiotics and temperatures. The method details provide further details.

METHOD DETAILS

Expression and purification of proteins

pET plasmids for ClbN, ClbB, ClbC (#114145), ClbH (#114156), and ClbI (#114151) were obtained from Addgene, and the proteins were purified as in the reference.[27] For the expression of the Clb proteins except ClbI and ClbH, the culture was induced with 0.1 mM IPTG at the OD_{600} ~0.5-0.6 and grown at 15°C for 16-18 hours. The cells were harvested, resuspended in the lysis buffer (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM $MgCl_2$, 10% v/v glycerol, protease inhibitor cocktail), and lysed by French Press at 4°C. The lysate was centrifuged, and the supernatant was loaded onto a HiTrap IMAC HP (Cytiva) pre-equilibrated with 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10 mM imidazole, 10% v/v glycerol, and 0.4 mM DTT. Protein was eluted with an imidazole gradient from 10 mM to 250 mM. The ClbC fractions were diluted by the low salt buffer containing 100 mM NaCl and loaded onto a HiTrapQ HP column (Cytiva), then eluted with 0.1-1 M NaCl gradient. The eluted protein was concentrated using a 50K MWCO Amicon Ultra Centrifugal Filter (Merck Millipore) and injected to HiLoad 16/600 Superdex 200 pg column (Cytiva) equilibrated in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, 10% v/v glycerol, and 1 mM DTT. For the ClbI and ClbH expression and purification, the same procedures were applied except using a higher NaCl concentration (500 mM) in the lysate buffer.

For the Sfp expression, the plasmid obtained from Addgene (#75015) was transformed in BL21(DE3), and a single colony from the transformed cells was grown overnight at 37°C for the large culture. The culture was induced with 1 mM IPTG and grown at room temperature for 6 hours. The cells were harvested, resuspended in the lysis buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 5 mM imidazole, 2 U/mL DNase I), and lysed by French Press (Avestin) at 4°C. The lysate was centrifuged to remove cell debris, and loaded onto a Hitrap IMAC HP column (Cytiva) pre-equilibrated with 50 mM Tris-HCl pH 8.0, 300 mM NaCl for Ni^{2+} affinity chromatography. The protein was eluted with 20 mM Tris-HCl pH 8.0, 500 mM NaCl, and an imidazole gradient from 5 mM to 250 mM, and dialyzed twice against 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 5 mM DTT. After dialysis, Sfp was concentrated by 10K MWCO Amicon Ultra centrifugal filter (Merck Millipore) and loaded onto HiLoad 16/600 Superdex 200 pg column (Cytiva) equilibrated in 10 mM Tris-HCl pH 8.0, 120 mM NaCl, and 5 mM DTT for size exclusion chromatography.

For the CoaA, CoaD, and CoaE expression (plasmids obtained from Addgene, #50386, #50388, #50390, respectively), the culture was induced with 0.1 mM IPTG (CoaA, CoaE) or 1 mM IPTG (CoaD) at OD_{600} ~ 0.6 and grown 16-18 hours at 37°C. The cells were harvested, resuspended in the lysis buffer (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM $MgCl_2$, 10% v/v glycerol, protease inhibitor cocktail), and lysed by French Press (Avestin). The lysate was centrifuged, and the supernatant was loaded onto a HiTrap IMAC HP column (Cytiva) and eluted with an imidazole gradient from 10 mM to 250 mM over 20 CV. The eluted protein was concentrated using a 10K MWCO Amicon Ultra centrifugal filter (Merck Millipore) and injected onto HiLoad 16/600 Superdex 200 pg column (Cytiva) equilibrated in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 10% v/v glycerol. The purified proteins above were supplemented by 15% glycerol, aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C until use.

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

The phosphopantetheinylation (ppantylation) of the Clb proteins

The purified Clb proteins were buffer-exchanged into a reaction buffer containing 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP. 20.4 μM of the buffer-exchanged apo-Clb proteins were reacted with a 20-fold excess of coenzyme A in the presence of 6.7 μM Sfp at room temperature for 1 hour. The complete conversion of apo-Clb proteins to holo form was confirmed by native mass spectrometry.

Synthesis of the BODIPY-CoA

BODIPY-CoA was synthesized as described previously.[28] Briefly, 300 μL of coenzyme A sodium salt hydrate (1 mg/mL, Sigma Aldrich) was added in 1.9 mL of reaction buffer (75 mM MES pH 6.0, 100 mM $Mg(OAc)_2$) containing 300 μL of DMSO. 220 μL of BODIPY-FL-Maleimide (25 mg/mL solution in DMSO, ThermoFisher) was added to the reaction buffer, and the resulting solution was stirred in the dark room at 4°C for 3 hours. The synthesis of the BODIPY-CoA was monitored by Thin Layer Chromatography (butanol: acetic acid: water, 5:2:4). After the reaction was completed, 10 mL of ethyl acetate was added to the reaction mixture three times for extraction to remove unreacted BODIPY-FL-maleimide.

BODIPY-CoA loading assay of the Clb proteins

BODIPY-CoA loading assay was initiated by mixing 5 μM Clb protein, 1 μM Sfp, 5 μM BODIPY-CoA in the reaction buffer containing 50 mM Tris-HCl pH 8.0, 200 mM NaCl, and 10 mM $MgCl_2$. The mixture was incubated in the dark at room temperature for 1 hour, and the reaction was quenched by adding 2x SDS-PAGE loading buffer containing 4.4% SDS, 111 mM Tris pH 6.8, 22.2% v/v glycerol, 0.22% bromophenol blue and boiling at 95°C for 5–10 minutes. The proteins were separated by 5-6% Tris-Acetate SDS-PAGE gel, and its fluorescence signal at λ = 488 nm and Coomassie blue staining were imaged.

Preparation of crosslinked Clb complexes for cryo-EM

The crosslink probe containing a *trans*-3-chloroacrylamide moiety was synthesized and analyzed by 1D NMR, HPLC, and LCMS in 99.8% purity and 100% ee (WuXi AppTec).[35] For the crosslinking, 100 μM of the crosslink probe was reacted with 66.7 μM of the buffer-exchanged Clb ACP protein in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP in the presence of 5 mM ATP, 2.5 μM Sfp, 0.5 μM CoaA, 0.5 μM CoaD and 0.5 μM CoaE. The reaction was carried out at room temperature for 3 hours, and the crosslinker attachment was examined by competitive BODIPY-CoA loading assay.

To prepare the ACP-bound ClbC complex, the full-length ClbC was attached by the crosslink probe in a one-pot reaction containing 10 μM ClbC, 3 μM CoaA, 3 μM CoaD, 3 μM CoaE, 1 μM Sfp, 200 μM crosslink probe, 8 mM ATP in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP. The reaction was incubated for 30 minutes at room temperature and then stopped by removing the substrates using a 7K Zeba desalting column (ThermoFisher). The crosslinked ClbC was purified in Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP. The pulled fractions containing crosslink ClbC were concentrated using a 50 K MWCO Amicon Ultra Centrifugal Filter (Merck Millipore) up to 1 mg/mL for cryo-EM grid preparation.

To prepare the ClbH-PCP-bound and ACP-bound ClbI complexes, the crosslinker-loaded ClbI ACP was added to the ClbI (S178C)ΔACP at a molar ratio of 3:1 and incubated at 10°C overnight (for ACP-bound ClbI complex) or at 4°C for 1 hour. Then, the crosslinked protein mixture was applied onto a 7K Zeba spin desalting column (ThermoFisher) to remove unreacted crosslinkers and substrates and purified in Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP. The pulled protein fractions were concentrated using a 50 K MWCO Amicon Ultra Centrifugal Filter (Merck Millipore) up to 1.5 mg/mL for cryo-EM grid preparation.

Cryo-EM sample preparation

For the apo-Clb proteins, the frozen proteins were thawed and purified by Superdex 200 10/300 GL column equilibrated in 20 mM Tris HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT. C Flat 1.2/1.3 Au 400, Quantifoil 1.2/1.3 Au300, Quantifoil 1.2/1.3 Cu400, or UltrAuFoil R1.2/1.3 300 grids were glow-discharged at negative polarity, 0.26 mbar, 15 mA for 25 seconds. Then, 3 μL of a sample was applied onto the grid and immediately blotted for 4 seconds and plunge-frozen in liquid ethane with 100% chamber humidity at 22°C using Vitrobot Mark IV (FEI).

Cryo-EM data acquisition and processing

Detailed data collection conditions are described in Table 1. Briefly, the frozen grids were screened on a 200 kV Glacios microscope, and the best grids were sent to either the New York Structural Biology Center (New York, USA) or the Institute of Membrane Proteins (Pohang, Republic of Korea) for data collection. Briefly, ~ 7200 - 15300 movies were acquired using a 300kV Titan Krios microscope (ThermoFisher) equipped with a K3 direct electron detector (Gatan). The pixel size ranged from 0.8 to 1.1 Å/pixel, and the defocus ranged from -0.8 μm to -2.5 μm were used for the data collection in either a super-resolution or a counting mode. A total of 50 frames were acquired per movie, with a cumulative electron dose of ~50-60 $e^-/Å^2$. The movies were aligned, summed, and dose-weighted using MotionCor2 in RELION 3.1 or Patch Motion Correction in cryoSPARC v3.[48–50] The dose-weighted images generated in RELION 3.1 were then transferred to cryoSPARC v3 for further processing. In the cryoSPARC, the contrast transfer function (CTF) of the summed images was determined using Patch CTF. Subsequently, particles were picked using a blob picker, extracted, and subjected to 2D classification. The resulting 2D averages were used for template picking, and the picked particles were sorted by 2D

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

classification and applied to heterogeneous refinement. The best class(es) were processed by either additional heterogeneous refinement or homogeneous refinement in cryoSPARC, and the final refined particles were transferred back to RELION 3.1 for particle polishing. Details for the cryo-EM data analysis for each dataset are described in Figures S3G, S5G, S7G, S9G, and S11.

Model building, refinement, and validation

The local resolution estimation and filtration were done by Blocres and Blocfilt commands in the Bsoft package (version 2.0.5), respectively.[61] For the apo-ClbC and apo-ClbI modeling, initial coordinates were generated by SWISS-MODEL and AlphaFold2[24,26] and fitted into the cryo-EM density using UCSF Chimera (version 1.11.12).[62] Then, the KS, LD, and AT domains were rigid-body refined in PHENIX (version 1.18.2),[64] and the overall structure was manually modified in Coot,[65] and subjected to the real-space refinement in PHENIX. Manual editing and real-space refinement were iterated until the desired results were achieved. For the refinement run, minimization_global, local_grid_search, and adp options were applied, and the secondary structure restraints were used. The nonbonded_weight parameter value was set to 1000 (default value: 100) to improve the MolProbity and clash scores.[67] For the refinement, the post-processed maps were used, but the local filtered map was also used for the last refinement iteration of the ACP-bound ClbI structure because it improved the modeling of the ACP region when inspected by eyes. The figures were made using either UCSF Chimera X (version 1.4) or PyMOL (version 2.4.0).[63]

Native mass spectrometry (nMS) analysis

The apo- and holo- ClbC and ClbI samples were buffer-exchanged into nMS solution (200 mM ammonium acetate, pH 7.5, 0.01% Tween-20) using Zeba spin desalting columns with a 40-kDa molecular weight cut-off (Thermo Scientific). Samples were diluted to a final concentration of 4-5 μM prior to nMS analysis. Each sample was loaded into a gold-coated quartz capillary tip that was prepared in-house and was electrosprayed into an Exactive Plus EMR instrument (Thermo Fisher Scientific) using a modified static nanospray source.[68] The MS parameters used included: spray voltage, 1.21 kV; capillary temperature, 150 – 200°C; S-lens RF level, 200; resolving power, 17,500 at m/z of 200; AGC target, 1 x 106; number of microscans, 5; maximum injection time, 200 ms; in-source dissociation (ISD), 10 – 30 V; injection flatapole, 8 V; interflatapole, 7 V; bent flatapole, 5 – 6 V; high energy collision dissociation (HCD), 200 V; ultrahigh vacuum pressure, 6.3 – 6.5 $\times$ 10^{-10} mbar; total number of scans, 100. Mass calibration in positive EMR mode was performed using cesium iodide clusters. Raw nMS spectra were visualized using Thermo Xcalibur Qual Browser (version 4.2.47). Data processing and spectra deconvolution were performed using UniDec version 4.2.0.[69,70] The following parameters were used for data processing: background subtraction (if applied), subtract curve 10; smooth charge state distribution, enabled; peak shape function, Gaussian.

The expected masses based on the protein sequence included ClbC (lost its N-terminal methionine): 99,068.9 Da and ClbI: 110,683.2 Da. Phosphopantetheinylation adds 322.3 Da to each protein subunit. The accuracy of the measured masses from the nMS analysis (calculated as the % deviation from the expected mass) ranged from 0.003% to 0.06%.

LC-MS sample preparation and analysis

For *in vitro* biosynthesis of the colibactin precursors, the frozen proteins were thawed and buffer-exchanged in 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP. The reaction mixture contained 4 μM ClbN, 4 μM ClbB, 4 μM ClbC, 10 μM ClbH, 125 μM CoA, 3 μM Sfp, 0.5 mM myristoyl-CoA, 0.5 mM malonyl-CoA, 0.3 mM L-Asn, 0.3 mM L-Ala, 0.3 mM SAM, 2.3 mM ATP, 1 mM NADPH, 1.7% v/v DMSO, and 10 μM ClbI if necessary. ClbN, ClbB, ClbC, and ClbH were incubated with CoA and Sfp for 1 hour at room temperature. After an initial incubation, the remaining substrate was added, and the reaction proceeded overnight. To isolate ClbI from the reaction mixture, the C-terminal his-tag was cleaved using PPX protease before the reaction. After the reaction, ClbI was separated from the reaction mixture using Ni^{2+}-NTA beads.

The separated proteins were added by 5-fold volume of ice-cold methanol, incubated on ice for 30 minutes, and centrifuged at 15,000 rpm for 15 minutes. The supernatant was saved for LC-MS analysis, and the protein pellet was hydrolyzed by adding 40 μL of 0.1 M KOH by heating at 70°C for 20 minutes. The samples were cooled on ice, 80 μL of 0.1 M HCl was added to the solutions for neutralization, and 120 μL of methanol was added. The hydrolyzed samples were then incubated at -20°C overnight and centrifuged at 15,000 rpm for 15 minutes.

The LC-MS analysis was performed on an Agilent1260 series HPLC system with a diode array detector and a 6130 series ESI mass spectrometer equipped with an analytical Kinetex C18 100 Å column (250 mm $\times$ 4.6 mm i.d., 5 μm; flow rate: 0.7 mL/min). Before injection, the pellet samples were filtered using 1.5 mL microcentrifuge filter tubes (0.45 μm, SPL Life Sciences, Korea), and subsequently dissolved in 240 μL of methanol for LC-MS analysis. The supernatant samples were initially concentrated using a centrifugal vacuum concentrator (HyperVAC-MAX, Gyrogen, Korea), and were subsequently dissolved in 240 μL of methanol. The samples were filtered using the same method as the pellet and then transferred to LC-MS vials for analysis. The synthesis of colibactin precursors **2** and **3**, which were used as standards for LC-MS analysis, is described in the Data S1 and S2.

Mass photometry experiment

The mass photometry experiment was performed on the Refeyn TwoMP system. For the measurement, coverslips were cleaned with water, ethanol, and isopropanol, and immersion oil was applied on the microscope objective lens before a coverslip was positioned on the microscope stage. 18 μL of buffer containing 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP was placed on the sample compartment to optimize the focus of the microscope. After locking the focus of the microscope, 2 μL of

Please cite this article in press as: Kim et al., Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin, Structure (2025), https://doi.org/10.1016/j.str.2025.04.017

100 nM Clb protein solution was added to the buffer on the silicon gasket and mixed with pipetting to achieve a final concentration of 10 nM. Data acquisition was started right after the addition of protein.

Multiple sequence alignment of Clb proteins and ClbH-PCP-like proteins

Amino acid sequences of the PKS/ NRPS proteins in the pks island were obtained from UniProt, and domain boundaries were determined using the InterPro server. The sequences were aligned by Clustal Omega and visualized using Jalview.[66] To investigate whether the C-terminal helix after the CP domain is widespread, pBLAST was performed on a non-redundant database in NCBI using ClbH-PCP. To exclude the same proteins in the result, the organisms that possibly contain the pks cluster, Enterobacteriales (taxid:91347), Klebsiella (taxid:570), *Escherichia coli* (taxid:562), and *Frischella perrara* (taxid:126701), were excluded on a database. Resulting proteins with the ClbH-PCP coverage at the C-terminus were selected and aligned by Clustal Omega. For clarity, other domains that do not align with the ClbH-PCP were deleted and aligned again with Clustal Omega. For secondary structure annotation, the JPred web service in Jalview was used.[66]

Molecular docking

Covalent docking simulations were performed using AutoDock4, MGLTools, and covalent docking preparation scripts.[46] The three-dimensional (3D) coordinates of the precursor molecules were initially generated using Coot. The apo-ClbI and precursor structures were subsequently processed with MGLTools and covalent docking preparation scripts to ensure compatibility with AutoDock4. A grid box of 20 Å $\times$ 20 Å $\times$ 20 Å was centered at the ligand binding site. Docking simulations were carried out using default parameters, with the precursor docked to chain A of apo-ClbI.

Crosslinker modeling

The 3D coordinates of the crosslinker molecule were built using Coot. The parameter file for covalently linked crosslinker was generated using Phenix. The manually programmed restraint file was input for the refinement step to establish the covalent linkages between ClbI-ACP S952, ClbH-PCP S1524, and ClbI S178.

QUANTIFICATION AND STATISTICAL ANALYSIS

Cryo-EM data collection and refinement statistics are summarized in Table 1. Root-mean-square deviations (r.m.s.d.) of Cα atoms between two structures (Figures 1B, 2D, and 4E) were calculated using PyMOL. σ values in Figure 5B were obtained using the data analysis software provided with the Refeyn TwoMP system.