Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

A Noncanonical Reducing-Domain Architecture Defines a New Class of Modular Polyketide Synthases

Jin Young Kang

jykang59@kaist.ac.kr

KAIST https://orcid.org/0000-0002-8493-7890

Jinwoo Kim

KAIST https://orcid.org/0000-0002-7205-9170

Byeongsub Lee

KAIST https://orcid.org/0009-0004-6837-4719

Jongseok Park

KAIST

Hyeongwoo Kim

KAIST

Minjae Kim

KAIST https://orcid.org/0000-0003-2098-1945

Gyu Sung Lee

Sungkyunkwan University

Chung Sub Kim

Sungkyunkwan University https://orcid.org/0000-0001-9961-4093

Hyun Uk Kim

Korea Advanced Institute of Science and Technology (KAIST) https://orcid.org/0000-0001-7224-642X

Article

Keywords:

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8739699/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: There is **NO** Competing Interest.

Abstract

Colibactin is a microbiome-derived genotoxin that damages the genomes of intestinal epithelial cells and has been linked to colorectal cancer. Its biosynthesis is encoded by a large gene cluster (*pks*/*clb*) that functions as a hybrid nonribosomal peptide synthetase - polyketide synthase assembly line, yet the structural basis of its assembly remains poorly understood. Here we show cryo-electron microscopy structures of colibactin biosynthetic enzyme ClbB that uncover a noncanonical architecture for a fully-reducing polyketide synthase module. Unlike metazoan fatty acid synthases and iterative polyketide synthases, ClbB arranges its condensing and reducing regions in a crossed configuration. Within its reducing region, the enoylreductase forms an inverted dimer and the dehydratase is positioned peripherally as monomers. Moreover, the dehydratase-enoylreductase unit is embedded within the ketoreductase fold. Bioinformatic analyses combined with deep-learning-based structure prediction indicate that this domain-embedding logic is broadly distributed across diverse bacterial phyla. These findings define a previously unrecognized class of reducing polyketide synthase modules and provide an expanded structural framework for understanding how modular polyketide synthases coordinate multi-step chemistry. The work also extends design principles for engineering megasynthases for chemical biosynthesis and synthetic biology.

Main Text

Diverse small molecules are produced by trillions of gut commensal microbiota, mediating host-microbe interactions[1]. While many microbiota-derived metabolites benefit the host, some have harmful effects and contribute to pathogenic conditions[2]. For instance, certain pathogenic *Proteobacteria* strains, including *Escherichia coli*, harbor the *pks* island, which encodes the biosynthetic enzymes of a genotoxin named colibactin[3–5]. Colibactin contains two thiazole rings that facilitate its intercalation into the duplex DNA and two cyclopropane rings that induce inter-strand DNA crosslinking in the gut epithelial cells, causing DNA double-strand breaks (DSBs) and cell cycle arrest[6–9]. As such, colibactin has been implicated in colorectal cancer, inflammatory bowel disease, and neonatal systemic infection[10–12].

The colibactin biosynthetic gene cluster comprises 19 genes, including three nonribosomal peptide synthetases (NRPSs), three polyketide synthases (PKSs), and two hybrid NRPS-PKS enzymes (Supplementary Fig. 1a)[6,13,14]. As modular assembly-line megaenzymes, NRPS and PKS synthesize secondary metabolites through a stepwise process, with each module incorporating and modifying a single building block. A minimal PKS module contains ketosynthase (KS), acyltransferase (AT), and acyl carrier protein (ACP) domains[15]. The AT domain selects an acyl extender unit and transfers it onto the phosphopantetheinyl arm of the holo-ACP via a thioester bond, while the KS domain catalyzes decarboxylative condensation between the ACP-tethered substrate and the upstream intermediate. Additional reducing domains - ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) - determine the degree of reduction of the β-carbon, and an optional methyltransferase (MT) introduces a methyl group to α-carbon or β-hydroxyl/keto oxygen[16,17] (Supplementary Note 1). Similarly, a minimal

NRPS module consists of condensation (C), adenylation (A), and peptidyl carrier protein (PCP) domains. The A domain activates and loads a specific amino acid onto the holo-PCP in an ATP-dependent manner, and the C domain catalyzes peptide bond formation with the upstream PCP-tethered intermediate. Upon the completion of the assembly, the product is released from the carrier protein (CP) domain - either PCP or ACP - via hydrolysis or catalysis by a thioesterase (TE)-mediated cleavage[18]. The hybrid NRPS-PKS system exploits the homology between PCP and ACP, allowing cross-recognition with both C and KS domains, thereby facilitating condensation reactions in both assembly lines[19].

The proposed mechanism of colibactin biosynthesis displays several intriguing features. For example, the pathway incorporates both *cis*-AT and *trans*-AT enzymes, utilizes an atypical aminomalonyl extender unit, and constructs a reactive cyclopropane moiety through a unique S-adenosyl methionine-dependent mechanism distinct from those found in other cyclopropane-containing natural products[20–23]. Notably, ClbB, the second assembly-line enzyme in the colibactin biosynthetic pathway, is a hybrid NRPS-PKS enzyme composed of a minimal NRPS module and a fully-reducing PKS module (Fig. 1a). Canonical fully-reducing PKSs typically feature a DH-ΨKR-ER-KR domain organization, in which the ER domain is positioned between the inactive ΨKR (or structural KR, KR_S) and catalytic KR (KR_C) (Fig. 1b).

In contrast, ClbB adopts a distinct domain configuration, ΨKR-KR-DH-ER, deviating from the conserved domain architecture (Fig. 1a, b)[24]. To date, only a few fully-reducing PKS module structures have been reported[25–30]. Most adopt a metazoan fatty acid synthase (mFAS)-like architecture, in which the DH domains form a central dimer, the ER dimer lies above it, and the ΨKR/KR domains extend outward without significant contact with the condensing domains (KS and AT) (Fig. 1c). To our knowledge, the only reported exception is module 5 of the juvenimicin biosynthetic pathway (JuvEIII), whose crystal structure revealed a canonical DH dimer yet unexpected DH-KR and ER-KR interfaces mediated by surface-associated interdomain linkers[30]. Meanwhile, all structurally characterized fully-reducing PKS modules reported to date retain the canonical DH-ΨKR-ER-KR sequence order, and no structure has been reported for a fully-reducing PKS module with the ΨKR-KR-DH-ER domain arrangement. Therefore, structural analysis of the ClbB PKS module reveals a fundamentally distinct structural class of fully-reducing PKS module.

Here, we determined the cryo-EM structures of the ClbB PKS module (hereafter referred to as $ClbB_{PKS}$) in its apo and NADPH-bound states at 3.6 Å and 2.5 Å, respectively. These structures reveal that the noncanonical domain sequence of $ClbB_{PKS}$ gives rise to a distinctive three-dimensional organization, redefining its overall architecture, domain oligomerization, and reaction chamber. This unprecedented configuration is further shaped by newly emerged linker regions that establish additional domain-domain interfaces. Bioinformatic analyses show that ClbB-type PKSs are widespread across diverse bacterial phyla. Together with AlphaFold-predicted models of ClbB-type PKSs, our structure uncovers novel insertion sites for DH, ER, and MT domains within the KR domain, defining a new class of reducing PKSs and expanding current understanding of PKS architectural diversity.

Results and Discussion

$ClbB_{PKS}$ defines a new structural class of fully-reducing PKSs

We first delineated the domain boundaries of ClbB through multiple sequence alignment (MSA) and subcloned its PKS ($ClbB_{PKS}$) and NRPS ($ClbB_{NRPS}$) regions for recombinant expression (Supplementary Fig. 2). The biosynthetic activities of the purified proteins were verified by liquid chromatography-mass spectrometry (LC-MS) analysis of their in vitro biosynthetic products (Supplementary Fig. 1b, c)[21,31]. In addition, the phosphopantetheinylation activity of $ClbB_{PKS}$ was assessed using a BODIPY-CoA loading assay (Supplementary Fig. 1d).

Our initial cryo-EM reconstruction of apo-$ClbB_{PKS}$ yielded a 3.6 Å-resolution map (Supplementary Figs. 3, 4)[32]. Although this revealed the overall architecture, the DH and ER regions were poorly resolved, precluding detailed modeling. To capture a substrate-bound state, we generated holo-$ClbB_{PKS}$ using *B. subtilis* Sfp, added the substrates malonyl-CoA and NADPH to it, and performed cryo-EM analysis. This approach markedly improved the density, yielding a 2.5 Å-resolution reconstruction with clear NADPH density but no traceable malonyl-CoA (Fig. 1d and Supplementary Figs. 5, 6). We therefore designate this reconstruction as the 'NADPH-bound $ClbB_{PKS}$' structure, which was used for all subsequent analyses unless otherwise indicated.

The structure reveals a canonical homodimeric KS-AT condensing region, consistent with previously reported modular PKS (mPKS) structures (Fig. 1d)[33]. In contrast, the reducing region of $ClbB_{PKS}$ adopts a spatial organization that is fundamentally distinct from the mFAS-like architecture shared by canonical fully-reducing PKSs (Fig. 1e and Supplementary Fig. 7)[25–30], thereby defining a previously unrecognized structural class. Four features distinguish $ClbB_{PKS}$ from canonical fully-reducing PKSs: First, the overall module geometry is unique. The condensing and reducing regions of $ClbB_{PKS}$ intersect at an angle of 74.8° in top view, whereas these two regions are nearly coaxially aligned in mFAS-like PKSs (Fig. 1c, e). Second, the ΨKR/KR domains are positioned above the KS and linker domain (LD), establishing direct contacts with the condensing region that are absent in the mFAS-like architecture. These interactions stabilize the KR domain, which exhibits more robust density than DH and ER domains in the apo structure (Fig. 1d and Supplementary Figs. 3e and 5d). Third, the DH domain is located at the periphery of the reducing region as a monomer, bridging the central ER dimer and the catalytic KR domains, in contrast to the central DH dimer observed in mFAS-like architectures (Fig. 1c-e and Supplementary Fig. 7). Fourth, although the ER dimer interface in $ClbB_{PKS}$ is structurally conserved, its orientation is inverted relative to that of mFAS ER domains, repositioning the entrance to its active site from the condensing-region side (mFAS) to the solvent-exposed side ($ClbB_{PKS}$) (Supplementary Fig. 8). Taken together, these four features identify $ClbB_{PKS}$ as the first structurally characterized member of a new class of fully-reducing PKS modules with the ΨKR-KR-DH-ER domain arrangement.

Multibody refinement of apo-ClbB$_{PKS}$ further revealed substantial mobility of the DH-ER didomain relative to the rigid KS-AT-KR core, visualizing stochastic multi-directional motions of the DH-ER in the absence of substrates (Supplementary Video 1)[34]. This flexibility contrasts with human FASN, in which the entire modifying region, including KR, DH, ER, and ΨMT domains, moves as a rigid body relative to the condensing region[35].

ClbB condensing region maintains a canonical organization

The condensing region of ClbB$_{PKS}$ (KS-LD-AT) adopts a domain organization similar to those of mPKS, iPKS, and mFAS, featuring a central KS dimer with a dimeric interface area of 4,166 Å^2, as analyzed by ePISA[25,27,33,36–38]. The KS domain exhibits the canonical thiolase-like αβαβα fold and harbors the conserved catalytic triad (C1254, H1390, H1428) within its active-site pocket (Supplementary Fig. 9a)[33,39,40].

The AT domain contains the catalytic residue S1694 and the malonyl-specific motifs $_{1792}$HAFH$_{1795}$, consistent with its substrate specificity (Supplementary Fig. 9b)[41]. Notably, the AT domains of ClbB$_{PKS}$ sit lower relative to those of other mPKS dimers, such as DEBS module 5 and ClbI, yielding a smaller interaxial angle between the two KS-LD-AT domains (Fig. 2a and Supplementary Fig. 10a, b). This deviation appears to stem from a distinct feature of the ClbB$_{PKS}$ LD. Whereas the LD typically adopts a 3α3β-fold, the ClbB$_{PKS}$ LD lacks the terminal α-helix, which is replaced by a β-strand resembling that of the ClbC LD (Fig. 2b and Supplementary Fig. 10c)[42]. This altered geometry may, in turn, influence the KR positioning through interactions between LD α1 and ΨKR β1 (Fig. 2c).

Following the AT domain, a ~ 60-aa post-AT linker (PAL) connects to the ΨKR domain. In mFAS, this region is considerably shorter (~ 30 aa) and proceeds directly to the DH domain. In contrast, the PAL of ClbB$_{PKS}$ extends toward the ΨKR domain positioned above the LD region of the same protomer. Notably, the C-terminal portion of the PAL engages the KS, KR, and post-ER linker (PEL), thereby stabilizing KR positioning (Fig. 2d). To explore how the upstream intermediate is delivered to the ClbB KS domain, we modeled a dimeric ClbB PCP-KS-LD-AT segment (residues 994–1915) using AlphaFold3. All ranked models consistently suggested the same docking interface at the lateral side of the ClbB KS dimer, mediated primarily by helix III of ClbB PCP (Supplementary Fig. 10d).

Interestingly, from the same dataset, we also obtained a monomeric ClbB$_{PKS}$ structure at 2.8 Å resolution using 41.3% of the total particles (Supplementary Fig. 6). Its overall structure closely resembles that of apo-ClbB$_{PKS}$, in which the KS-AT-KR tridomain is well resolved, whereas the DH-ER didomain remains unresolved (Supplementary Fig. 11a). We interpret this monomeric class as more likely representing a damaged or partially disassembled population generated during grid preparation, given that no detectable monomeric species was observed in the size-exclusion chromatography profile of purified ClbB$_{PKS}$ (Supplementary Fig. 11b). Such disassembly of protein complexes during cryo-EM grid

preparation is well documented and commonly attributed to exposure to the air-water interface[43,44]. Notably, NADPH density is absent in the monomeric reconstruction, further supporting the interpretation that this population may be catalytically compromised.

To further assess the native oligomeric state of ClbB, we performed mass photometry on $ClbB_{NRPS}$, $ClbB_{PKS}$, and full-length ClbB (Supplementary Fig. 12). While $ClbB_{NRPS}$ exists exclusively as a monomer, both monomeric and dimeric species were observed for $ClbB_{PKS}$ and full-length ClbB. Measurements over the 6.25–100 nM concentration range further indicated that $ClbB_{PKS}$ would be predominantly dimeric under cryo-EM conditions (~ 3 μM), supporting the interpretation that the monomeric cryo-EM population likely arises during sample preparation (Supplementary Fig. 12f). Notably, full-length ClbB exhibits a higher dissociation constant than $ClbB_{PKS}$ (Supplementary Fig. 12e). Given that type I PKS modules are typically organized around a dimeric KS core, whereas NRPS modules more often function as monomers, these observations raise the intriguing possibility that PKS and NRPS activities of ClbB differ in their dependence on oligomeric state. Further studies will be required to determine how NRPS and PKS activities in ClbB are influenced by this dynamic monomer–dimer equilibrium[23].

ClbB KR exhibits divergent motifs and a putative ΨKR-associated ACP-docking interface

In the NADPH-bound $ClbB_{PKS}$ structure, four NADPH molecules are bound within the conserved active-site pockets of the KR and ER domains (Fig. 3a, b). Superimposition of the apo and NADPH-bound $ClbB_{PKS}$ structures reveals a slight expansion of the KS-AT-KR tridomains (Supplementary Fig. 13). The stabilization of the DH-ER didomain upon NADPH binding likely arises from a salt bridge between the 10'-phosphate of the bound NADPH (absent in NADH) and R2386 from the KR-to-DH linker (Fig. 3a). Alanine substitution of R2386 reduced the in vitro biosynthetic yield by ~ 65%, underscoring its functional importance (Fig. 4e).

Despite its distinct spatial arrangement compared with fully-reducing mPKSs, the $ClbB_{PKS}$ ΨKR/KR domain adopts a canonical short-chain dehydrogenase/reductase (SDR) fold, retaining the catalytic triad (S2321-Y2334-K2293) (Fig. 3a). Meanwhile, the $ClbB_{PKS}$ KR lacks the LDD motif characteristic of B-type KRs and the W motif indicative of A-type KRs (Supplementary Fig. 2e). These motifs are known to orient the ACP-tethered substrates and to determine the stereochemistry of the resulting β-hydroxyl group. Their absence thus suggests that $ClbB_{PKS}$ employs an alternative strategy for substrate processing, potentially linked to its atypical domain organization. An N2320-S2338 hydrogen bonding pair near the active site is retained through a reciprocal substitution of the highly conserved Ser-Asn pair in canonical mPKS KRs, indicating local adaptation of the KR active site in ClbB (Fig. 3a and Supplementary Fig. 2e).

Unlike other mPKSs, the putative entrance of the $ClbB_{PKS}$ KR active site is enclosed by the KR-to-DH linker, DH domain, and PEL, which together may form an ACP docking platform for ketoreduction

(Supplementary Fig. 14a, b)[45]. To further investigate the KR-ACP interaction and the potential roles of the KR-to-DH linker, DH domain, and PEL in ACP docking, we generated AlphaFold3 models of the $ClbB_{PKS}$ reducing region (ΨKR/KR-DH-ER) in complex with a single $ClbB_{PKS}$ ACP and aligned the resulting models to the cryo-EM structure of $ClbB_{PKS}$. Among the five ranked models, only one exhibited a plausible KR-ACP binding mode, in which ACP was oriented toward the NADPH-binding pocket while avoiding major clashes with neighboring domains (Supplementary Fig. 14c).

Guided by these observations, we generated 25 alanine substituents of $ClbB_{PKS}$ and analyzed their in vitro biosynthetic products using LC-MS (Fig. 3c, d). Among these mutants, H2127A showed the largest reduction in product formation, followed by P2128A and E3013A; additional decreases were observed for L2126A, H2279A, I2329A, and R3050A. Mapping these residues onto the structure delineates a potential ACP-docking surface located near the ΨKR-KR interface, rather than the KR-to-DH linker, DH domain, or PEL (Supplementary Fig. 14d). Because a genuine KR-ACP structure is not yet available, further structural studies designed to capture the KR-ACP complex will be required to test this model directly.

ClbB DH and ER domains also display divergent structural and sequence features

The DH domain follows the KR domain via the KR-to-DH linker and is positioned at the periphery of the reducing region as a monomer, unlike dimeric DHs found in other mPKS structures[26–30]. Despite its unusual placement, the $ClbB_{PKS}$ DH domain adopts the canonical pseudodimeric hot-dog fold and retains the catalytic dyad H2421 and D2585 (Supplementary Fig. 15) [46]. Consistent with its structural divergence, the active-site motifs of $ClbB_{PKS}$ DH show several sequence deviations relative to canonical mPKS DHs. Notably, the GYxYGPxF motif begins with His (H2540) instead of Gly, and the terminal residue of the Dxxx(Q/H) motif is replaced with Ser (S2589), reflecting its evolutionary divergence. Following the DH domain, the ER is connected via a 20-aa DH-to-ER linker, whose density is relatively weak, likely due to solvent exposure.

In both ER domains, clear NADPH densities are observed, stabilized by the conserved pyrophosphate motif, $_{2828}$GGVGLA$_{2833}$ (Fig. 3b). Although the ER retains the characteristic MDR (medium-chain dehydrogenase/reductase) fold, the positions corresponding to the conserved Lys-Asp pair in canonical ER domains are occupied by Ser (S2917) and Asn (N2941), respectively, further underscoring the distinctive characteristics of ClbB[47]. In addition, the position corresponding to a Tyr associated with formation of a (2S)-alkyl-branched product is occupied by L2729 in ClbB, whereas the conserved Arg that forms a salt bridge with the adenine ribose phosphate of NADPH is replaced by H2870, which may instead form a hydrogen bond with the phosphate moiety (Fig. 3b and Supplementary Fig. 2g)[48].

To assess the influence of these residues on catalytic activity, we generated alanine-substitution mutants of L2729, H2870, S2917, and N2941, and compared their product yields with that of wild-type

$ClbB_{PKS}$ (Supplementary Fig. 16). Because K2918 is adjacent to S2917, which occupies the position corresponding to the conserved lysine, we also included K2918 in the mutagenesis analysis to test whether it contributes to catalysis in place of S2917. LC-MS analysis showed that the K2918A mutant caused the largest decrease in product formation (~ 60%), raising the possibility that K2918 may instead fulfill the role of the conserved lysine. L2729A also showed a large reduction, suggesting an important contribution of L2729 to catalysis despite the absence of a stereochemical outcome in the native ClbB product. The S2917A mutation, which replaces the conserved Lys, also reduced product yield to ~ 65% of the wild-type level. By contrast, N2941, which occupies the conserved Asp position, had little effect on activity. The H2870A mutation reduced production by ~ 10%, indicating a modest contribution to catalysis. Overall, these mutagenesis results did not identify a single residue whose substitution completely abolishes ER activity, which is broadly consistent with previous studies showing that putative catalytic residues in PKS ER domains contribute to enzyme activity to different extents depending on the enzyme context[48–50]

In $ClbB_{PKS}$, the inverted ER dimer reorients the active site entrance toward the solvent-exposed surface (Supplementary Fig. 8). This arrangement juxtaposes the DH and ER active sites and may facilitate ACP progression from DH to ER during successive β-keto processing steps. By contrast, if the ER dimer were oriented as in mFAS, the region surrounding the ER active site would likely be more restricted within the ClbB architecture because of its proximity to the KS dimer and KR domain, potentially occluding the ACP-docking surface. Thus, the inverted ER arrangement may support catalysis by improving ACP access to the ER domain and facilitating transfer of the ACP-tethered intermediate from DH to ER. Direct testing of this model will require future biochemical studies that perturb the DH-ER spatial arrangement or the residues lining the proposed ACP transfer route.

ClbB reducing domains adopt distinct oligomerization states from mPKSs

While the KS domains of mPKS and mFAS invariably form dimers through their conserved interfaces, the oligomerization patterns of the reducing domains are more variable (Supplementary Fig. 7)[25–30,37,38]. For example, DH domains from full-length iPKS, mFAS, and mPKS are typically dimeric, whereas ER domains can exist as either monomers or dimers[25–27,30]. In contrast, the ClbB DH domain does not dimerize; instead, it establishes novel interfaces with the adjacent KR and ER domains, forming inter-domain contact areas of 326.5 $Å^2$ and 432.7 $Å^2$, respectively. These interfaces contribute to the overall assembly of the reducing region. Structural modeling based on canonical, dimeric DHs revealed steric clashes with neighboring regions, indicating that the DH dimerization is structurally disfavored in ClbB (Supplementary Fig. 17). Consistently, size-exclusion chromatography of the excised DH domain confirmed its monomeric nature even in isolation, in the absence of steric interference from other domains (Supplementary Fig. 18a,b).

In contrast, the ClbB ER domain forms a continuous β-sheet across the dimer interface, resembling the arrangement observed in mFAS and iPKSs (Fig. 1c). The isolated ER domain exists as a stable dimer in solution, both in the presence and absence of NADPH, demonstrating that ER dimerization is an intrinsic property that does not depend on other domains or cofactors (Supplementary Fig. 18a,c). Together, these results suggest that the structural diversity among reducing domains arises not only from the noncanonical domain order of $ClbB_{PKS}$ domains but also from their distinct oligomeric states.

Post-ER linker consolidates the characteristic organization of the $ClbB_{PKS}$ reducing domains

The post-ER linker (PEL) is a 105-aa segment connecting the ER to the ACP, with its N-terminal 54 residues resolved in the density map (Fig. 4a,b). The resolved density reveals two α-helices and a β-strand that weave through the interdomain spaces, establishing extensive contacts (in total, 3576 $Å^2$) with the ΨKR, KR, ER, and linker elements. The N-terminal region of PEL (PEL-N; residues 3000–3020, colored in teal) comprises a single α-helix interfacing with the ER, KR, DH, and the KR-to-DH linker. The C-terminal portion (PEL-C; residues 3021–3053, colored in dark salmon) contains an α-helix and a β-strand arranged antiparallel and bent by ~ 100° relative to the α-helix of PEL-N.

Remarkably, PEL-C occupies the position of the terminal α-helix (αG) and β-strand (βG) of the KR domain, thereby completing its fold (Fig. 4c and Supplementary Fig. 19)[51]. Consequently, the DH-ER didomain is inserted within the KR domain between the βF and αG. In canonical mPKSs, by contrast, the ER domain follows ΨKR β8, and the KR lies downstream of the ER, making this DH-ER insertion unique (Fig. 4d). The remaining ~ 40 unresolved residues of the PEL likely provide the flexibility required for ACP to shuttle the colibactin intermediate across domains and deliver it to the downstream enzyme, ClbC.

To assess the structural role of the PEL in maintaining integrity and catalytic activity, we generated ΔPEL-N and ΔPEL-C deletion mutants, along with alanine substitutions at interdomain interface residues (Figs. 2c and 4a,b). Catalytic activities were assayed by one-pot in vitro biosynthesis followed by LC–MS analysis of the colibactin precursor **1** (Fig. 4e, f). Deletion of either PEL-N or PEL-C abolished detectable activity, confirming their essential structural roles. Mutations at the PEL interfaces further revealed differential contributions of individual contacts (Fig. 4e). PEL-N interface substitutions I2792A and R3014A reduced production by approximately 30% and 55%, respectively, whereas PEL-C substitutions E3027A and R3030A caused relatively moderate reductions of ~ 20% and ~ 30%, respectively. In addition, alanine mutation of R1974, which forms a cation-π interaction with W3042 from PEL-C, as well as a salt bridge with E1542 from LD, reduced production by ~ 60%, underscoring the importance of the KR-LD and KR-PEL-C interfaces (Fig. 2c). Collectively, these results demonstrate that PEL-mediated interactions are critical for maintaining both structural integrity and catalytic activity of $ClbB_{PKS}$.

Bioinformatic analyses reveal ClbB-type PKSs across diverse bacterial phyla

Structural analysis of $ClbB_{PKS}$ revealed a distinctive architecture resulting from the DH-ER insertion into the KR domain and the emergence of new linker segments. To investigate whether ClbB-type PKSs are present across bacterial genomes, we conducted large-scale bioinformatic searches using antiSMASH-DB v.4 and InterPro (Supplementary Fig. 20)[52–55]. We first queried both databases for PKS modules containing the ΨKR-KR-DH(-ER)-ACP domain configuration and removed redundant sequences from the results. The initial hits (13 fully-reducing, 571 partially-reducing) were clustered using sequence similarity network, and ~ 10% from each cluster (one representative with fewer than 10 members) were subjected to high-throughput structure prediction using Boltz-2 to assess whether the DH(-ER) region is indeed inserted within the KR domain, as observed in ClbB (Supplementary Fig. 20)[56]. Predicted structures were manually inspected, ultimately yielding 13 fully-reducing and 161 partially reducing ClbB-type PKSs distributed across multiple bacterial phyla, including Pseudomonadota, Myxococcota, Actinomycetota, and Cyanobacteriota (Fig. 5a). Clustering of the cognate BGCs using BiG-SCAPE identified 174 BGCs grouped into 137 gene-cluster clans, further underscoring the substantial diversity of ClbB-type PKS-containing BGCs (Supplementary Fig. 21).

To quantitatively assess the sequence features that distinguish ClbB-type PKSs, we generated sequence logos for the KR, DH, and ER motif regions and compared them with those of canonical reducing PKSs (Supplementary Figs. 22, 23). These analyses recapitulated the characteristic absence of the LDD and W motifs in ClbB-type KR domains, the conserved substitutions in the DH motifs, and the altered active-site-associated residues in the ER domain, thereby providing quantitative support for the sequence divergence described above.

We next asked whether sequence-level features associated with the noncanonical architecture of ClbB-type PKSs could be identified beyond structural prediction and quantified the lengths of three interdomain linker regions (KS–AT, AT–KR, and KR–DH) across three groups from the initial hits: ClbB-type PKSs, non-ClbB-type PKSs containing an intervening ACP domain between AT and KR (+ ACP), and non-ClbB-type PKSs lacking such an ACP insertion (no ACP) (Supplementary Fig. 24). Kruskal–Wallis tests followed by Dunn's post hoc tests with Bonferroni correction showed highly significant differences among groups for all three linkers (KS-AT: $P = 1.9 \times 10^{-55}$; AT-KR: $P = 3.0 \times 10-^{34}$; KR-DH: $P = 3.9 \times 10^{-50}$). Notably, ClbB-type modules exhibited the shortest KR–DH linker, consistent with insertion of the DH(–ER) unit within the KR domain, whereas the AT–KR linker was markedly longer in the non-ClbB-type (+ ACP) group, consistent with the presence of an intervening ACP domain between AT and KR. Together, these analyses provide quantitative support for the architecture-associated sequence features that distinguish ClbB-type PKSs from canonical reducing PKSs.

We subsequently predicted the dimeric 3D structures of representative modules from the curated ClbB-type PKSs using AlphaFold3[57]. As an initial test case, we examined TubF, a tubulysin biosynthetic enzyme with the ΨKR-MT-KR-DH-ER-ACP domain configuration. The predicted model revealed the same

DH-ER didomain insertion between the KR βF strand and αG helix observed in ClbB, while its MT was positioned between the βB and the αB (Fig. 5b, c, and Supplementary Fig. 25a). We then extended this analysis to eight additional modules (four fully-reducing and four partially-reducing), all sharing the ΨKR-KR-DH(-ER)-ACP domain arrangement, and found that all retained the DH(-ER) insertion in the same site (Supplementary Figs. 25–27). Interestingly, despite this conserved local configuration, the predicted global architecture substantially varied, suggesting that structural divergence across the ClbB-type PKSs is greater than that of canonical reducing PKSs (Fig. 5b-e and Supplementary Figs. 25–27).

The conserved DH-ER insertion of ClbB-type PKSs further underscores the role of the PEL region, suggesting that it is not merely a passive linker but a conserved architectural element of this PKS class. Although sequence alignment of PEL regions across ClbB-type PKSs shows limited primary-sequence conservation, structural prediction indicates that their overall secondary-structure topology is well maintained, with a variable PEL-N region and a more constrained PEL-C region that consistently completes the KR fold (Supplementary Fig. 28). The enrichment of hydrophobic residues in PEL-C further supports its role in stabilizing the hydrophobic core of the KR domain. These features suggest that evolutionary diversification of ClbB-type PKSs may have occurred through flexible remodeling of PEL-N while preserving the structurally integrated role of PEL-C. They also point to the PEL as a potentially useful element for future engineering, including DH or DH-ER insertion into KR-only PKS modules.

Phylogenetic analyses support KR-DH co-evolution in ClbB-type PKSs

We next asked whether the noncanonical KR–DH organization of ClbB-type PKSs is associated with a recognizable evolutionary transition from the canonical DH–KR order. However, we found no evidence for a stepwise evolutionary transition from canonical DH-KR modules to noncanonical KR-DH modules, such as duplicated or partially rearranged DH-containing intermediate modules. The KR phylogeny also did not reveal a continuous transitional clade linking canonical and noncanonical modules (Fig. 6a). To test whether KR and DH nonetheless remained evolutionarily linked, we performed a cophylogenetic analysis using PACo (Procrustean Approach to Cophylogeny), which showed highly significant congruence between KR and DH phylogenies for both canonical and noncanonical modules (Fig. 6 and Supplementary Fig. 29)[58]. These results support close KR-DH co-evolution in the noncanonical arrangement, although they do not identify a specific transitional pathway from the canonical DH-KR order to the noncanonical KR-DH order. In addition, whereas the KR-DH linker sequence itself is highly variable across ClbB-type PKSs, linker length showed a clearer pattern, with ER-containing modules tending to have shorter KR–DH linkers than ER-lacking modules (Supplementary Fig. 30), suggesting that linker length is a more informative evolutionary feature of this region than primary-sequence conservation.

To provide broader phylogenetic context for the distribution of ClbB-type PKSs, we reconstructed a maximum-likelihood phylogeny of ClbB-type KR domains (Supplementary Fig. 31). The resulting tree

showed partial clustering by phylum, particularly for Actinomycetota and Pseudomonadota, consistent with lineage-associated diversification. At the same time, the phylogeny does not unambiguously distinguish vertical inheritance from horizontal transfer. Thus, the broad phylogenetic distribution and conserved domain insertion of ClbB-type PKSs suggest, but do not prove, vertical inheritance from an ancestral module followed by diversification within bacterial lineages.

A distinct reducing-domain architecture supports a discrete reaction chamber model

The unique spatial organization of the $ClbB_{PKS}$ reducing domains creates discrete β-keto processing chambers within each protomer (Fig. 7 and Supplementary Fig. 32a). Structural modeling of ACP-domain complexes, based on previous studies, suggests that the ACP can reach active sites within the same protomer across a linear distance of ~ 30–80 Å, except for the KS domain, whose catalytic cysteine residues are on the opposite protomer. In contrast, accessing catalytic domains in the opposing protomer would require a span of ~ 90–120 Å, exceeding the range allowed by the ACP-tethering linker. These geometric constraints are consistent with a model in which each ACP may operate within a single reaction chamber.

Based on our structural and biochemical analyses, together with previous studies of PKS systems, we propose a putative ACP trajectory involving the AT, KS, KR, DH, and ER domains in $ClbB_{PKS}$ (Fig. 7 and Supplementary Fig. 32b). Cavity mapping indicates that the catalytic pockets of the KS and KR domains are oriented toward a confined reaction chamber, whereas the entrances to the AT, DH and ER active sites are directed outward (Supplementary Fig. 33). This geometry suggests spatial separation between substrate-loading and terminal-processing steps while confining key intermediate catalytic steps to the central chamber. These features are consistent with a continuous-access model in which ACP may sample catalytic pockets along a smooth S-shaped trajectory from the AT to the ER domains through the central chamber of the complex (Fig. 7). Because the proposed ACP docking poses are based on homology modeling rather than experimentally resolved ACP-bound structures, the exact docking geometry remains to be established, and alternative ACP trajectories are also possible.

In addition to this overall geometric arrangement, comparison of the apo and NADPH-bound structures suggests a possible role for cofactor-dependent conformational dynamics in the catalytic cycle (Supplementary Fig. 32b). In the NADPH-bound state, a salt bridge between R2386 and the 10'-phosphate of KR-bound NADPH contributes to the stabilization of the DH-ER didomain, whereas $NADP^+$ release after ketoreduction may increase DH-ER mobility, as observed in the multibody refinement of the apo-$ClbB_{PKS}$. In particular, the stabilized state may favor ketoreduction by supporting a more ordered ACP-docking interface at the KR domain, whereas the flexible state following cofactor release may promote subsequent β-keto processing steps by reducing the likelihood of ACP re-engagement with KR and increasing the range of conformations accessible for ACP binding with the DH and ER domains. This model differs from the previously proposed gating model called 'turnstile mechanism,' in which the KS

active site is transiently occluded after chain elongation during the modifying reaction in the catalytic cycle. Instead, the flexibility of DH-ER didomain in $ClbB_{PKS}$ is more consistent with a continuous-access, dynamic-biasing model in which conformational dynamics bias ACP progression through subsequent catalytic steps while the KS and other catalytic active sites remain accessible. Furthermore, although this noncanonical architecture could, in principle, reduce diffusion loss during ACP shuttling or influence module-to-module communication, additional biochemical and kinetic studies will be required to determine whether it confers functional advantages beyond structural stabilization[48,51].

In summary, our cryo-EM structures of $ClbB_{PKS}$ reveal a fundamentally different strategy for full-chain β-keto processing in mPKSs. Rather than conforming to the canonical arrangement of reducing domains, $ClbB_{PKS}$ employs a reconfigured architecture in which the DH-ER didomain is embedded within the KR domain, while newly emerged linkers reshape the reducing region into discrete reaction chambers. Evolutionary and deep-learning-based structural analyses show that this architectural logic is widely conserved across ClbB-type PKSs, despite substantial divergence in their overall arrangements. These findings uncover an overlooked but broadly distributed design principle for reducing PKS modules, offering new mechanistic insight into the evolution of mPKS logic and conceptual foundation for engineering next-generation PKS assembly lines.

Materials and Methods

Protein expression and purification

pET plasmids containing Sfp (#75015) and ClbN (#51497) were obtained from Addgene, and the proteins were purified as in Zha, *ACS Chem. Biol.*, 2016[21]. pET plasmids containing $ClbB_{PKS}$ and $ClbB_{NRPS}$ were cloned from a plasmid obtained from Addgene ClbB (#114155). For the ClbN, $ClbB_{NRPS}$, $ClbB_{PKS}$, and ClbB expression, the culture was induced with 0.5 mM IPTG at the OD_{600} ~ 0.5–0.6 and incubated with shaking at 15 ºC for 16–18 hours. The cells were pelleted, resuspended in the lysis buffer (50 mM Tris-HCl pH 8.0, 250 mM NaCl, 10 mM $MgCl_2$, 10% v/v glycerol, homemade protease inhibitor cocktail), and lysed by French Press at 4ºC. The lysate was centrifuged twice at 15,000 rpm to remove cell debris, 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 50 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM $MgCl_2$, 10% v/v glycerol, 0.4 mM DTT, and an imidazole gradient from 10 mM to 250 mM. The fractions containing $ClbB_{PKS}$ were tracked by SDS-PAGE, and diluted by the low salt buffer (20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM $MgCl_2$, 10% v/v glycerol, and 1 mM DTT) and loaded onto a HiTrapQ HP column (Cytiva) pre-equilibrated with the low salt buffer. Protein was eluted with 20 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, 10% v/v glycerol, 1 mM DTT, and 0.1-1 M NaCl gradient. The eluted protein was concentrated using a 100K 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 size exclusion chromatography.

For the Sfp expression, 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 4ºC. 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. The purified proteins were supplemented by 15% glycerol, aliquoted, flash-frozen in liquid nitrogen, and stored at -80 ºC until use.

Synthesis of the BODIPY-CoA

BODIPY-CoA was synthesized as described previously[31]. 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 ℃ 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 $ClbB_{PKS}$, 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 95ºC 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.

LC-MS sample preparation and analysis

The frozen Clb proteins and Sfp were thawed and buffer-exchanged in 50 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP. The reaction mixture containing 5 µM Clb-proteins, 3 µM Sfp, 125 µM CoA, 500 µM Myristoyl-CoA, 500 µM Malonyl-CoA, 2 mM NADPH, 0.3 mM L-Ala, 0.3 mM L-Asn, and 2.3 mM ATP. ClbN, ClbB, $ClbB_{NRPS}$, and $ClbB_{PKS}$ were incubated with CoA and Sfp for 1 h at room temperature. After an initial phosphopantetheinylation, the remaining substrates were added, and the reaction was left to proceed overnight at room temperature.

After the reaction, 5-fold volume of ice-cold methanol was added to the mixture and incubated on ice for 30 minutes to precipitate the protein fractions. Pellet was collected via 4°C, 15,000 rpm centrifugation for 15 minutes. The supernatant was saved for subsequent LC-MS analysis, and the pellet was

hydrolyzed by adding 40 µL of 0.1 M KOH and incubation at 70 ºC for 20 minutes. For neutralization, 80 µL of 0.1 M HCl was added to solution, incubated on ice, and 120 µL of methanol was added to the samples. The hydrolyzed samples were then incubated at -20 ºC overnight for precipitation and centrifuged at 15,000 rpm for 15 minutes. The supernatant fraction was collected and transferred to the LC-MS vials for analysis. LC-MS analyses were performed on two systems. The first consisted of an UltiMate 3000 Standard LC System with ISQ EC mass spectrometer (ThermoFisher), connected with Gemini C18 110 Å (250 mm x 4.6 mm i.d., 5 µm). The conditions were as follows: 0.7 mL/min flow rate, 10 µl injection, a linear increase of solvent B from 10% to 100% over 20 min followed by solvent B 100% for 5 min (solvent A: water + 0.1% formic acid, solvent B: acetonitrile + 0.1% formic acid). The second system consisted of an ACQUITY UHPLC I-Class PLUS coupled to a Xevo G2-XS QTof mass spectrometer (Waters), equipped with an ACQUITY UHPLC BEH C18 column 130 Å (50 mm x 2.1 mm i.d., 2.1 µm). Separation conditions were as follows: 0.6 mL/min flow rate, 10 µl injection, solvent B 10% for 0.5 min, a linear increase of solvent B from 10% to 100% over 4.5 min followed by solvent B 100% for 4 min, then returning to 10% at 10 min (solvent A: water + 0.1% formic acid, solvent B: acetonitrile + 0.1% formic acid).

Cryo-EM sample and grid preparation

For apo-$ClbB_{PKS}$, the frozen stocks were thawed on ice, and buffer exchanged by Superdex 200 Increase 10/300 GL column (Cytiva) to 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM DTT. The fractions containing the protein were pooled and concentrated to 1 mg/mL using a 100K MWCO Amicon Ultra Centrifugal Filter (Merck Millipore). For substrate-bound holo-$ClbB_{PKS}$, apo-$ClbB_{PKS}$ was buffer exchanged to 20 mM Tris-HCl pH 8.0, 200 mM NaCl, 10 mM $MgCl_2$, and 1 mM TCEP, instead of DTT, by Superdex 200 Increase 10/300 GL column (Cytiva). Subsequently, 5 µM of $ClbB_{PKS}$ was phosphopantetheinylated with 2 µM Sfp and 200 µM CoA for one hour at room temperature. The reaction was quenched by 7K Zeba desalting columns (ThermoFisher), and purified by size-exclusion chromatography. The pooled fractions containing holo-$ClbB_{PKS}$ were concentrated again and reacted with 400 µM malonyl-CoA and 2 mM NADPH for one hour at room temperature and used as a cryo-EM sample without quenching.

For cryo-EM grid preparation, UltrAuFoil 1.2/1.3 300 mesh grids, or Quantifoil 1.2/1.3 300 gold mesh grids were glow-discharged at negative polarity, 0.26 mbar, 15 mA for 25 seconds, respectively. Then, 3 µL of the sample was applied onto the grid and blotted for 3 s or 4 s, and plunge-frozen in liquid ethane using Vitrobot Mark IV (FEI) in 100% humidity and 22°C setting.

Cryo-EM data acquisition and processing

The best grids were selected in 200 kV Glacios microscope, and sent to either the New York Structural Biology Center (New York, USA) or the Institute of Membrane Proteins (Pohang, Republic of Korea) for high-quality data collection. Briefly, 15,345 and 10,054 movies were acquired using a 300 kV Titan Krios microscope (ThermoFisher) equipped with a K3 direct electron detector (Gatan) for apo- and NADPH-bound $ClbB_{PKS}$, respectively. The pixel size was 1.065 and 1.1 Å/pixel, and the defocus ranged from − 0.8

μm to -2.4 μm were used for the data collection in 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[59,60]. The images generated in RELION were then transferred to CryoSPARC v3.1.0. Subsequently, the contrast transfer function (CTF) of each image was determined using Patch CTF, and particles were picked using a blob picker, extracted, and 2D classified. The best classes were used as templates for template picking, then the picked particles were sorted by 2D classification and applied to heterogeneous refinement. The best class was processed by additional heterogeneous refinement in CryoSPARC, and the sorted particles were transferred back to RELION 3.1 by PyEM for particle polishing. Details for the cryo-EM data analysis for each dataset are described in Supplementary Figs. 4 and 6.

Mass photometry experiment

The mass photometry experiment was conducted on the Refeyn TwoMP system. For measurement, the immersion oil was placed on the objective lens. The coverslip was washed by deionized water, ethanol, and isopropanol, and 18 μL of buffer containing 20 mM Tris-HCl (pH 8.0), 200 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT was applied. The focus was optimized using the buffer, and 2 μL of 62.5–1000 nM ClbB-construct protein solution was added to the 18 μL-buffer drop on the gasket. The data acquisition was commenced after pipette mixing.

Model building, refinement, and validation

The local resolution estimation and filtration were conducted using Blocres and Blocfilt functions in the Bsoft package (version 2.0.5), respectively[61]. For the modeling, the initial coordinates of fragment domains (KS-LD-AT, ΨKR-KR, DH, ER) were generated by AlphaFold2 and fitted into the cryo-EM density using UCSF Chimera (version 1.11.12)[32]. Then, the overall structure was manually modified, and linkers connecting each domain were built *de novo* in Coot[62]. Subsequently, the initial structure was subjected to the real-space refinement in PHENIX. Manual editing and real-space refinement cycles were iterated until satisfied. For the real-space refinement, minimization_global, local_grid_search, and adp options were applied, and the secondary structure restraints were used in later steps. Finally, the nonbonded_weight parameter value was set to 1000 (default value: 100) at the last run to improve the MolProbity and clash scores[63]. For the refinement, the post-processed maps were mainly used, but the local filtered map was also used because it retained better density near the periphery of the KR and DH domains. The figures were made using either UCSF Chimera X (version 1.4) or PyMOL (version 2.4.0)[64].

Identification of ClbB-type PKS modules

The amino acid sequences containing the KS-AT-KR-DH-ER or KS-AT-KR-DH domain architectures were identified using InterPro[55]. Specifically, the following domain accession IDs were used as search parameters: IPR014030 (β-ketoacyl synthase-like, N-terminal), PF02801 (β-ketoacyl synthase-like, C-terminal), IPR032821 (polyketide synthase, C-terminal extension), PF00698 (acyltransferase domain), PF08659 (ketoreductase domain), IPR049552 (polyketide synthase dehydratase domain, N-terminal),

IPR049551 (polyketide synthase dehydratase domain, C-terminal), and IPR013154 (alcohol dehydrogenase-like, N-terminal). The search was conducted with the option of “Order of domain matters: Yes”.

To identify biosynthetic gene clusters (BGCs) containing the desired module architecture, protein sequences retrieved from InterPro were mapped to their respective host species, and corresponding whole genomes were collected from the NCBI database[65]. These genomes were analyzed using antiSMASH v7.0.06 to predict and annotate BGCs[52]. BGCs containing modules with KS–AT–KR–DH or KS–AT–KR–DH–ER domain architectures were selected from the resulting GenBank files for further analysis. Additionally, BGCs retrieved from antiSMASH-DB v4 were combined with the previously identified BGCs to compile a comprehensive dataset[53].

From the initial 584 module sequences, we constructed a sequence similarity network (SSN). From the SSN, 10% of nodes from each cluster, together with all singletons, were sampled for structure prediction using Boltz-2. The resulting models were evaluated using explicit criteria to identify genuine ClbB-type PKS modules and exclude false-positive hits from the sequence-based search. Specifically, a module was classified as ClbB-type only if (i) the DH(-ER) domain was positioned between βF strand and αG helix of the KR domain, as defined by structural alignment with the cryo-EM structure of $ClbB_{PKS}$, and (ii) the overall KR fold was structurally continuous, with the PEL completing the C-terminal part of the KR domain. Modules were excluded if the DH(-ER) was positioned outside this insertion site. Among the excluded models, the majority exhibited either KS-AT-ACP-KR-DH-PPTase architecture, likely representing iterative modules, or a KS-AT-KR-DH-ACP architecture lacking the DH(-ER) insertion within KR. Application of these criteria yielded 174 ClbB-type modules.

Sequence similarity network and gene cluster analyses

To examine sequence-level relationships among these modules, a sequence similarity network (SSN) was constructed using all 584 KS–AT–KR–DH(–ER)-containing modules, where pairwise similarity was calculated using BLASTP with the following thresholds: (i) percent identity > 60%, (ii) E-value < 1×10^{-30}, and (iii) sequence coverage > 0.9 (Fig. 5a). Modules clustered primarily by taxonomic origin.

To analyze BGCs containing modules with ClbB-like domain architectures, the corresponding host species were subjected to taxonomic and phylogenetic analyses. Taxonomic information was retrieved from the NCBI Taxonomy database based on scientific names, and a phylogenetic tree was constructed using MEGA to explore evolutionary relationships among the species[65,66].

To group biosynthetic gene clusters (BGCs) into gene cluster families (GCFs) and gene cluster clans (GCCs), BiG-SCAPE was employed with default parameters[67]. The resulting GCC assignments were visualized as a network using Cytoscape v3.10.99, with each node representing a BGC containing a PKS module order of KS-AT-KR-DH(-ER), and each node is connected when the similarity between two nodes is higher than 0.3[68].

Phylogenetic analyses of canonical and noncanonical KR and DH domains

To analyze the evolutionary origin of ClbB-like domain architectures, KR and DH domains were collected from noncanonical PKS modules and canonical bacterial PKS modules. Noncanonical modules were extracted from antiSMASH-annotated GenBank files based on the presence of the KR–DH domain organization. Canonical reference modules were collected from the MIBiG 4.0 database and filtered to retain only entries whose host organisms belonged to the bacterial lineage[69]. For the canonical dataset, host taxonomy was reassigned using the NCBI taxonomy database via taxopy, and non-bacterial entries were excluded[70]. Phylum names were normalized before downstream phylogenetic analysis.

Canonical reference modules were defined as antiSMASH-annotated PKS modules from MIBiG in which KS, AT, DH, and KR domains were located on the same protein in the conventional KS–AT–DH–KR order. Additional domains, including ER or terminal tailoring domains, were allowed if the core KS–AT–DH–KR order was preserved.

To investigate the evolutionary relationships among noncanonical and canonical KR and DH domains, three protein sequence datasets were prepared: combined KR domains, paired KR domains, and paired DH domains. The combined dataset contained both noncanonical domains and canonical reference domains, and was used to assess the phylogenetic placement of noncanonical domains relative to canonical bacterial PKS domains. The paired KR and DH datasets contained matched KR–DH domain pairs from the same modules and were used for downstream comparison of KR and DH evolutionary histories. The combined KR dataset was used to construct the phylogeny shown in Fig. 6a, whereas the paired KR and DH datasets were used for the cophylogenetic (PACo) analysis in Fig. 6b.

Multiple sequence alignments were performed independently for each dataset using MAFFT with the --auto option[71]. Poorly aligned and highly gapped regions were removed using trimAl with the -gappyout option[72]. The trimmed alignments were used for maximum-likelihood phylogenetic reconstruction with IQ-TREE v3.1.1[73]. For each alignment, the best-fit amino acid substitution model was selected independently using ModelFinder Plus (-m MFP)[74]. In all cases, Q.PFAM + F+I+R5 was selected as the best-fit model. Branch support was assessed using 1,000 ultrafast bootstrap replicates and 1,000 SH-aLRT replicates[75,76]. IQ-TREE analyses were performed with automatic thread allocation and a fixed random seed (20260518) to ensure reproducibility.

Cophylogenetic analysis of KR and DH domains

To perform a Procrustes-based cophylogenetic analysis to assess whether paired KR and DH domains share congruent evolutionary histories, consensus trees inferred from the paired KR and paired DH datasets were used as input[58]. KR and DH leaves were matched by module identifier, and only modules present in both trees were retained. Patristic distance matrices were calculated independently for the KR and DH trees and transformed into Euclidean coordinates by principal coordinates analysis. The

resulting coordinate matrices were compared by Procrustes analysis, and the Procrustes m^2 statistic was used as a measure of phylogenetic incongruence, with lower m^2 values indicating greater congruence between KR and DH evolutionary histories. Statistical significance was assessed by randomly permuting KR–DH associations 999 times to generate a null distribution of m^2 values. The analysis was performed for all paired modules, noncanonical modules only, and canonical modules only.

For visualization of phylogenetic relationships between KR and DH domains, tanglegrams were generated from the paired KR and DH trees (Supplementary Fig. 29). Tanglegrams were additionally untangled by iteratively rotating internal nodes while preserving the tree bipartitions and branch lengths. This procedure was used only to improve visual readability and did not alter the quantitative cophylogeny statistics.

Analysis of KR-DH linker architecture

To characterize structural variation in the KR–DH interdomain region of noncanonical modules, the amino acid length of the linker between KR and DH domains was calculated for each module. Linker length distributions were compared according to the presence or absence of an ER domain and host phylum. Linker length distributions were compared across the three groups independently using two-sided Mann–Whitney U tests. Effect size for the ER-associated difference was estimated using rank-biserial correlation. To assess whether the apparent phylum-level difference in linker length was confounded by ER-domain presence, the association between host phylum and ER presence was tested using Fisher's exact test. Linker length differences between Actinomycetota and Pseudomonadota were further examined within ER-negative modules as a stratified comparison. For visualization, linker length distributions were plotted as boxplots stratified by ER presence, phylum, and their combination.

Comparison of interdomain linker lengths among PKS module types

To compare interdomain linker architectures across PKS module types, modules were classified into three groups on the basis of their antiSMASH domain organization: ClbB-type modules, in which a DH(–ER) unit is inserted within the KR domain; non-ClbB-type modules carrying an intervening ACP domain between the AT and KR domains (non-ClbB, +ACP); and non-ClbB-type modules lacking this ACP insertion (non-ClbB, no ACP). Group assignment was based on the linear order and connectivity of the KS, AT, DH, ER, KR, and ACP domains within each module.

For each module, the amino acid lengths of three interdomain linker regions (KS-AT, AT-KR, and KR-DH) were derived from the annotated domain boundaries. Each linker was defined as the residue span between the C-terminal boundary of the upstream domain and the N-terminal boundary of the downstream domain; accordingly, the KR–DH linker of ClbB-type modules captured the spacing between the KR domain and the internally positioned DH domain.

Because linker lengths were not assumed to follow a normal distribution, each linker region was compared across the three groups independently using the Kruskal–Wallis test. For linker regions with a significant omnibus result, pairwise differences between groups were evaluated using two-sided Dunn's post hoc tests with Bonferroni correction for multiple comparisons, and an adjusted $P < 0.05$ was considered significant. Median linker lengths were summarized per group and linker region, and the distributions were visualized as boxplots stratified by group and linker type.

All statistical analyses were performed in Python using SciPy and scikit-posthocs.

Declarations

Acknowledgments: We thank 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 cryo-EM data collection, and Dr. Ed Eng and Charlie Dubbeldam 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, and Joongwon Lee at KAIST for helping with the Boltz-2 structure predictions. Serial mass photometry was performed using a Mass photometer (Model: TwoMP, Refeyn) at Chungnam National University. The work is supported by grants from POSCO TJ Science Fellowship and National Research Foundation of Korea (NRF-RS-2024-00440614, NRF-RS-2024-00344154, NRF-2021R1C1C100656014) to JYK, National Research Foundation of Korea (NRF-RS-2024-00352229) to HUK, and National Research Foundation of Korea (NRF-2021R1C1C1011045, NRF-2022R1A6A1A03054419), and BK21 FOUR Project to CSK. The use of cryo-EM facilities of the CELINE consortium was supported by the National Research Foundation of Korea (NRF-RS-2024-00440614).

Author Contributions

J.K. and J.Y.K. conceptualized and initiated the project. J.K. and J.P. performed protein purification, biochemical assays, cryo-EM data collection, and data processing. H.W. and M.K. assisted J.K. in protein purification, biochemical assays, and cryo-EM grid preparation. J.K. and B.L. designed and performed bioinformatic analysis. J.K. and G.S.L. performed LC-MS analysis, and C.S.K. provided funding and facilities for MS experiments. J.Y.K. supervised the project, analyzed cryo-EM data with J.K., and obtained funding. J.K. and J.Y.K. wrote the original draft of the manuscript, and J.K., C.S.K., H.U.K., and J.Y.K. revised and edited the manuscript.

Competing interest declaration

The authors declare that they have no competing interests.

Data Availability

The cryo-EM maps and model coordinates have been deposited on Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB), respectively (apo-ClbB: EMD-69442; NADPH-bound ClbB: EMD-69437 and 24CH). Mass spectrometry and mass photometry data are presented in Supplementary Figs. 1c and 12, respectively.

References

1. Donia, M. S. & Fischbach, M. A. Small molecules from the human microbiota. *Science (1979).* **349**, (2015).
2. Hou, K. *et al.* Microbiota in health and diseases. *Signal Transduct. Target. Ther.* **7**, (2022).
3. Faïs, T., Delmas, J., Barnich, N., Bonnet, R. & Dalmasso, G. Colibactin: More than a new bacterial toxin. *Toxins (Basel).* **10**, (2018).
4. Balskus, E. P. Colibactin: Understanding an elusive gut bacterial genotoxin. *Nat. Prod. Rep.* **32**, 1534–1540 (2015).
5. Putze, J. *et al.* Genetic structure and distribution of the colibactin genomic island among members of the family Enterobacteriaceae. *Infect. Immun.* **77**, 4696–4703 (2009).
6. Nougayrède, J.-P. *et al.* Escherichia coli Induces DNA Double-Strand Breaks in Eukaryotic Cells. *Science (1979).* **313**, 848–851 (2006).
7. Secher, T., Samba-Louaka, A., Oswald, E. & Nougayrède, J. P. Escherichia coli Producing Colibactin Triggers Premature and Transmissible Senescence in Mammalian Cells. *PLoS One* **8**, 1–17 (2013).
8. Cuevas-Ramos, G. *et al.* Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 11537–11542 (2010).
9. Carlson, E. S. *et al.* The specificity and structure of DNA cross-linking by the gut bacterial genotoxin colibactin. *Science (1979).* **390**, (2025).
10. Buc, E. *et al.* High Prevalence of Mucosa-Associated E. coli Producing Cyclomodulin and Genotoxin in Colon Cancer. *PLoS One* **8**, (2013).
11. McCarthy, A. J. *et al.* The genotoxin colibactin is a determinant of virulence in Escherichia coli K1 experimental neonatal systemic infection. *Infect. Immun.* **83**, 3704–3711 (2015).
12. Dubinsky, V., Dotan, I. & Gophna, U. Carriage of Colibactin-producing Bacteria and Colorectal Cancer Risk. *Trends Microbiol.* **28**, 874–876 (2020).
13. Xue, M. *et al.* Structure elucidation of colibactin and its DNA cross-links. *Science (1979).* **365**, (2019).
14. Wilson, M. R. *et al.* The human gut bacterial genotoxin colibactin alkylates DNA. *Science (1979).* **363**, (2019).
15. Soohoo, A. M., Cogan, D. P., Brodsky, K. L. & Khosla, C. Structure and Mechanisms of Assembly-Line Polyketide Synthases. *Annu. Rev. Biochem.* **93**, 471–498 (2024).
16. Skiba, M. A. *et al.* Domain Organization and Active Site Architecture of a Polyketide Synthase C-methyltransferase. *ACS Chem. Biol.* **11**, 3319–3327 (2016).

17. Skiba, M. A. *et al.* Structural Basis of Polyketide Synthase O-Methylation. *ACS Chem. Biol.* **13**, 3221–3228 (2018).
18. Süssmuth, R. D. & Mainz, A. Nonribosomal Peptide Synthesis—Principles and Prospects. *Angew. Chem. Int. Ed.* **56**, 3770–3821 (2017).
19. Smith, J. L. & Sherman, D. H. An Enzyme Assembly Line. *Science (1979).* **321**, 1304–1305 (2008).
20. Brachmann, A. O. *et al.* Colibactin biosynthesis and biological activity depend on the rare aminomalonyl polyketide precursor. *Chem. Commun.* **51**, 13138–13141 (2015).
21. Zha, L., Wilson, M. R., Brotherton, C. A. & Balskus, E. P. Characterization of Polyketide Synthase Machinery from the pks Island Facilitates Isolation of a Candidate Precolibactin. *ACS Chem. Biol.* **11**, 1287–1295 (2016).
22. Zha, L. *et al.* Colibactin assembly line enzymes use S-adenosylmethionine to build a cyclopropane ring. *Nat. Chem. Biol.* **13**, 1063–1065 (2017).
23. Bonhomme, S., Contreras-Martel, C., Dessen, A. & Macheboeuf, P. Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin. *Structure* **31**, 700-712.e4 (2023).
24. Sandmann, A., Sasse, F. & Müller, R. Identification and Analysis of the Core Biosynthetic Machinery of Tubulysin, a Potent Cytotoxin with Potential Anticancer Activity. *Chem. Biol.* **11**, 1071–1079 (2004).
25. Maier, T., Leibundgut, M. & Ban, N. The Crystal Structure of a Mammalian Fatty Acid Synthase. *Science (1979).* **321**, 1315–1322 (2008).
26. Herbst, D. A., Jakob, R. P., Zähringer, F. & Maier, T. Mycocerosic acid synthase exemplifies the architecture of reducing polyketide synthases. *Nature* **531**, 533–537 (2016).
27. Wang, J. *et al.* Structural basis for the biosynthesis of lovastatin. *Nat. Commun.* **12**, 1–10 (2021).
28. Wang, J. *et al.* C–N bond formation by a polyketide synthase. *Nat. Commun.* **14**, (2023).
29. Jiang, Z. *et al.* Visualizing acyl carrier protein interactions within a crosslinked type I polyketide synthase. *Nat. Commun.* **16**, (2025).
30. McCullough, T. M. *et al.* Structure of a modular polyketide synthase reducing region. *Structure* **31**, 1109-1120.e3 (2023).
31. La Clair, J. J., Foley, T. L., Schegg, T. R., Regan, C. M. & Burkart, M. D. Manipulation of Carrier Proteins in Antibiotic Biosynthesis. *Chem. Biol.* **11**, 195–201 (2004).
32. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
33. Tang, Y., Kim, C. Y., Mathews, I. I., Cane, D. E. & Khosla, C. The 2.7-Å crystal structure of a 194-kDa homodimeric fragment of the 6-deoxyerythronolide B synthase. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 11124–11129 (2006).
34. Nakane, T., Kimanius, D., Lindahl, E. & Scheres, S. H. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. *Elife* **7**, e36861–e36879 (2018).

35. Schultz, K. *et al.* Snapshots of acyl carrier protein shuttling in human fatty acid synthase. *Nature* **641**, 520–528 (2025).
36. Herbst, D. A. *et al.* The structural organization of substrate loading in iterative polyketide synthases. *Nat. Chem. Biol.* **14**, 474–479 (2018).
37. Bagde, S. R., Mathews, I. I., Fromme, J. C. & Kim, C.-Y. Modular polyketide synthase contains two reaction chambers that operate asynchronously. *Science (1979).* **374**, 723–729 (2021).
38. Cogan, D. P. *et al.* Mapping the catalytic conformations of an assembly-line polyketide synthase module. *Science (1979).* **374**, 729–734 (2021).
39. Mindrebo, J. T. *et al.* Gating mechanism of elongating β-ketoacyl-ACP synthases. *Nat. Commun.* **11**, 1727 (2020).
40. Mindrebo, J. T. *et al.* Structure and Mechanistic Analyses of the Gating Mechanism of Elongating Ketosynthases. *ACS Catal.* **11**, 6787–6799 (2021).
41. Park, H., Kevany, B. M., Dyer, D. H., Thomas, M. G. & Forest, K. T. A polyketide synthase acyltransferase domain structure suggests a recognition mechanism for its hydroxymalonyl-acyl carrier protein substrate. *PLoS One* **9**, (2014).
42. Kim, M. *et al.* Structural study on human microbiome-derived polyketide synthases that assemble genotoxic colibactin. *Structure* **33**, 1208-1223.e5 (2025).
43. Glaeser, R. M. Proteins, interfaces, and cryo-EM grids. *Curr. Opin. Colloid Interface Sci.* **34**, 1–8 (2018).
44. D'Imprima, E. *et al.* Protein denaturation at the air-water interface and how to prevent it. *Elife* **8**, e42747 (2019).
45. Liu, C. *et al.* Substrate-bound structures of a ketoreductase from amphotericin modular polyketide synthase. *J. Struct. Biol.* **203**, 135–141 (2018).
46. Keatinge-Clay, A. Crystal Structure of the Erythromycin Polyketide Synthase Dehydratase. *J. Mol. Biol.* **384**, 941–953 (2008).
47. Keatinge-Clay, A. T. The structures of type I polyketide synthases. *Nat. Prod. Rep.* **29**, 1050–1073 (2012).
48. Zheng, J., Gay, D. C., Demeler, B., White, M. A. & Keatinge-Clay, A. T. Divergence of multimodular polyketide synthases revealed by a didomain structure. *Nat. Chem. Biol.* **8**, 615–621 (2012).
49. Kwan, D. H. *et al.* Prediction and Manipulation of the Stereochemistry of Enoylreduction in Modular Polyketide Synthases. *Chem. Biol.* **15**, 1231–1240 (2008).
50. Kwan, D. H. & Leadlay, P. F. Mutagenesis of a modular polyketide synthase enoylreductase domain reveals insights into catalysis and stereospecificity. *ACS Chem. Biol.* **5**, 829–838 (2010).
51. Keatinge-Clay, A. T. & Stroud, R. M. The Structure of a Ketoreductase Determines the Organization of the β-Carbon Processing Enzymes of Modular Polyketide Synthases. *Structure* **14**, 737–748 (2006).
52. Blin, K. *et al.* AntiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. *Nucleic Acids Res.* **51**, W46–W50 (2023).

53. Blin, K., Shaw, S., Medema, M. H. & Weber, T. The antiSMASH database version 4: additional genomes and BGCs, new sequence-based searches and more. *Nucleic Acids Res.* **52**, D586–D589 (2024).
54. Jones, P. *et al.* InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
55. Blum, M. *et al.* InterPro: The protein sequence classification resource in 2025. *Nucleic Acids Res.* **53**, D444–D456 (2025).
56. Passaro, S. *et al.* Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. *bioRxiv* (2025) doi:10.1101/2025.06.14.659707.
57. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
58. Balbuena, J. A., Míguez-Lozano, R. & Blasco-Costa, I. PACo: A Novel Procrustes Application to Cophylogenetic Analysis. *PLoS One* **8**, (2013).
59. Scheres, S. H. W. RELION: Implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
60. Zheng, S. Q. *et al.* MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
61. Heymann, J. B. Guidelines for using Bsoft for high resolution reconstruction and validation of biomolecular structures from electron micrographs. *Protein Sci.* **27**, 159–171 (2018).
62. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
63. Williams, C. J. *et al.* MolProbity: More and better reference data for improved all-atom structure validation. *Protein Sci.* **27**, 293–315 (2018).
64. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
65. Sayers, E. W. *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
66. Tamura, K., Stecher, G. & Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **38**, 3022–3027 (2021).
67. Navarro-Muñoz, J. C. *et al.* A computational framework to explore large-scale biosynthetic diversity. *Nat. Chem. Biol.* **16**, 60–68 (2020).
68. Shannon, P. *et al.* Cytoscape: A software Environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
69. Zdouc, M. M. *et al.* MIBiG 4.0: advancing biosynthetic gene cluster curation through global collaboration. *Nucleic Acids Res.* **53**, D678–D690 (2025).
70. Von Meijenfeldt, F. A. B., Arkhipova, K., Cambuy, D. D., Coutinho, F. H. & Dutilh, B. E. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biol.* **20**,

(2019).

71. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
72. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).
73. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
74. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermiin, L. S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
75. Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010).
76. Thi Hoang, D. *et al.* UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol* **35**, 518–522 (2017).

Figures

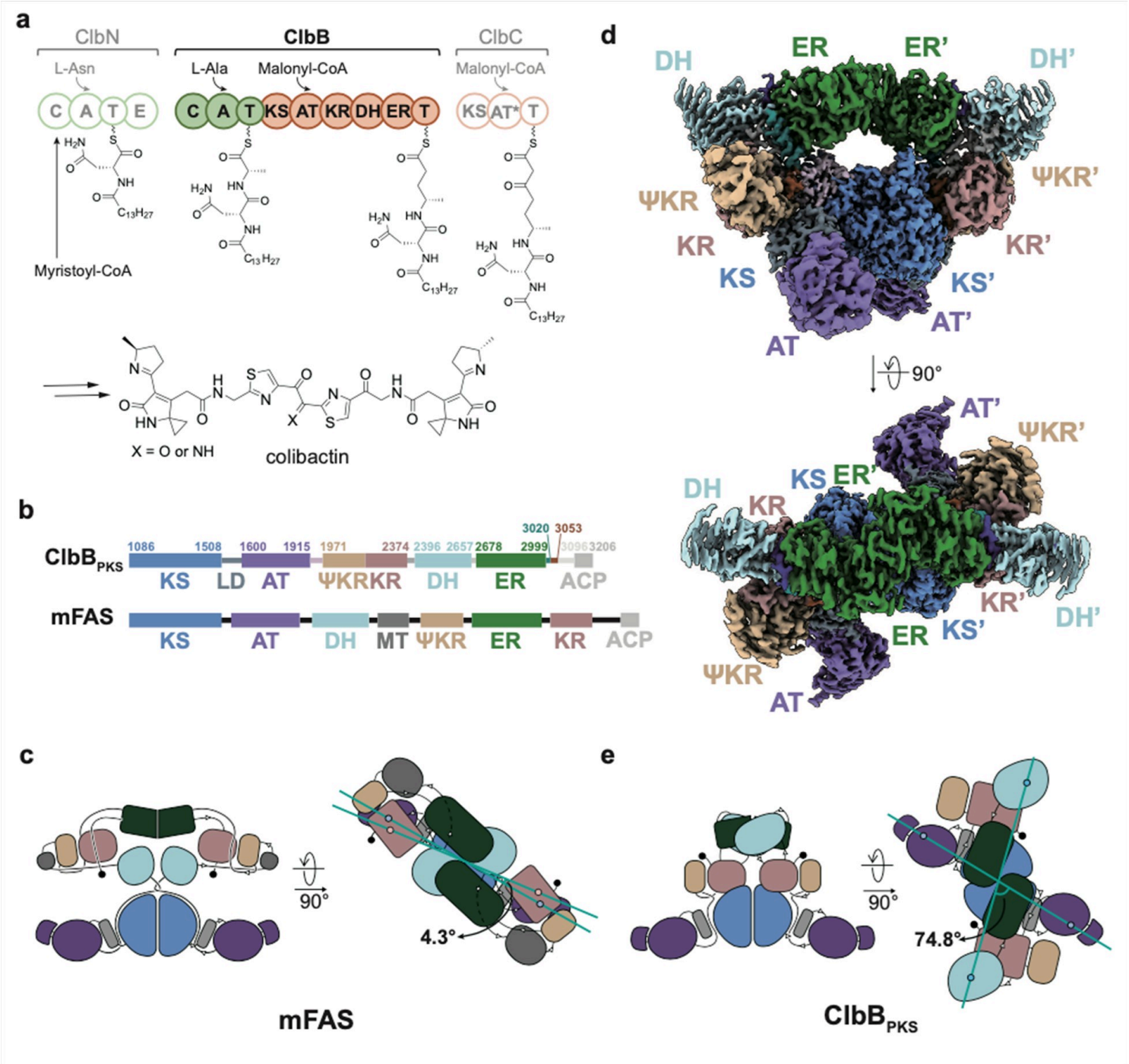

Figure 1

Cryo-EM structure of NADPH-bound ClbB$_{PKS}$ and schematic comparison with mFAS. a, Early steps of the colibactin biosynthetic pathway including ClbB. The 1D domain organization of ClbB is shown. **b,** Domain organization of ClbB$_{PKS}$ and mFAS. Residue numbers for domain boundaries are shown for ClbB$_{PKS}$. **c**, Schematic representation of mFAS. The axes of the condensing and reducing regions are defined by lines connecting the centers-of-mass of the AT and KR domains, because they are the outermost boundaries of the condensing and reducing regions. The inter-axial angles are indicated. **d,** Local-filtered cryo-EM map of NADPH-bound ClbB$_{PKS}$ in front (upper) and top (lower) views. **e,** Schematic representations of ClbB$_{PKS}$. The axes of the condensing and reducing regions are defined by lines

connecting the centers-of-mass of the AT and DH domains in $ClbB_{PKS}$, because they are the outermost boundaries of the condensing and reducing regions. The inter-axial angles are indicated. Abbreviation: C, condensation; A, adenylation; T, thiolation; KS, ketosynthase; LD, linker domain; AT, acyltransferase; KR, ketoreductase; DH, dehydratase; ER, enoylreductase; ACP, acyl carrier protein.

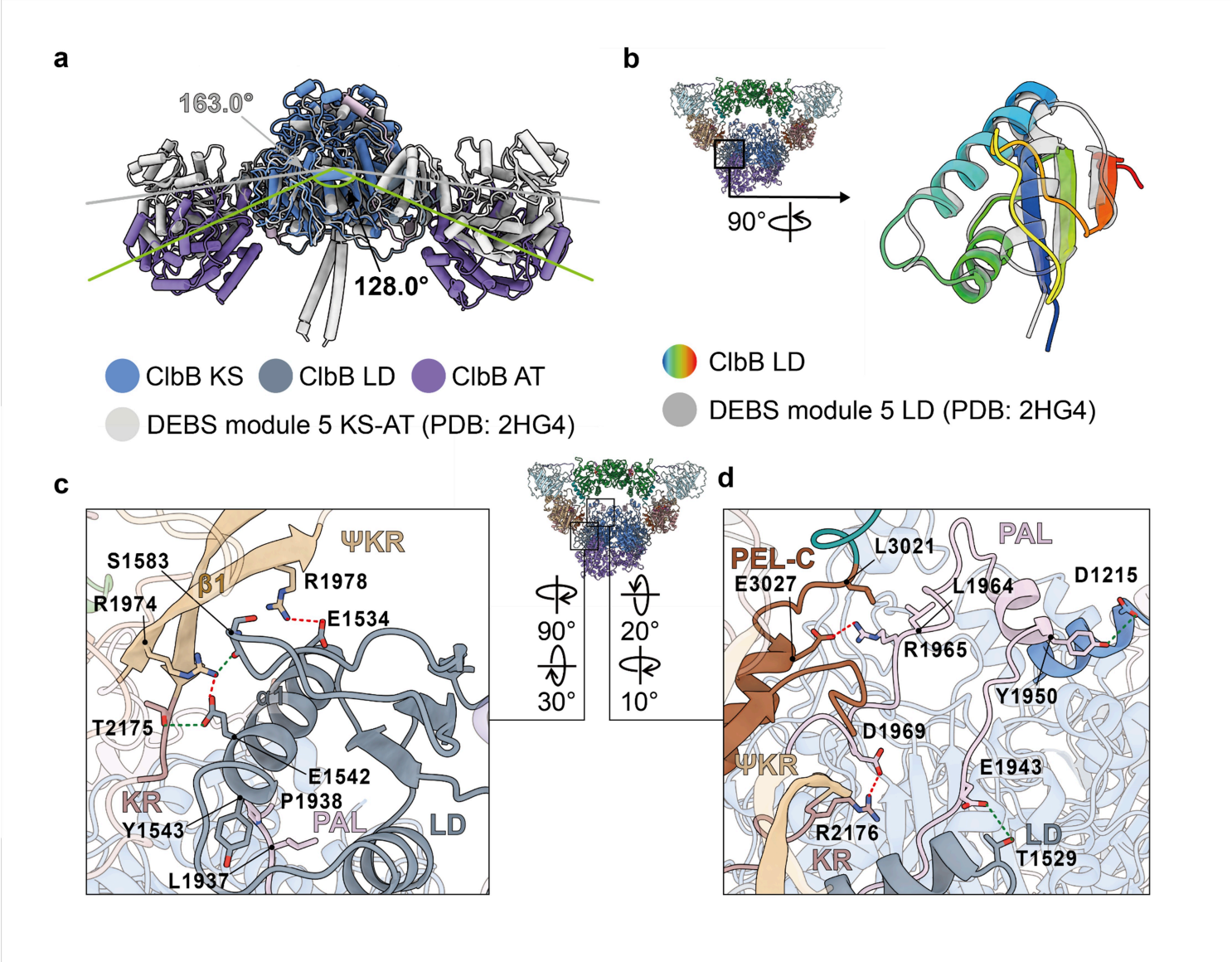

Figure 2

The condensing region of $ClbB_{PKS}$. a, Superposition of ClbB and DEBS module 5 KS-LD-AT (PDB 2HG4), aligned by the KS dimer. Structures are colored as indicated. The inter-axial angle between the two KS-LD-AT domains is indicated in pale green (ClbB) and gray (DEBS module 5). **b,** Superposition of the LDs from ClbB and DEBS module 5 (PDB 2HG4). ClbB LD is colored in rainbow (N-terminal, blue; C-terminal, red); DEBS module 5 LD is light gray with 50% transparency. **c,** Close-up view of the ClbB LD-KR interface. Key residues and interactions are indicated as hydrogen bonds (green dotted lines) or salt bridges (red dotted lines). **d,** Close-up view of the post-AT linker (PAL). Hydrogen bonds (green dotted lines) and salt bridges (red dotted lines) are indicated.

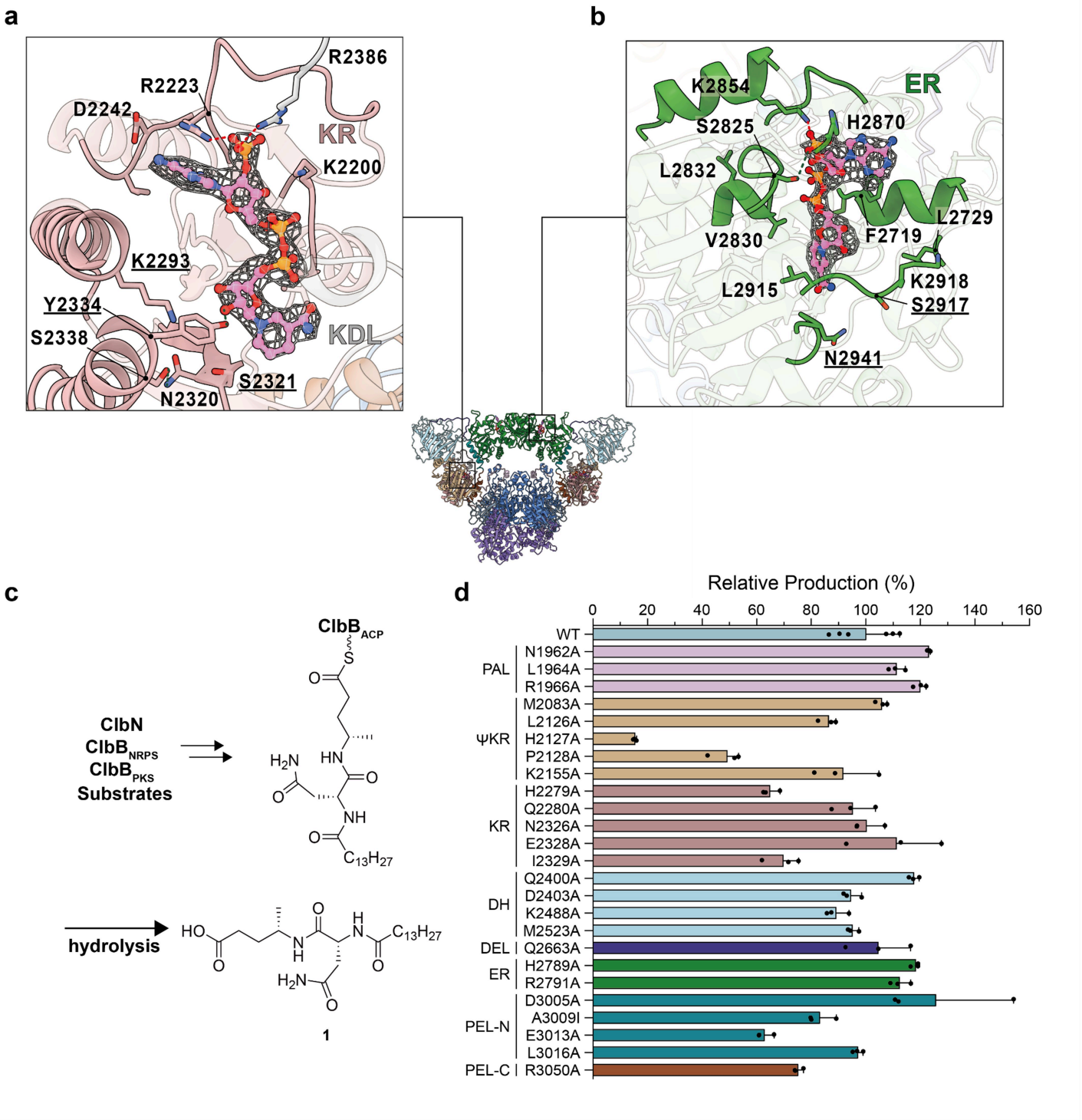

Figure 3

NADPH-binding sites of ClbB$_{PKS}$ and LC-MS analysis of ClbB$_{PKS}$ KR mutants.

Close-up views of the NADPH-binding sites of **a,** KR; and **b**, ER. The local-filtered NADPH density is shown. Residues involved in NADPH binding**,** including the cofactor-binding motif ($_{2828}$GGVGLA$_{2833}$), as well as catalytic motifs are highlighted. Known catalytic residues are underlined. **c**, Chemical Structure of the colibactin precursor produced in the in vitro biosynthesis reaction. **d**, Relative production of the

intermediate (**1**) shown in **c**, quantified by LC-MS. All in vitro biosynthesis reactions were performed in triplicate.

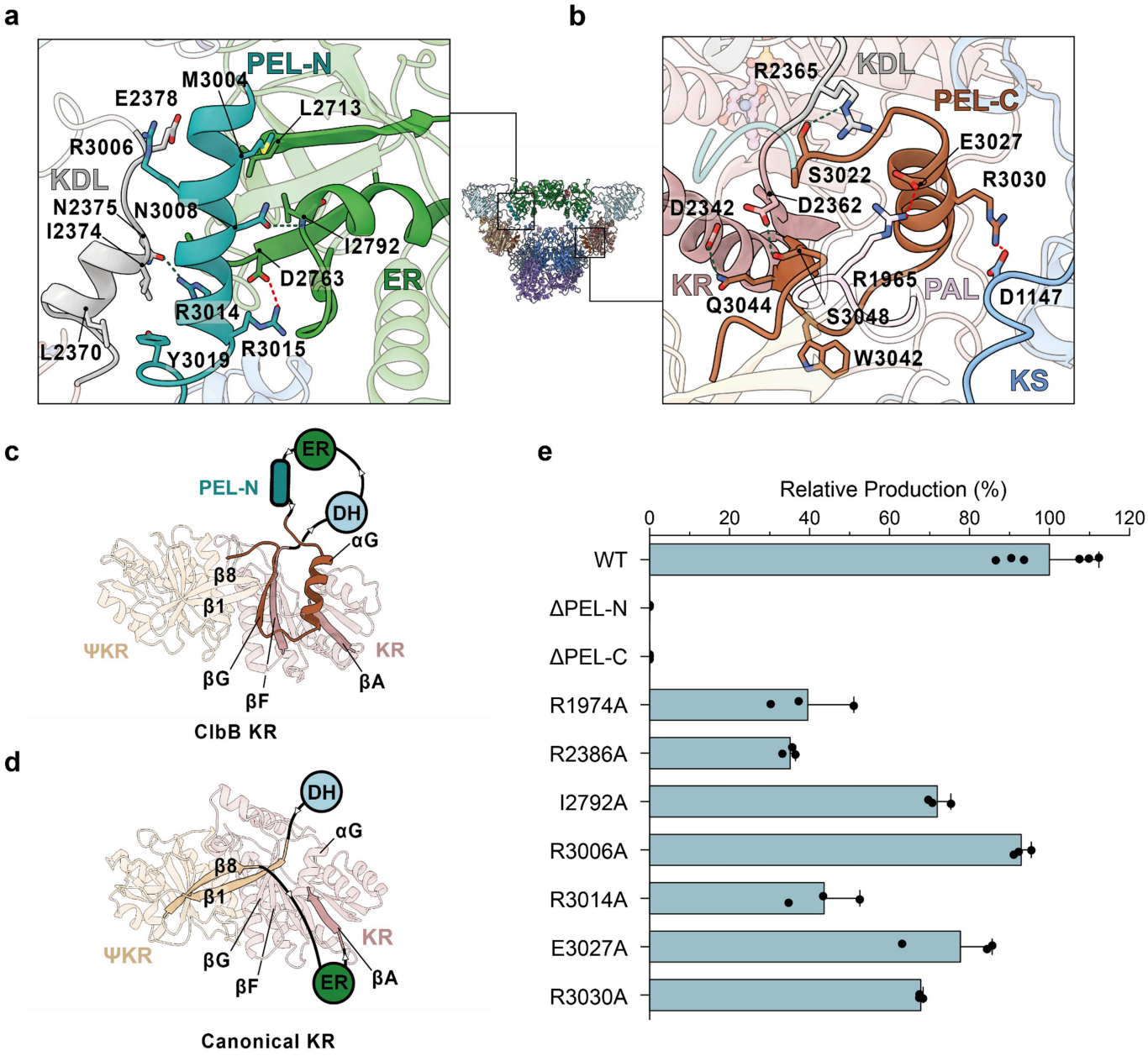

Figure 4

PEL interactions, DH-ER insertion within the KR domain, and LC-MS analysis of $ClbB_{PKS}$ ER/PEL mutants.

Close-up views of **a,** N-terminal post-ER linker (PEL-N), and **b,** the C-terminal post-ER linker (PEL-C). Key residues and interactions are highlighted (hydrogen bond, green; salt bridge, red). **c,** Schematic

illustration of insertion of the DH-ER didomain into the KR domain in ClbB. Structural elements adjacent to the insertion site are labeled. **d**, Schematic illustration of the canonical connection between KR and DH/ER domains in mPKSs. **e,** LC-MS analysis of products from in vitro colibactin precursor biosynthesis using wild-type and mutant $ClbB_{PKS}$.

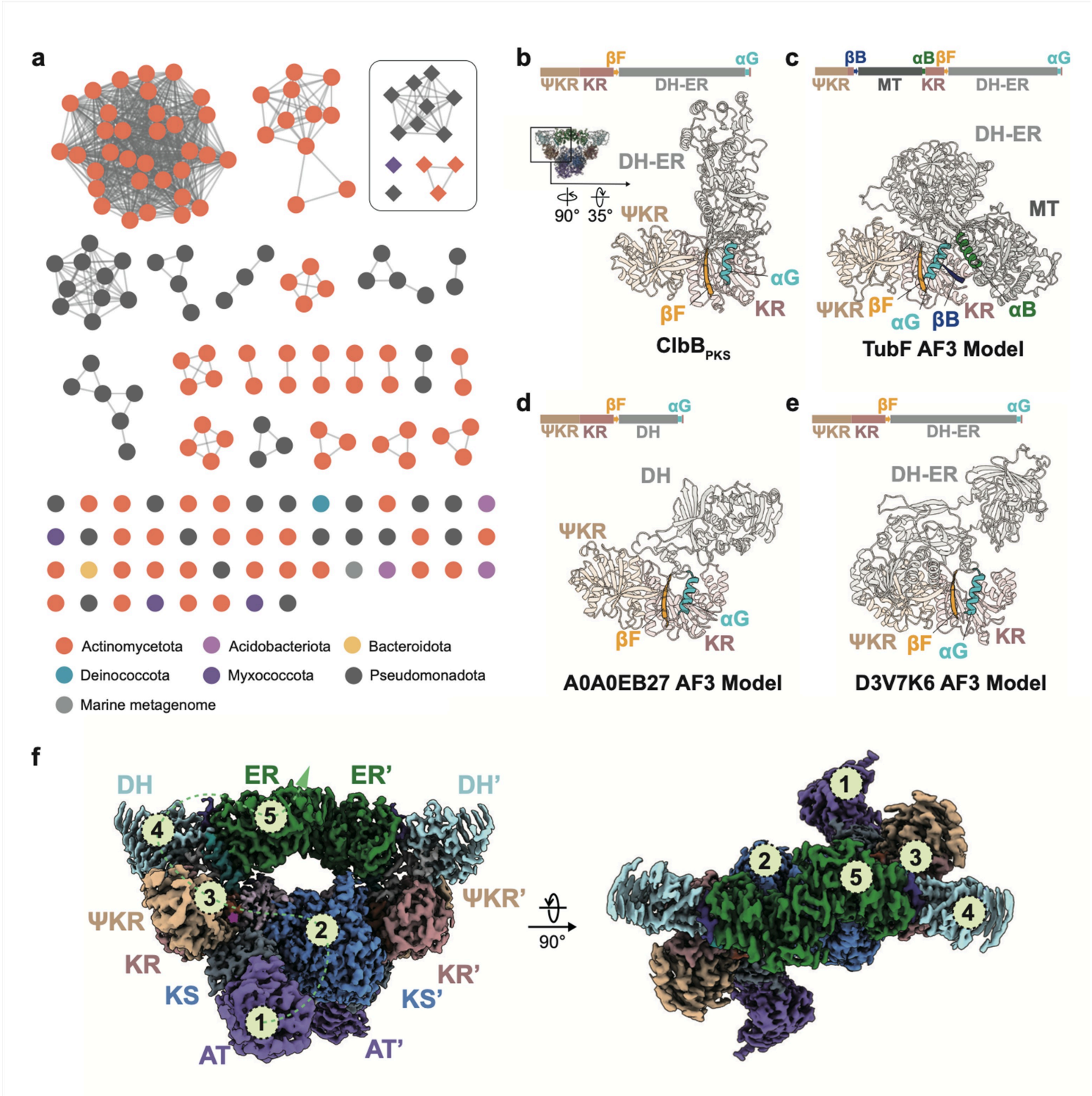

Figure 5

Bioinformatic and AlphaFold3 analyses of ClbB-type PKSs. **a,** Sequence similarity network of ClbB-type PKSs. Each node represents a KS–AT–KR–DH(–ER)–containing module. Nodes in circles indicate modules without an ER domain, and nodes in diamonds show modules with an ER domain and are included in a box in the right upper corner. Nodes are connected when the following three criteria are met: (i) percent identity > 60%, (ii) E-value < 1×10^{-30}, and (iii) sequence coverage > 0.9. Reducing-domain organization of **b,** $ClbB_{PKS}$, **c,** TubF, **d,** A0A0EB27, and **e,** D3V7K6. The ΨKR and KR domains are colored as in **Fig. 4c** with 50% transparency. The N- and C-terminal DH-ER insertion sites are colored orange and cyan, respectively, whereas the N- and C-terminal MT insertion sites are colored blue and green, respectively.

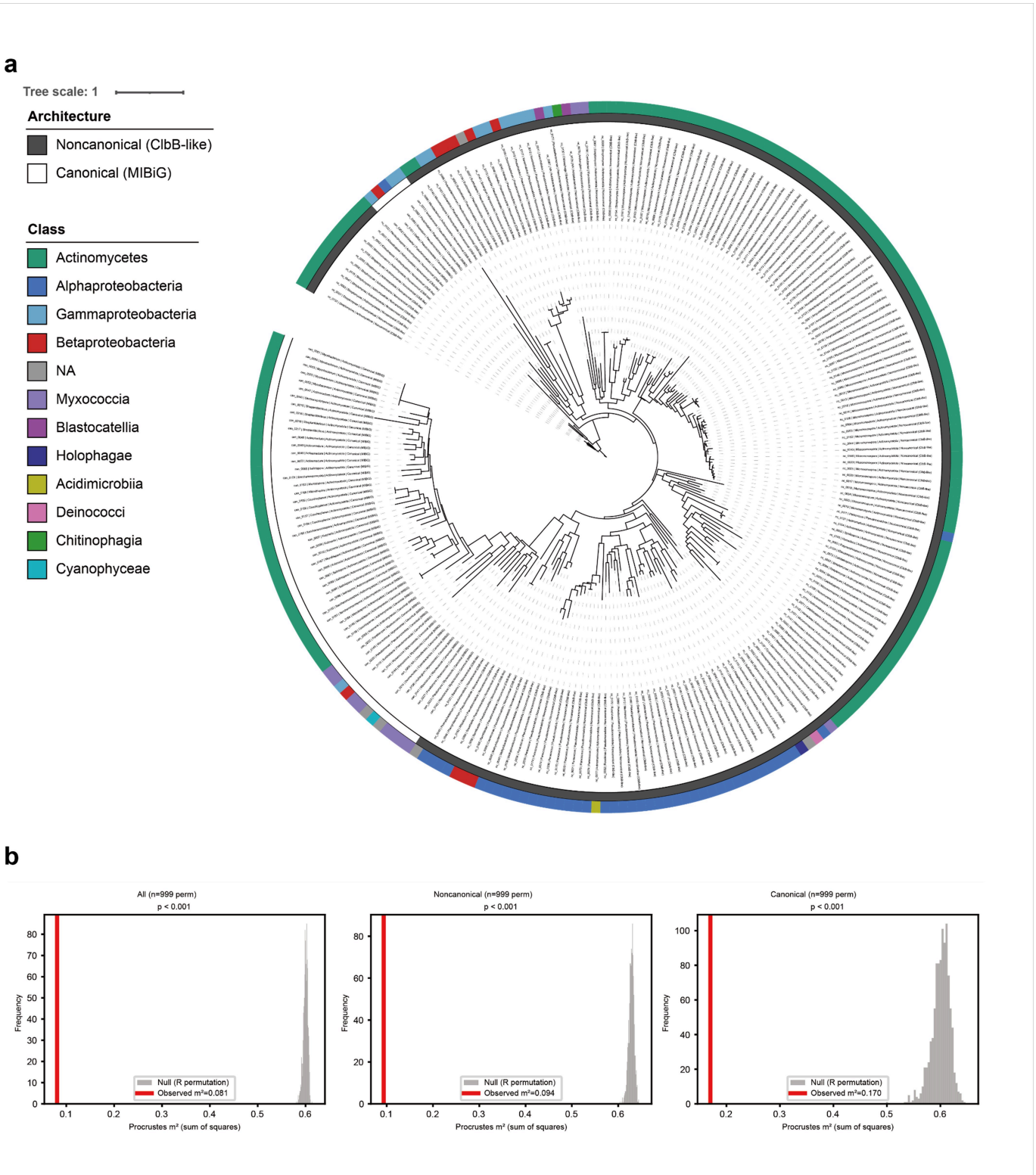

Figure 6

Phylogenetic and Cophylogenetic analysis of KR and DH domains. a, Maximum-likelihood phylogeny of KR domains from noncanonical, ClbB-like PKS modules and canonical bacterial reference modules obtained from MIBiG 4.0. Concentric rings indicate module architecture (noncanonical ClbB-type, dark gray; canonical MIBiG, white) and host taxonomic class (colored as in the legend). NA, unassigned class. **b,** Procrustean cophylogenetic test (PACo) of congruence between paired KR and DH domain trees. Gray

histograms show the null m² distribution from 999 permutations; red lines mark the observed m². Lower m² indicates greater congruence. For all paired modules (n = 236), noncanonical only (n = 179), and canonical only (n = 57), the observed m² (0.081, 0.094, 0.170) fell below the null distribution (all P = 0.0010), indicating significantly congruent KR and DH evolutionary histories.

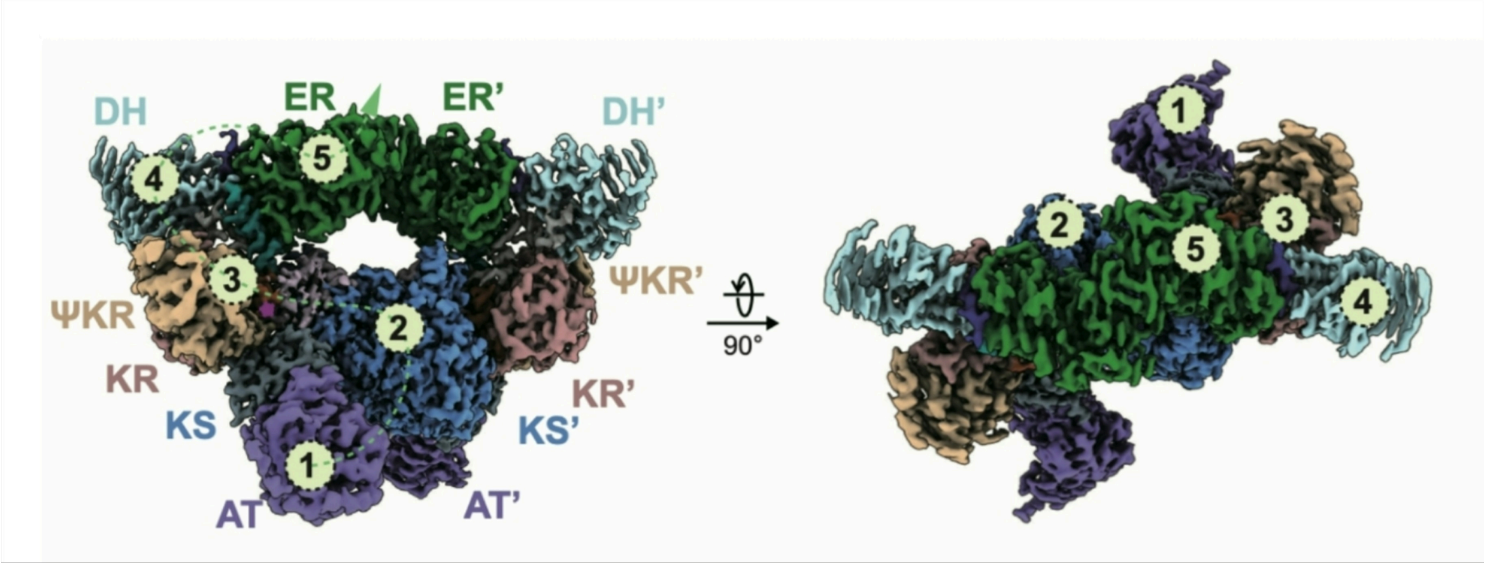

Figure 7

Putative ACP trajectory in ClbB$_{PKS}$. Putative ACP docking sites on ClbB$_{PKS}$ are indicated by pale green circles. The position of the last modeled residue is marked by a purple circle with a dotted outline. The green dotted arrow marks the predicted ACP trajectory. Numbers within the circles denote the order of catalytic reactions. Active sites on the opposite protomer are shown on the right in a top view.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInfo.docx
- D1300070879valreportfullannotateP1.pdf
- 20260715nrreportingsummary.pdf
- 20260715supplerevised.docx
- Supplementaryvideo1.mp4
- NCOMMS26016672ARS.pdf