Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin

Sarah Bonhomme, Carlos Contreras-Martel, Andréa Dessen, Pauline Macheboeuf

Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin

Sarah Bonhomme[1], Carlos Contreras-Martel[1], Andréa Dessen[1], Pauline Macheboeuf[1, 2*]

[1]Univ. Grenoble Alpes, CNRS, CEA, Institut de Biologie Structurale (IBS), Bacterial Pathogenesis Group, F-38000 Grenoble, France.
[2]Lead contact
*Correspondence : pauline.macheboeuf@ibs.fr

Abstract

The genotoxin colibactin produced by *Escherichia coli* is involved in the development of colorectal cancers. This secondary metabolite is synthesized by a multi-protein machinery, mainly composed of non-ribosomal peptide synthetase (NRPS)/polyketide synthase (PKS) enzymes. In order to decipher the function of a PKS-NRPS hybrid enzyme implicated in a key step of colibactin biosynthesis, we conducted an extensive structural characterization of the ClbK megaenzyme. Here we present the crystal structure of the complete *trans*-AT PKS module of ClbK showing structural specificities of hybrid enzymes. Additionally, we report the SAXS solution structure of the full-length ClbK hybrid that reveals a dimeric organization as well as several catalytic chambers. These results provide a structural framework for the transfer of a colibactin precursor through a PKS-NRPS hybrid enzyme and can pave the way for re-engineering PKS-NRPS hybrid megaenzymes to generate diverse metabolites with many applications.

Introduction

Escherichia coli is both a commensal of the normal gut and the most common causative agent of infectious diseases caused by a Gram-negative organism. This versatility is linked to the production of a vast array of virulence factors and effectors that manipulate different host cell functions. Some human-associated *E. coli* strains that belong to the phylogenetic group B2 have been shown to trigger high genotoxicity by inducing DNA damage *in vitro* [1] and *in vivo* [2], leading to tumorigenic effects. Strains that generate this effect carry the 54-kb *pks* pathogenicity island and have mainly been associated to the development of colorectal cancers [3,4] and less frequently to prostate and bladder cancers [3–6]. The genotoxic action of these strains is correlated to the production of the genotoxin colibactin by the *pks* biosynthetic gene cluster (BGC) (Fig. 1A). Despite colibactin being an unstable molecule that long resisted structural determination, its recently characterized structure reveals two cyclopropane rings involved in DNA alkylation [7] and two thiazole rings that are suspected to act as DNA intercalating agents [8]. Moreover, the heterodimeric structure of colibactin has been shown to be responsible for inter-strand crosslinks [9].

Colibactin, similarly to many other natural products, is assembled by an assembly line composed of non-ribosomal peptide synthetase (NRPS) and polyketide synthase (PKS) megaenzymes (Fig. 1A) encoded by the *pks* BGC. The NRPS and PKS enzymes share the same assembly line logic and display a common general organization. Each NRPS or PKS megaenzyme is assembled in one or several module(s), a module being defined as a unit that adds one monomer (an amino acid for NRPS and an acyl unit for PKS) to the final natural metabolite. Each module is further divided into domains, each one catalyzing a specific enzymatic activity. In the NRPS system, the adenylation (A) domain activates an amino acid that is further transferred to a phosphopantheteine (PPant) arm of a peptidyl carrier protein (PCP) to form a thioester intermediate [10]. The condensation (C) domain catalyzes the formation of an amide bond between the upstream aminoacyl group from the donor PCP domain and the downstream aminoacyl group bound to the acceptor PCP. Alternatively, a cyclization (Cy) domain can catalyze both the condensation reaction and a cyclodehydration reaction to form a thiazoline or an oxazoline ring [11]. Tailoring domains can further chemically modify the amino acid unit along its incorporation [12].

In the PKS system, the acyltransferase (AT) domain is responsible for selecting the extender acyl substrate that is further transferred to the PPant arm of the acyl carrier protein (ACP). Next, the ketosynthase (KS) domain catalyzes the trans-thioesterification of the donor polyketide chain from the upstream ACP domain on an active site cysteine residue. Once acylated, the KS domain catalyzes a Claisen type condensation reaction between the acceptor substrate attached to the

downstream ACP domain and the donor acyl chain to form a β-ketothioester that can be further modified by tailoring domains like ketoreductases, dehydratases, enoylreductases or methyltransferases [13].

Modular PKSs can be divided into two distinct classes. In *cis*-AT PKSs, each module contains a dedicated AT domain responsible for a specific acyl chain recruitment [14]. In contrast, the *trans*-AT PKSs employ a standalone AT domain that supplies the ACP domains in the assembly line with an acyl unit [15]. Interestingly, the *pks* BGC is a mixed system with some *cis*-AT PKS modules and some *trans*-AT modules (Fig. 1A).

The multi-domain organization of NRPS and PKS machineries is at the source of the inherent flexibility of the multienzymes making the structural characterization of such proteins very challenging, in particular in what relates to the visualization of the intrinsically flexible carrier protein domain. The structures of NRPS megaenzymes have been extensively studied for years but were first limited to isolated domains until recently, when the structures of full modules or even dimodular NRPS systems were solved (Drake et al., 2016; Kreitler et al., 2019; Miller et al., 2016; Reimer et al., 2016, 2019; Tarry et al., 2017; reviewed in Bonhomme et al., 2021). NRPSs are highly flexible monomers, whose structural elucidation has required the use of substrates, substrate analogues and dead-end inhibitors [22]. Very recently, a breakthrough in the NRPS field revealed the unexpected dimerization of certain NRPS modules [23–25], raising the question of the diversity of NRPS quaternary structures. PKS structures have also long been limited to isolated domains or didomains until recently, when high resolution structures of modular *cis*-AT PKSs including the ACP domains were solved [26–29]. Additionally, several high-resolution structures of iterative PKSs were also recently solved [30,31]. Modular PKSs have been shown to function as homodimeric enzymes with divergent module shapes ranging from elongated [30,32–34] to arch-shaped [28]. In comparison to *cis*-AT PKSs, there is poor structural knowledge on *trans*-AT PKSs [35,36] and until now, no high resolution structure of a full module from *trans*-AT PKSs is available.

Because NRPS and PKS enzymes share the same assembly line organization, many hybrid systems containing both NRPS and PKS modules have been identified. These systems dramatically increase the complexity of the synthesized hybrid molecules that incorporate both acyl and aminoacyl units [37]. In these mixed organizations, two different types of interface have been described: the tethered type where the mixed modules are present on the same polypeptide and the separate type where the modules are located on different polypeptides [37]. In this work, the term 'NRPS-PKS hybrid' designates a tethered-type interface composed of a NRPS module in N-terminus fused to a PKS module in C-terminus, while the term 'NRPS/PKS hybrid' defines a separate type interface composed of a PKS polypeptide following an NRPS polypeptide in the assembly line. Structural information on hybrids from the tethered type is limited to the structures of domains [38] or a single module [31]; thus, a full-length tethered hybrid is yet to be structurally characterized. The colibactin-producing BGC encodes nineteen proteins that are responsible for the biosynthesis of the toxin and amongst those, eight of them belong to the NRPS/PKS family with three NRPSs, three PKSs and two hybrid proteins of the tethered type, namely ClbK and ClbB (Fig. 1A).

We focused our attention on the ClbK PKS-NRPS hybrid that incorporates an unstable α-aminoketone group and one of the two thiazole rings into colibactin (Fig. 1A) and thus, may play an important role in its genotoxic potential. ClbK is composed of a N-terminal *trans*-AT PKS module with a KS domain, a non-functional AT* domain and an ACP domain (Fig. 1B). The AT* is complemented by the standalone *trans*-AT ClbG which loads an unusual aminomalonyl unit onto the ClbK ACP domain (Fig. 1C) [39,40]. The C-terminal NRPS module of ClbK is composed of

cyclization, adenylation, oxidase and peptidyl carrier protein domains (Fig. 1B), responsible for both the incorporation of a cysteine unit to form the second thiazoline ring of colibactin and the oxidation of both thiazolines into thiazole rings [41]. Remarkably, the ClbK PKS module presents both a separate interface with the upstream NRPS module of ClbJ (Fig. 1A and 1C) and a tethered interface with its own NRPS module at the C-terminus of the ClbK polypeptide, increasing the complexity of its assembly.

Here we present the structural characterization of the ClbK PKS-NRPS hybrid with the crystal structure of a complete module from a *trans*-AT PKS and the overall architecture of the full-length dimodular hybrid megaenzyme which could pave the way for the comprehension of colibactin biosynthesis.

Results

Crystal structure of the ClbK PKS module

To initiate the structural characterization of the ClbK PKS-NRPS hybrid, we crystallized the ClbK-PKS module (residues 1-787) composed of KS (1-423), linker (LD, 424-511), truncated AT* (512-659), post-AT* linker (PAL, 660-694) and ACP (701-787) domains (Fig. 2A). Crystals diffracted X-rays to 2.98 Å resolution at the ESRF ID30A-1 beamline (France). We performed numerous molecular replacement trials with several PKS domains or didomains without success before using the AlphaFold2 prediction algorithm on the ClbK-PKS sequence [42]. The four best-ranked models from AlphaFold2 displayed a compact stable KS-AT* core and a flexible ACP domain (Fig. S1A) that was removed to generate a successful molecular replacement solution. Several rounds of refinement and reconstruction allowed us to fully reconstruct the structure of ClbK-PKS (Fig. 2B). Statistics are summarized in Table 1.

ClbK-PKS crystallized in space group C2221 with four molecules in the asymmetric unit. The crystalline arrangement shows a dimeric interface between two KS domains that is similar to the dimerization interfaces observed in classical PKS modules [26,27,31–34] (Fig. 2B and 2C), indicating that most probably two dimers crystallized in the asymmetric unit. The overall electron density allowed the unambiguous tracing of each monomer containing KS-LD-AT*-PAL domains and one ACP domain per dimer (Fig. S1B). The superposition of each homodimer of the asymmetric unit shows a rmsd value of 0.22 Å over C alphas with little movement from the KS-AT* domains whereas the ACP domain adopts slightly different localizations (red and orange in Fig. S1A). The overall dimeric arrangement of KS-AT* is centered around the two-fold rotational axis with the two AT* domains located at opposite sides of the KS dimer (Fig. 2B). The rather rigid KS-AT* core architecture is similar to the canonical extended conformation of DEBS *cis*-AT PKS modules from the erythromycin assembly line [27,32,33] (Fig. 2C) and OzmQ, a NRPS/PKS hybrid *trans*-AT PKS from the oxazolomycin gene cluster [43] but is strikingly different from the unique arch-shaped dimer of the full PKS module 5 of the pikromycin synthase (PikAIII) where the two AT domains are bent downward forming the two legs of an arch [28] (Fig. 2D).

Originality of the KS domain

The KS domain (blue in Fig. 2) adopts a thiolase fold with a five-layered αβαβα arrangement [44] and displays a classical catalytic triad composed of Cys167, His303 and His343 (Fig. 3A). The KS domain forms the dimer interface (2070 $Å^2$), as seen in the dimerization interface of other KS dimers [14]. Interestingly, the ClbK KS sequence lacks the helices that form the docking domains (DD) that are usually present in *cis*-AT PKS such as DEBS module 5, for example (Fig. 2C and Fig. S2).

The main structural difference between KS domains resides on variable loops located at the bottom of the KS, at the so-called "clasping loop" and "active site cap" identified by Lohman and coworkers (Fig. 3A and 3B) [43]. In ClbK-PKS, the clasping loop is located between residues 40-51 (Fig. S2). This region which is usually unstructured, folds into strands β2 and β3 in ClbK, (Fig. 3A, blue in Fig. 3B and pink in Fig. S2). In the crystal structure of the RhiE didomain composed of a KS domain and an inactive branching (B) domain, this clasping loop interacts with the B domain [35]. Similarly, in the cryo-EM structure of DEBS module 1, the clasping loop interacts with the downstream KR domain [27]. We thus suggest that the ClbK-PKS β2 and β3 strands might favor an interaction with one of the downstream ClbK domains. Adjacent to the clasping loop, the active site cap folds as an alpha-helix in DEBS module 5, as well as *trans*-AT PKS BaeKS5 (wheat in Fig. 3B) and other *cis*-AT PKSs but folds as the β8 strand in ClbK-PKS (cyan in Fig. 3 and pink in Fig. S2) to form a 3-stranded β-sheet with β2 and β3 from the clasping loop. In two other KS from hybrid systems, such as the NRPS/PKS OzmQ and NRPS-PKS TAS1 [38], the cap also folds as a strand (cyan box in Fig. S2 and wheat in Fig. 3B) that could be a specific feature for NRPS/PKS hybrids and that will be discussed below.

The examination of the KS surface revealed the presence of four substrate entrances, two on each side of the KS dimer (Fig. 3C). First, we noticed the presence of the classical side entrance that is seen in all PKS KS structures and that forms a tunnel at the KS dimer interface allowing access to the active Cys167 (Fig. 3C). Interestingly, the structure of ClbK-PKS also revealed the presence of a bottom entrance as seen in the PikAIII structure [28] that opens at roughly 90º from the side entrance through the bottom of the KS domain (Fig. 3C). This bottom entrance is usually closed by the helical active site cap in *cis*-AT PKSs (Fig. 3D) but the β8 strand at this position moves out of the way in ClbK-PKS and leads to the opening of this gate (Fig. 3C).

The ClbK KS domain accepts two bulky substrates, i.e the growing colibactin substrate that contains a thiazole ring and the large aminomalonyl extender substrate (Fig. 1C); both must enter the KS cavities. In *trans*-AT PKSs, the specificity of substrate recognition has been correlated to a relatively conserved residue immediately before the active site cysteine that is usually an alanine or a methionine [45]. However, ClbK-PKS displays a serine residue at the position preceding the active site Cys167. Indeed, it has been found that in *trans*-AT PKSs that incorporate a thiazole or an oxazole intermediate, this residue is rather a small glycine or serine that would prevent steric clashes within the binding pocket to incorporate the heterocycle (Fig. S3) [45]. Additionally, another possible adaptation of KS to accommodate large amino acid-containing substrates is the folding of the active site cap into a beta strand mentioned above. As already hypothesized by the authors of the TAS1 structure, we propose that the absence of the active site cap helix could accommodate bulky substrates such as amino acid-containing substrates provided by NRPS modules [38].

The truncated acyltransferase domain and its surrounding linkers

Classical AT domains from *cis*-AT and *trans*-AT PKSs are composed of large (~240 residues) and small (~60 residues) subdomains. The conserved catalytic residues are located at the interface between the small and the large AT subdomains (see DEBS module 5 in Fig. S4). The sequence analysis of the ClbK AT* domain indicated the absence of active site AT residues, classifying it as a non-functional AT* domain [1]. Surprisingly, the ClbK-PKS structure reveals that the AT* domain folds as an independent domain composed mainly of α-helices and it shares several features with classical AT domains (Fig. S4). The AT* lacks the entire small subdomain as well as some sections of the large subdomain with an overall 140-residue truncation (Fig. S4). Therefore, the ClbK-AT active site is completely disrupted with none of the three essential active site motifs present,

providing an explanation for its non-functionality and the necessity for the *trans*-AT domain ClbG to load the ClbK ACP domain during colibactin biosynthesis [40].

These truncated AT domains are common in *trans*-AT PKSs such as those observed in BaeKS5 and the NRPS/PKS hybrid *trans*-AT PKS OzmQ (Fig. S4). BaeKS5 possesses three conserved helices as substitutes for the AT domain, forming the LINKS motif, suggested as being involved in inter-modular quaternary arrangements [46]. ClbK-PKS does not possess this motif but rather a larger insertion that could derive from an evolutionary degradation of an existing AT domain as suggested for the NRPS/PKS hybrid OzmQ [43].

In *cis*-AT PKS, the AT domain is flanked by a KS-AT linker domain (LD) and a post-AT linker (PAL). Interestingly, despite the absence of a functional AT domain, these linkers are also well conserved in *trans*-AT PKS. The LD of ClbK shows a βααβα conformation (Fig. S4), lacking the last β-strand that forms the conserved three-stranded β-sheet present in most LDs (Fig. S2 and S4). At the C-terminus, the last helix of the AT* domain and the PAL are very conserved. The PAL is formed by an a-helix that stacks on the LD and a long loop that wraps around the KS domain in order to further connect the next ACP domain (Fig. 2).

The ACP domain is positioned at the bottom entrance of the KS

The crystal structure of ClbK-PKS displays one ACP molecule per dimer (Fig. 2) suggesting that the second ACP is probably not in a unique position. The attribution of one ACP to either dimer chain is not straightforward as no electron density is visible between the C-terminus of PAL (residue 697) and the N-terminus of ACP (residue 705). The distance between these two terminal residues varies from 17.6 Å to 20 Å and both positions are compatible with the fitting of a 9-residue extended linker. The second PKS dimer of the asymmetric unit composed of chains C and D does not show any electron density for residues 697-705 either and the overall positioning of this second ACP is slightly different from the one in the AB dimer, highlighting its flexibility. Indeed, as mentioned previously, the two ACPs do not superimpose completely and are rather rotated 19° and translated 13 Å from each other leading to one ACP domain being closer to the KS (Fig. S1A).

The overall structure of the ACP domain consists of a right-handed helical bundle made of well-defined H1, H2, H3 and H4 α-helices separated by loops (Fig. S5A). Interestingly, we noticed the presence of a H0 helix that has been shown to be much more common in *cis*-AT PKS systems than in *trans*-AT or NRPS/PKS hybrids [47]. The overall structure of the ACP domain is highly similar to the one within MLSB module 7 [47] (rmsd 1.45 Å), DEBS module 2 [48] (rmsd 3.53 Å) and the PCP domain of TycC module 3 [49] (rmsd 1.21 Å) (Fig. S5A, B, C). The conserved Ser747, post-translationally modified with a 20 Å-long PPant arm, is located at the N-terminus of H2. ClbK-PKS was expressed in BAP1 *E. coli* cells that overexpress a phosphopantetheinyl transferase and the presence of the PPant arm was confirmed by mass spectrometry. However, we could not detect electron density for the PPant arm, suggesting that it is flexible.

Remarkably, the sequence alignment of ClbK ACP with different sequences of ACP and PCP domains show that the motif including the catalytic serine at the N-terminus of helix H2 resembles the PCP motif with a GGHS instead of a GXDS in ACPs [48,49] (Fig. S5D). Moreover, the 10 residue-long loop I that connects helices H1 and H2 is rather short compared to other ACP domains and is closer to the length of the loop seen in PCP domains (Fig. S5). Additionally, the DALI structural similarity server displayed mostly PCP domains as structures that were the most closely related to ClbK ACP classifying it as being more similar to a PCP than to an ACP. Given the hybrid nature of ClbK, we hypothesized that the PCP-like features of ClbK ACP could facilitate the transfer of

the hybrid substrate from the upstream KS domain to the cyclization domain of the downstream NRPS module on the same polypeptide. Notably, the ACP domain from HMWP1 and EpoA, that also deliver their substrates to Cy domains, lack these PCP-like features (Fig. S5D).

In ClbK-PKS structure, the ACP is located in an unusual position, i.e., at the bottom of the KS dimer (Fig. 2B and Fig. 4A). Residues from ACP helices H1 and H2 make hydrogen bonds with the abovementioned KS β2 and β3 strands and the interface covers 284 $Å^2$ between the two domains. Although we cannot exclude that the position of ACP is only promoted by the crystallographic packing, several elements suggest its position is relevant. Interestingly, ACP is positioned next to the KS bottom entrance as also observed in the PikAIII structure bound to methylmalonyl (Dutta et al., 2014) (Fig. 4B). In this work, the authors blocked this bottom entrance and noticed a dramatic decrease in PikAIII activity, thus validating the functionality of the entrance. The fitting of ClbK-PKS crystal structure into the cryo-EM map of PikAIII loaded with the MM acceptor substrate reveals that the ClbK ACP molecule fits well in the density that was attributed to the ACP5 domain of PikAIII (Fig. 4C). However, in the ClbK-PKS crystal structure, the conserved ACP serine is located 38 Å from the KS active site, exceeding the 20 Å length of the PPant arm, suggesting that the ACP is not in a catalytically active conformation (Fig. 4A). The unusual location of the ClbK ACP domain could reflect either its trapping before or after the condensation reaction.

ClbK possesses a second dimerization element in addition to the KS domain

The crystal structure of ClbK-PKS reveals a KS-based dimerization architecture similar to KS domains from other modular PKSs. Notably, it has been suggested that the sole KS-based dimerization surface could be too weak to maintain the dimeric state of a whole PKS-NRPS hybrid [37]. To investigate this issue, we determined the oligomeric state of several multi-domain fragments of ClbK by size exclusion chromatography coupled with multi-angle laser light scattering (SEC-MALLS). We first performed SEC-MALLS measurements on the ClbK-PKS module and surprisingly found that it is mostly monomeric in solution with a low proportion of dimers (Fig. S6H). We hypothesized that the absence of dimerizing elements in the N-terminal or C-terminal regions of the KS (Fig. S2) could explain the dissociation of dimeric ClbK-PKS into a monomeric form in solution [50]. We next performed SEC-MALLS on full-length ClbK ($ClbK_{fl}$), which shows a peak of aggregates and a peak corresponding to a dimer (Fig. S6B). This observation suggested that the dimerization of $ClbK_{fl}$ is only partially due to its PKS module and requires the contribution of another dimerization element within the NRPS module. Although NRPSs are usually monomeric, a few examples of dimeric NRPS modules have been described recently with dimerization either through a cyclization-A_{core} interface or through an oxidase domain [23–25]. To investigate the second dimerization element, we designed the construct ClbK-PKS-A_{core}, including residues 1-1659 that incorporate the cyclization, the A_{core} domains and part of the A_{sub} subdomain at the C-terminus of the PKS module (Fig. 1B). ClbK-PKS-A_{core} is also monomeric (Fig. S6F), indicating that the second element facilitating the dimerization of ClbK is most probably located C-terminally to the A_{core} subdomain. Finally, we designed the construct ClbK-PKS-A_{sub} (residues 1-2023) that includes the second half of the A_{sub} subdomain as well as the oxidation domain. The SEC-MALLS measurement on ClbK-PKS-Asub displays a peak with aggregates and a peak corresponding to a dimer (Fig. S6D) suggesting that the second element of dimerization might be the oxidation domain from the NRPS module of ClbK.

Structural characterization of full-length ClbK by SAXS

A full dimodular PKS-NRPS hybrid is yet to be structurally characterized. To fulfill this gap, we performed size exclusion chromatography coupled with small angle X-ray scattering (SAXS) on

$ClbK_{fl}$ on beamline BM29 at the ESRF (Grenoble, France). In order to exclude the contribution from the first peak of aggregate (Fig. S6B), we selected the signal from the second half of the dimeric species using CHROMIXS from the ATSAS suite [51] before subtracting the buffer signal. The Guinier representation of the resulting scattering curve is linear, confirming the absence of inter-particle interactions (Fig. 5A), and generated an R_g of 7.95 ± 0.02 nm. Calculation of the distance distribution function (P(*r*)) using GNOM yielded a maximum particle diameter (D_{max}) of about 30 nm (Fig. 5B). The fine analysis of the P(*r*) curve (Fig. 5B) and the Kratky plot (Fig. 5C) suggest that $ClbK_{fl}$ is a folded modular protein where modules might be connected by flexible linkers (Table 2).

We next calculated *ab initio* three-dimensional envelopes from the experimental scattering data using DAMMIN with the imposition of a prolate shape and P2 symmetry. We generated 20 *ab initio* models that were subsequently aligned, averaged and filtered into eight clusters, before applying a final round of *ab initio* modeling onto each cluster using DAMMIN to yield a distribution of eight models presented in Fig. 5D. All of the models show good correlation with the experimental data ($\chi 2$ = 1.06 ± 0.10) and a mean value of normal spatial discrepancy NSD = 1.05 ± 0.10. The molecular weight estimated from their volumes is 479 kDa, in agreement with the dimeric form of $ClbK_{fl}$. The SAXS envelope possesses an overall volume of about 1,321 nm^3, and the dimeric PKS module with only 231 nm^3 occupies less than 1/5th of the structure. In this perspective, our attempts to model the ClbK-PKS module into the $ClbK_{fl}$ SAXS envelope were unsuccessful. However, we noticed that the overall shape of the models resembles the X shape of the mammalian fatty acid synthase mFAS [34], the iterative PKS LovB-LovC complex from the lovastatin biosynthetic pathway [31] and the H-shaped dimeric NRPS PchE [23]. Additionally, the global arrangement of the dimodular $ClbK_{fl}$. suggests the presence of at least two chambers that could be catalytic. Notably, the different shapes represented in Fig. 5D reflect the conformational changes inherent to this molecule. Overall, this represents an original structural information of a full dimodular hybrid system.

Discussion

Colibactin is a natural product produced by several *Enterobacteriaceae* including *E. coli*, that displays several bioactivities including severe genotoxic and cytotoxic effects that can lead to the development of colorectal cancers. In addition, colibactin has also been reported to show antibacterial, anti-inflammatory and analgesic effects [52]. Recently, its antibacterial activity was linked to microbiome regulation by induction of prophage production, limiting the growth of competing bacteria [53]. These different behaviors could be explained by the diversity and the complexity of colibactin molecules produced by the *pks* BGC [54]. Notably, the colibactin assembly line displays rare biosynthesis mechanisms such as the incorporation of the unusual AM unit and the presence of three non-functional AT domains in ClbC, ClbK and ClbO [55]. The incorporation of AM is a key step in the production of colibactin metabolites; thus, the structural comprehension of one of its incorporating enzymes, ClbK, is crucial to the understanding of this enzymatic process. Here we present the high-resolution structure of the full *trans*-AT PKS module of the ClbK PKS-NRPS hybrid, which contains the three essential domains for chain elongation i.e., the ketosynthase, a truncated acyltransferase and the acyl carrier protein domains. However, we found out that the acyltransferase domain was non-functional due to degenerative truncations, therefore, the non-functionality of the AT* domain raises questions regarding the loading of the extender unit on the ACP domain of ClbK.

An aminomalonyl (AM) synthesizing cluster (ClbD, E, F, G and H proteins) [39] was identified in the *pks* BGC by homology to the proteins involved in the biosynthesis of the aminomalonyl unit in zwittermicin A and the guadinomine biosynthesis pathways [56,57] (Fig. 1). The standalone *trans*-AT ClbG has been hypothesized to complement the degenerative AT* from ClbK and thus deliver the AM extender unit directly to the ClbK-PKS module [40]. The docking position of a *trans*-AT enzyme on its *trans*-AT PKS is still unclear, including the question of whether it can dock onto the post-KS region that would thus act as an AT docking domain [58] or directly onto the ACP domain [15]; however, evidence points to the second hypothesis. In particular, the extensive work on several *trans*-AT PKSs [46] identified the LINKS region in the place of the AT domain, that could be implicated in lateral self-oligomerization of KS domains to form megacomplexes such as the Bacillaene megacomplex in *B. subtilis* [59]. Thus, this region should not be implicated in the recognition of a *trans*-AT domain for these KS. More generally, in KS domains from *trans*-AT PKSs, the diversity in size and sequence of the insertion at the AT position suggest that these regions do not play a docking role for *trans*-AT domains. Additionally, certain *trans*-AT domains have been shown to charge ACPs without the presence of the KS or the LD [60–62]. These results are in favor of a colibactin biosynthesis pathway model where ClbG would dock directly onto the ACP domain of ClbK to deliver its extender AM unit.

Surprisingly, the ClbK-PKS structure displays the ACP in a position close to the KS bottom entrance as seen in the structure of PikAIII loaded with methylmalonyl [28]. However, the ACP catalytic serine is not correctly positioned to deliver a substrate to the KS domain suggesting that the ACP could be in a transitional position before or after KS-catalyzed condensation. The bottom entrance has never been observed in *trans*-AT PKSs [15] and has only been seen once in *cis*-AT PKSs [28]. In some structures, the bottom entrance is occluded by closing loops as seen in DEBS module 5 [32] but in a few others, the region closing the bottom entrance is disordered, making the presence of the bottom entrance difficult to identify. Thus, other structures of KS domains with an open bottom entrance must be obtained to confirm its relevance among PKSs. The side entrance of the ClbK KS domain is also open, suggesting a similar discrimination mechanism as described for the KS domain of PikAIII. We can thus hypothesize that the presence of both entrances could allow for an effective channeling mechanism for substrates to travel through different catalytic tunnels. During colibactin biosynthesis, the side entrance could be devoted to the incoming growing colibactin molecule attached to the PCP domain of ClbJ. This donor substrate could then be transferred to the KS catalytic cysteine before being processed by the KS domain. The bottom entrance could accept the ClbK ACP domain loaded with the AM extender substrate. A high-resolution structure of the PKS module charged with the AM unit is needed to confirm this hypothesis.

The ClbK PKS module is surrounded by two NRPS modules i.e., the upstream NRPS ClbJ module and the downstream NRPS module of ClbK itself, making ClbK a "double-hybrid" enzyme. Hence, it raises the question of particular adaptations at the interface between NRPS and PKS modules. For example, the leinamycin hybrid LnmI possesses a NRPS module followed by a *trans*-AT PKS module with a KS-ACP-KS domain architecture [63]. The first KS lacks one of the two catalytic histidines and is thus condensation-incompetent (KS^0). This KS^0 facilitates a crosstalk mechanism between NRPS and PKS modules by accepting an incoming NRPS substrate from a PCP domain and transferring it to an ACP domain. The ClbK-PKS crystal structure describes a different type of interface between an upstream NRPS module and a downstream PKS module with a direct transfer from the ClbJ-NRPS module to the ClbK-PKS module. We noticed adaptations of the KS domain of ClbK such as its active site cap that folds as a β-strand, thus widely opening the KS bottom entrance and allowing bulky amino acid-containing substrates to leave the KS. At the PKS-NRPS tethered interface, the ClbK ACP domain possesses PCP-like features that could facilitate the

substrate transfer from a PKS module to a NRPS module on the same polypeptide. This ACP specificity could also raise the question of module redefinition as it was suggested recently in *trans*-AT PKSs [64]. According to the Keatinge-Clay module definition, the ACP would belong to the same module as the downstream Cy domain and this definition is consistent with the PCP-like features of ClbK-ACP.

Finally, the oligomeric state of PKS-NRPS hybrid systems is difficult to predict since both partners usually do not show the same oligomeric status. Indeed, most PKSs have been shown to be dimeric [13] and the NRPS enzymes have long been thought to adopt a monomeric behavior [65] until very recently. The PKS-NRPS hybrid HWMP1 forms both monomers and dimers in solution, suggesting that the KS-based dimerization is not strong enough to maintain the dimeric state of the full-length protein [65]. We showed here that the PKS-NRPS $ClbK_{fl}$ hybrid is dimeric both using SEC-MALLS and SAXS experiments. Interestingly, ClbK-PKS is monomeric in solution and dimeric in the crystal structure. The dimerization surface in the PKS module is established through the large KS interface, similarly to classical dimeric PKS systems, but the monomeric behavior in solution suggests that the $ClbK_{fl}$ PKS-NRPS hybrid possesses a second dimerization element in the NRPS module. Very recently, the structures of FmoA3 and PchE [23,24] showed a head-to-tail dimerization through the adenylation-cyclization interface of each monomer, specifically involving the adenylation A_{core} subdomain. However, we ruled out this possible mode of dimerization in ClbK given that the construct ClbK-PKS-A_{core} including the cyclization and the A_{core} domains is a monomer in solution. We thus focused our attention on the oxidase domain that is responsible for the oxidation of thiazolines to form thiazole rings [66], implicated in the bioreactivity of certain molecules such as epothilone [67], bleomycin [68] and indigoidine [69]. In ClbK, the oxidase domain is inserted within the adenylation domain, between the A8 and A9 motifs, forming an interrupted A domain [70]. Several non-NRPS oxidase domains have also been shown to be dimeric, like the MbcC [71] or ThcOx oxidase domains from the cyanobactin pathway [72] that possess 27% and 30% identity with the oxidase domain of ClbK, respectively. In fact, many flavine-cofactor-containing proteins that share high sequence similarity display dimeric structures. Very recently, the structure of the BmdC oxidase domain in complex with the BmdB NRPS module from the bacillamide biosynthetic gene cluster was solved [25]. The oxidase domain is responsible for the dimerization of the whole NRPS module, which has the same domain composition as ClbK-NRPS. We thus performed SEC-MALLS measurements on a ClbK-PKS-A_{sub} construct that harbors the oxidation domain and determined that it is a dimer in solution, lending support to our hypothesis that the second element of dimerization in $ClbK_{fl}$ is most probably the NRPS oxidase domain. Thus, we propose a speculative model for the ClbK PKS-NRPS hybrid where the upstream PKS module is linked to the downstream NRPS module through the dimeric oxidase domains with elongated arms composed of the other NRPS domains. However, the distribution of SAXS *ab initio* shapes reflects the inter-domain flexibility inherent to this molecule and suggests that it must be subjected to vast rearrangements during catalysis.

This structural characterization of a PKS-NRPS hybrid enzyme from the colibactin pathway will pave the way for a better comprehension of the very complex assembly line leading to the production of this genotoxin. In addition, this structural information will be helpful for the biosynthesis of new molecules by biochemical engineering. Indeed, NRPS/PKS enzymes synthesize very diverse metabolites and many efforts have been made these past decades to rationally engineer NRPS/PKS assembly lines. For now, this engineering has focused on pure NRPS or PKS systems; however, the combination of the two, as nature does, could generate highly diverse metabolites with many applications. Future engineering on hybrid megaenzymes could thus benefit from this work.

Acknowledgments

This work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UAR 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-0005-002) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). We thank Aline Le Roy for assistance and access to the Protein Analysis On Line (PAOL) platform and Caroline Mas for access to the SEC-MALLS platform. We thank Sylvain Engilberge for his help running the AlphaFold2 program. We thank Petra Pernot for her help during SAXS data collection at ESRF and Frank Gabel for his help during SAXS data analysis. We also thank Pr. Chaitan Khosla (Stanford University) for his kind gift of BAP1 cells.
Work on NRPS/PKS enzymes in the Bacterial Pathogenesis Laboratory is supported by grants from Le Comité de Savoie de la Ligue Contre le Cancer (M32917, to P.M.) and the FINOVI foundation (AO12-20, to A.D.). S.B. is the recipient of a PhD fellowship from the Ecole Nationale Supérieure (ENS) de Lyon.

Author Contributions

Conceptualization: SB, AD, PM; Methodology: SB, PM; Investigation: SB, PM, CCM; Visualization: SB, PM; Supervision: PM; Writing-original draft: PM; Writing-review & editing: SB, AD, PM

Declaration of interests

Authors declare that they have no competing interests.

Inclusion and Diversity

We support inclusive, diverse, and equitable conduct of research.

Figure titles and legends

Fig. 1. Biosynthesis of colibactin by the *pks* BGC. (A) Scheme of the colibactin biosynthetic mechanism [9]. Non-ribosomal peptide synthetases (NRPSs, in yellow), polyketide synthases (PKSs, in salmon) and hybrid NRPS/PKS cooperate to build colibactin. The aminomalonyl unit incorporated by the ClbK-PKS module is shown is red. As this aminoketone group between the thiazole rings is sensitive to oxidation and hydrolysis, the amino group is not found in the final colibactin molecule. Domain abbreviations: A = adenylation, ACP = acyl carrier protein, AT = acyltransferase, AT* = atypical AT, C = condensation, Cy = cyclization, DH = dehydratase, E = epimerization, ER = enoylreductase, KS = ketosynthase, KR = ketoreductase, Ox = oxidase, PCP = peptidyl carrier protein. **(B)** Domain organization of ClbK. **(C)** Catalytic activities of the PKS module of ClbK. The KS domain first catalyzes the transthioesterification, i.e. the loading of the donor substrate from ClbJ-PCP to a conserved KS cysteine (1). The ACP domain receives an aminomalonyl acceptor substrate from the free-standing AT ClbG (2). The KS domain then catalyzes the decarboxylative condensation between donor and acceptor substrates (3). The condensed product becomes a donor substrate for the Cy domain of ClbK-NRPS (4).

Fig. 2. Overall architecture of ClbK-PKS. (A) Linear domain organization of ClbK-PKS. **(B)** Crystal structure of ClbK-PKS at 2.98 Å resolution with each domain colored uniquely in accordance to (A). The Ks domain of one monomer is colored in blue and the KS from the other

dimer is colored in cyan. **(C)** Structural alignment of ClbK-PKS (colored like in A and B) and DEBS module 5 (in grey, PDB 2HG4). (**D)** Superposition of ClbK-PKS onto the EM density for PikAIII (in grey surface, EMD-5653).

Fig. 3. KS active site accessibility. (A) Monomeric view of the KS domain with the β2 and β3 clasping loop colored in blue and the β8 active site cap colored in cyan, both are boxed at the bottom of the KS. The active site residues are represented with red spheres for Cys167 and green spheres for His303 and His343. **(B)** Rearrangement of the active site cap and the clasping loop. The clasping loop is shown in blue for ClbK and grey for the other structures and is composed of β2 and β3 in ClbK. It is usually unstructured except for ClbK, TAS1 and OzmQ. The active site cap is colored in cyan in ClbK and in wheat in the other structures. It folds as an helix in *cis*-AT PKS and *trans*-AT PKS BaeKS5. In hybrid enzymes (ClbK, OzmQ and TAS1) it folds as a beta strand (β8 in ClbK). **(C)** Representation of the tunnel cavities of the KS domain of ClbK. The cavity is seen on the side (up) and on the bottom (bottom) of the KS with the catalytic Cys167 colored in red. β8 of the active site cap is colored in cyan **(D)** Representation of the tunnel cavities of DEBS module 5 (PDB 2HG4). The bottom entrance is closed by the helix from the active site cap, colored in wheat.

Fig. 4. The ACP domain is located at the bottom entrance of the KS domain. (A) The representation of the KS cavities highlights the ACP position close to the bottom entrance of KS. The distance between the KS Cys167 and ACP Ser747 residues (in red) is however too long (38 Å) to be in a catalytically active position. **(B)** Representation of the EM envelope of PikAIII loaded with methylmalonyl (MM-PikAIII) (EMD-5653) with the fitted DEBS domains highlighting the positioning of the KS5, AT5, ACP5 and KR5 domains as observed in [28]. **(C)** Superposition of ClbK-PKS onto the same MM-PikAIII EM envelope showing the similarity of positioning of ClbK ACP and PikAIII ACP5.

Fig. 5. SEC-SAXS analyses of $ClbK_{fl}$ and ab initio model generated from SAXS data. (A) The Guinier plot is linear, indicating the absence of inter-particle interactions. **(B)** The pair distance distribution function, P(*r*), reveals $ClbK_{fl}$ has a maximum particle diameter (Dmax) around 30 nm. **(C)** The Kratky plot indicates the protein is folded. **(D)** 20 *ab initio* models were generated by DAMMIN with an imposed P2 symmetry. These models were aligned, averaged and filtered before a final step of refinement with DAMMIN, yielding the distribution of 8 clusters for which their representative models are presented here. The refined models are displayed with Pymol.

Table 1. Data collection and refinement statistics. Statistics for the highest-resolution shell are shown in parentheses.

Table 2. SAXS data collection and analysis.

Table 1

DATA COLLECTION ClbK-PKS (PDB: 8CJH)	
Resolution range (Å)	48.98 – 2.98 (3.08 – 2.98)
Space group	C2221
Unit cell parameters (Å)	187.64 233.37 256.59 90 90 90
Total reflections	370039 (28856)
Unique reflections	78289 (5592)
Multiplicity	4.7 (5.2)
Completeness (spherical) (%)	68.5 (15.8)
Completeness (ellipsoidal) (%)	93.6 (72.7)
Mean I/sigma(I)	9.5 (1.6)
Wilson B-factor (Å^2)	121.67
R-merge	0.138 (1.050)
CC1/2	0.996 (0.522)
REFINEMENT	
R-work	0.217 (0.355)
R-free	0.238 (0.202)
Protein residues	2892
rms bond lengths (Å)	0.008
rms angles (°)	1.28
Residues in most favored/allowed regions of the Ramachandran plot (%)	99.76
Average B-factor (Å^2)	96.71

Statistics for the highest-resolution shell are shown in parentheses.

Table 2

SAXS data-collection parameters	
Instrument	BM29 beamline at ESRF, Grenoble, France with Pilatus3 2M detector in vacuum
Wavelength	0.992 Å
Beam size (μm)	200 x 200
Camera length (m)	2.827
q measurement range (Å^{-1})	0.0025 – 0.6
Exposure time	Continuous 1 s data-frame measurements of SEC solution
Sample configuration	SEC-SAXS
Sample temperature	20°C
Software employed for SAXS data reduction, analysis and interpretation	

SAXS data reduction	chromixs
Guinier analysis	Primus
Shape/bead modelling	DAMMIN
Atomic structure modelling	dadimodo
Three-dimensional graphic model representations	Pymol
Structural parameters	
Guinier analysis	
I(0) (cm^{-1})	217.1 +/- 0.6
R_g (nm)	7.95 +/- 0.03
P(r) analysis	
I(0) (nm^{-1})	219.0 +/- 0.4
R_g (nm)	8.2 +/- 0.02
d_{max} (nm)	30.9
q range (nm^{-1})	0.04 – 1.78
χ^2 (total estimate from GNOM)	0.8297
Porod volume (nm^{-3}) (ratio Vp/calculated M)	1341.55
Shape model-fitting results	
DAMMIN	
q range for fitting (nm^{-1})	0.04 – 1.78
Symmetry, anisotropy assumptions	P2
$\chi^{2,}$ CORMAP P-values	1.01, 0.14

STAR Methods

RESOURCE AVAILABILITY

Lead Contact

Further information and requests for resources and reagents should be directed to the lead contact, Pauline Macheboeuf (pauline.macheboeuf@ibs.fr)

Materials Availability

The plasmids generated in this study are available from the lead contact upon request.

Data and Code Availability

- All X-ray diffraction data and small angle X-ray data have been deposited in the PDB and SASBDB respectively and are publicly available as of the date of publication. Accession numbers are listed in the key resources table.
- 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.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

In this study BL21 (DE3) Bap1 competent *E. coli* cells (gift from Pr. Khosla) were used. These were grown either in LB medium or auto-induction medium (see method details section) at 37°C until mid-log phase and 18°C overnight for protein expression.

METHOD DETAILS

Cloning of ClbK constructs

All genes were PCR amplified from pBAC-pks, a kind gift from JP Nougayrède, and further cloned using the InFusion cloning kit (Takara Bio); cloning success was confirmed by sequencing. $ClbK_{fl}$ (UNIPROT Q0P7K1) was cloned into a pET15b vector that contains a N-terminal thrombin cleavable hexa-histidine tag using *$clbK_{fl}$_NdeI_for* and *$clbK_{fl}$_BamHI_rev* primers

Multi-domain fragments of ClbK were PCR amplified from the pET15b-ClbK plasmid and recombined in the pETM11 plasmid that contains a N-terminal hexa-histidine tag and a Tev cleavage site using *clbK-fragments_NcoI_for* primer as the forward primer and a specific primer as the reverse primer *(clbK-PKS_XhoI_rev* and *ClbK-PKS-A_{core}_XhoI_rev* for the ClbK-PKS and ClbK-PKS-A_{core} fragments respectively).

Expression of $ClbK_{fl}$, ClbK-PKS-A_{core}, ClbK-PKS-A_{sub} and ClbK-PKS

All proteins were expressed in *Escherichia coli* BL21(DE3) Bap1 cells, a kind gift of Pr. Khosla. Expression was carried out in LB medium (Lennox) for $ClbK_{fl}$ and ClbK-PKS-A_{sub} whereas ClbK-PKS (residues 1-787) and ClbK-PKS-A_{core} (residues 1-1659) were expressed using auto-induction medium (For 0.5 L, 10 g tryptone, 5 g yeast extract, 1 mM MgSO4, 0.1 M $(NH_4)_2SO_4$, 0.2 M KH_2PO_4, 0.2 M Na_2HPO_4, 0.5 % glycerol, 0.05 % glucose, 0.2 % α-lactose). In both cases, cells were grown at 37°C until $OD_{600} \approx 0.5$ and then cooled down to 18°C. In LB medium, after two hours of temperature equilibration, proteins were induced with 0.2 mM isopropyl-β-D-

thiogalactopyranoside (IPTG) at 18°C for 16-20 hours whereas in auto-induction medium, expression was carried out at 18°C for 20-24 hours.

Protein purification

$ClbK_{fl}$ expressing cells were harvested by centrifugation and resuspended in buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 5 mM dithiothreitol (DTT)). The lysis buffer was supplemented with benzonase (Sigma) and protease inhibitors (Complete EDTA-free (Roche)) to prevent protein degradation. Cells were lysed using a Constant Systems Cell disrupter (4x20-25 kpsi) and the supernatant from the centrifugation step was incubated with Protino-Ni-IDA (Macherey-Nagel) resin. The protein was eluted using the appropriate lysis buffer supplemented with 250 mM imidazole and 1 mM EDTA was directly added to the $ClbK_{fl}$ elution fractions. Eluted $ClbK_{fl}$ was dialyzed against (50 mM HEPES pH 7.5, 100 mM NaCl, 5 mM DTT, 1 mM EDTA) and further purified using an anion exchange chromatography step using a monoQ 5/50 GL column (GE Healthcare). The protein was eluted using a linear salt gradient. Finally, the protein was submitted to a polishing gel filtration step using a Superdex 200 Increase 10/300 GL column (GE Healthcare) in Buffer B (50 mM HEPES pH 7.5, 200 mM NaCl, 5 mM DTT).

ClbK-PKS-A_{core} and ClbK-PKS-A_{sub} were purified the same way as $ClbK_{fl}$ using the lysis buffer C (50 mM HEPES pH 7.5, 300 mM NaCl, 5 mM DTT) supplemented with benzonase and protease inhibitors. The buffer D (25 mM HEPES pH 7.5, 150 mM NaCl, 5 mM DTT) was used for the final gel filtration step.

ClbK-PKS protein was lysed using buffer C supplemented with benzonase and protease inhibitors and purified by a two-step purification including the Protino-Ni-IDA resin for affinity chromatography and a gel filtration step using buffer D. This protein was used for further SEC-MALLS experiments.

For X-ray crystallography, the hexa-histidine tag of ClbK-PKS was cleaved after affinity chromatography by incubation with TEV protease (1:100) during a dialysis step against buffer C and further eluted using a reverse affinity chromatography. The protein was submitted to a final gel filtration step in buffer E (25 mM HEPES pH 7.5, 50 mM NaCl, 5 mM DTT). This protein was used for crystallization and was subjected to electrospray mass spectrometry to further confirm the cleavage of the tag and the presence of the PPant arm.

Crystallography

ClbK-PKS crystals grew within a few days in 0.1 M MES pH 6.5, 19 % PEG 3350 at 20°C using the hanging-drop setup of the high throughput crystallography facility from the Integrated Structural Biology Grenoble (ISBG). They were soaked in a cryo-protectant solution (0.1 M MES pH 6.5, 18 % PEG 3350, 25 % glycerol), mounted in a cryo-loop and flash-cooled in liquid nitrogen. Diffraction data were collected at the ID30A-1 beamline at the European Synchrotron Radiation Facility (ESRF, Grenoble, France).

X-ray diffraction images were indexed and scaled using XDS [74] and the resolution limit was determined using STARANISO [81]. The structure was solved by molecular replacement with Phaser [75] using a truncated model of monomeric ClbK-PKS generated by AlphaFold2 [42] using an in-house computer cluster. The structure was completed by cycles of manual model building with COOT [76], phase optimization with PARROT [77] and automatic model-building with BUCCANEER [78] from the CCP4 program suite [79]. Several cycles of manual model building and macromolecular refinement were performed with REFMAC [80] to a R_{work} of 22.9% and a R_{free} of 26.3%. The final structure

contains two dimers (A, B &C, D) of PKS that contain the KS, AT*, LD, PAL domains. Because we could not confidently connect the two ACP domains to any of the four protein chains, we chose to provide these with unique protein chains (E&F). Figures containing protein structures were generated with PYMOL (http://www.pymol.org). Final refined model coordinates and structure factors were deposited at the Protein Data Bank (PDB, http://www.rcsb.org), ID code: 8CJH.

Size-exclusion chromatography coupled to small angle X-ray scattering (SEC-SAXS)
Data were collected at BM29 beamline (ESRF, Grenoble, France), at a wavelength λ=0.992 Å and a Pilatus2M detector was placed 2.82 m away from samples. The momentum transfer s was defined by $s= 4\pi \sin(\theta) /\lambda$. $ClbK_{fl}$ was injected at a flow rate of 0.75 mL/min on a Superdex 200 Increase 10/300 GL column equilibrated in buffer B, at 20°C. Analysis was performed using the ATSAS software package version 3.0.3 [51]. First, CHROMIXS was used for selection of sample and buffer regions, buffer subtraction and data averaging. Guinier and Kratky plots were displayed using PRIMUS. The pair distance distribution function was evaluated from scattering data using GNOM. 20 *ab initio* models were generated with DAMMIN (slow mode, P2 imposed symmetry). Following model comparison and averaging with DAMAVER program suite, a final step was performed with DAMMIN to yield a distribution of eight refined models that fit the experimental data. Volumes of the individual DAMMIN models and estimated molar mass were extracted directly from the PDB files. The small angle data has been deposited in the SASBDB data bank under the code: SASDQ87.

Size-exclusion chromatography coupled to Multi-Angle Laser Light Scattering (SEC-MALLS)
SEC-MALLS analyses were performed using a system composed of a Wyatt DAWN Heleos II MALS, a Wyatt Optilab T-rex refractometer, an Hitachi Elite LaChrom L-2400 UV detector and an Hitachi Elite LaChrom Pump L-2130. A Superdex 200 Increase 10/300 GL column was equilibrated with several column volumes of buffer before injection of concentrated samples at a flow rate of 0.5 mL/min. Before each experiment, bovine serum albumin was injected to confirm the correct calibration of the system. Molar masses were estimated from scattering data using Astra 6 software (Wyatt).

QUANTIFICATION AND STATISTICAL ANALYSIS

Data collection and refinement statistics for the structures described in this work are available in the key resources table section and in Table 2.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Bacterial and virus strains		
Escherichia coli BL21 Bap1 (DE3)	[73]	N/A
Chemicals, peptides, and recombinant proteins		
$ClbK_{fl}$	This study	N/A
ClbK-PKS-A_{core}	This study	N/A
ClbK-PKS-A_{sub}	This study	N/A
ClbK-PKS	This study	N/A
LB Broth (Lennox)	AthenaES	Cat#0102
tryptone	US Biological	Cat#T8750
Yeast extract	Euromedex	Cat#UY2010
Magnesium sulfate	Sigma	Cat#M7506

Potassium phosphate	Sigma	Cat#P0662
Sodium phosphate	Sigma	Cat#S8282
glycerol	VWR	Cat#24388.295
glucose	Sigma	Cat#G7021
α-lactose	Sigma	Cat#L2643
IPTG	Neo Biotech	Cat#NB-45-00030
HEPES	Euromedex	Cat#10-110-C
NaCl	Euromedex	Cat#1112-A
DTT	Euromedex	Cat#EU0006-D
MES	Sigma	Cat#M5057
PEG 3350	Sigma	Cat#88276
imidazole	Sigma	Cat#56750
cOmplete EDTA-free protease inhibitor	Roche	Cat#56619600
Benzonase	Sigma	Cat#E1014-5KU
Protino-Ni-IDA resin	Macherey-Nagel	Cat#745210
TEV	Home-purified clone	N/A
InFusion cloning kit	Takara	Cat#638920
Deposited data		
ClbK-PKS	This study	PDB: 8CJH
SAXS data on $ClbK_{fl}$	This study	SASBDB: SASDQ87
Oligonucleotides		
clbKfl_NdeI_for CGCGCGGCAGCCATATGACTTACAGTGAAAGCGATATTGC	This study	N/A
$clbK_{fl}$_BamHI_rev GTTAGCAGCCGGATCCTCACTCATCGCCCTTTTCTCCC	This study	N/A
clbK-fragments_NcoI_for TTCAGGGCGCCATGGGTACTTACAGTGAAAGCGATATTGCCA	This study	N/A
clbK-PKS_XhoI_rev GTGGTGGTGCTCGAGTCAGCCTTGTTCGAGCAGATAGC	This study	N/A
clbK-PKS-A_{core}_XhoI_rev GTGGTGGTGCTCGAGTCAATCGTTGCGCCCCAAAAAG	This study	N/A
Recombinant DNA		
pBAC-pks	Nougayrède et al., (2006)	N/A
pET15b-$ClbK_{fl}$	This study	N/A
pETM11-ClbK-PKS-A_{core}	This study	N/A
pETM30-ClbK-PKS-A_{sub}	This study	N/A
pETM11-ClbK-PKS	This study	N/A
Software and algorithms		
XDS	[74]	https://xds.mr.mpg.de/html_doc/XDS.html
Phaser	[75]	https://www/ccp4.ac.uk

AlphaFold2	[42]	https://www.deepmind.com/research/highlighted-research/alphafold
Coot	[76]	https://www/ccp4.ac.uk
PARROT	[77]	https://www/ccp4.ac.uk
BUCCANEER	[78]	https://www/ccp4.ac.uk
CCP4	[79]	https://www/ccp4.ac.uk
REFMAC	[80]	https://www/ccp4.ac.uk
PYMOL		http://www.pymol.org
ATSAS suite	[51]	https://www.embl-hamburg.de/biosaxs/software.html
Astra 6	Wyatt	https://www.wyatt.com/products/software/astra.html
Other		
MonoQ 5/50 GL	GE Healthcare	Cat#17-5166-01
Superdex 200 Increase 10/300 GL	GE Healthcare	Cat#28-9909-44

References

1. Nougayrède, 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. 10.1126/science.1127059.

2. Cuevas-Ramos, G., Petit, C.R., Marcq, I., Boury, M., Oswald, E., and Nougayrède, J.-P. (2010). Escherichia coli induces DNA damage in vivo and triggers genomic instability in mammalian cells. Proc Natl Acad Sci U S A *107*, 11537–11542. 10.1073/pnas.1001261107.

3. 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. 10.1038/s41586-020-2080-8.

4. 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., et al. (2020). Colibactin DNA-damage signature indicates mutational impact in colorectal cancer. Nat Med *26*, 1063–1069. 10.1038/s41591-020-0908-2.

5. Shrestha, E., Coulter, J.B., Guzman, W., Ozbek, B., Hess, M.M., Mummert, L., Ernst, S.E., Maynard, J.P., Meeker, A.K., Heaphy, C.M., et al. (2021). Oncogenic gene fusions in nonneoplastic precursors as evidence that bacterial infection can initiate prostate cancer. Proc Natl Acad Sci U S A *118*, e2018976118. 10.1073/pnas.2018976118.

6. Chagneau, C.V., Massip, C., Bossuet-Greif, N., Fremez, C., Motta, J.-P., Shima, A., Besson, C., Le Faouder, P., Cénac, N., Roth, M.-P., et al. (2021). Uropathogenic E. coli induces DNA damage in the bladder. PLoS Pathog *17*, e1009310. 10.1371/journal.ppat.1009310.

7. 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. 10.1126/science.aar7785.

8. Roy, R.S., Gehring, A.M., Milne, J.C., Belshaw, P.J., and Walsh, C.T. (1999). Thiazole and oxazole peptides: biosynthesis and molecular machinery. Nat Prod Rep *16*, 249–263. 10.1039/a806930a.

9. 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. 10.1126/science.aax2685.

10. Bonhomme, S., Dessen, A., and Macheboeuf, P. (2021). The inherent flexibility of type I non-ribosomal peptide synthetase multienzymes drives their catalytic activities. Open Biol *11*, 200386. 10.1098/rsob.200386.

11. Bloudoff, K., Fage, C.D., Marahiel, M.A., and Schmeing, T.M. (2017). Structural and mutational analysis of the nonribosomal peptide synthetase heterocyclization domain provides insight into catalysis. Proc Natl Acad Sci U S A *114*, 95–100. 10.1073/pnas.1614191114.

12. Walsh, C.T., Chen, H., Keating, T.A., Hubbard, B.K., Losey, H.C., Luo, L., Marshall, C.G., Miller, D.A., and Patel, H.M. (2001). Tailoring enzymes that modify nonribosomal peptides during and after chain elongation on NRPS assembly lines. Curr Opin Chem Biol *5*, 525–534. 10.1016/s1367-5931(00)00235-0.

13. Keatinge-Clay, A.T. (2012). The structures of type I polyketide synthases. Nat Prod Rep *29*, 1050–1073. 10.1039/c2np20019h.

14. Dodge, G.J., Maloney, F.P., and Smith, J.L. (2018). Protein-protein interactions in "cis-AT" polyketide synthases. Nat Prod Rep *35*, 1082–1096. 10.1039/c8np00058a.

15. Kosol, S., Jenner, M., Lewandowski, J.R., and Challis, G.L. (2018). Protein-protein interactions in trans-AT polyketide synthases. Nat Prod Rep *35*, 1097–1109. 10.1039/c8np00066b.

16. Drake, E.J., Miller, B.R., Shi, C., Tarrasch, J.T., Sundlov, J.A., Allen, C.L., Skiniotis, G., Aldrich, C.C., and Gulick, A.M. (2016). Structures of Two Distinct Conformations of holo-Nonribosomal Peptide Synthetases. Nature *529*, 235–238. 10.1038/nature16163.

17. Kreitler, D.F., Gemmell, E.M., Schaffer, J.E., Wencewicz, T.A., and Gulick, A.M. (2019). The structural basis of N-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. Nat Commun *10*, 3432. 10.1038/s41467-019-11383-7.

18. Miller, B.R., Drake, E.J., Shi, C., Aldrich, C.C., and Gulick, A.M. (2016). Structures of a Nonribosomal Peptide Synthetase Module Bound to MbtH-like Proteins Support a Highly Dynamic Domain Architecture. J Biol Chem *291*, 22559–22571. 10.1074/jbc.M116.746297.

19. Reimer, J.M., Aloise, M.N., Harrison, P.M., and Martin Schmeing, T. (2016). Synthetic cycle of the initiation module of a formylating nonribosomal peptide synthetase. Nature *529*, 239–242. 10.1038/nature16503.

20. Reimer, J.M., Eivaskhani, M., Harb, I., Guarné, A., Weigt, M., and Schmeing, T.M. (2019). Structures of a dimodular nonribosomal peptide synthetase reveal conformational flexibility. Science *366*. 10.1126/science.aaw4388.

21. Tarry, M.J., Haque, A.S., Bui, K.H., and Schmeing, T.M. (2017). X-Ray Crystallography and Electron Microscopy of Cross- and Multi-Module Nonribosomal Peptide Synthetase Proteins Reveal a Flexible Architecture. Structure *25*, 783-793.e4. 10.1016/j.str.2017.03.014.

22. Gulick, A.M., and Aldrich, C.C. (2018). Trapping interactions between catalytic domains and carrier proteins of modular biosynthetic enzymes with chemical probes. Nat Prod Rep *35*, 1156–1184. 10.1039/c8np00044a.

23. Wang, J., Li, D., Chen, L., Cao, W., Kong, L., Zhang, W., Croll, T., Deng, Z., Liang, J., and Wang, Z. (2022). Catalytic trajectory of a dimeric nonribosomal peptide synthetase subunit with an inserted epimerase domain. Nat Commun *13*, 592. 10.1038/s41467-022-28284-x.

24. Katsuyama, Y., Sone, K., Harada, A., Kawai, S., Urano, N., Adachi, N., Moriya, T., Kawasaki, M., Shin-Ya, K., Senda, T., et al. (2021). Structural and Functional Analyses of the Tridomain-Nonribosomal Peptide Synthetase FmoA3 for 4-Methyloxazoline Ring Formation. Angew Chem Int Ed Engl *60*, 14554–14562. 10.1002/anie.202102760.

25. Fortinez, C.M., Bloudoff, K., Harrigan, C., Sharon, I., Strauss, M., and Schmeing, T.M. (2022). Structures and function of a tailoring oxidase in complex with a nonribosomal peptide synthetase module. Nat Commun *13*, 548. 10.1038/s41467-022-28221-y.

26. 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. 10.1126/science.abi8532.

27. Cogan, D.P., Zhang, K., Li, X., Li, S., Pintilie, G.D., Roh, S.-H., Craik, C.S., Chiu, W., and Khosla, C. (2021). Mapping the catalytic conformations of an assembly-line polyketide synthase module. Science *374*, 729–734. 10.1126/science.abi8358.

28. Dutta, S., Whicher, J.R., Hansen, D.A., Hale, W.A., Chemler, J.A., Congdon, G.R., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., et al. (2014). Structure of a modular polyketide synthase. Nature *510*, 512–517. 10.1038/nature13423.

29. Whicher, J.R., Dutta, S., Hansen, D.A., Hale, W.A., Chemler, J.A., Dosey, A.M., Narayan, A.R.H., Håkansson, K., Sherman, D.H., Smith, J.L., et al. (2014). Structural rearrangements of a polyketide synthase module during its catalytic cycle. Nature *510*, 560–564. 10.1038/nature13409.

30. Herbst, D.A., Huitt-Roehl, C.R., Jakob, R.P., Kravetz, J.M., Storm, P.A., Alley, J.R., Townsend, C.A., and Maier, T. (2018). The structural organization of substrate loading in iterative polyketide synthases. Nature Chemical Biology *14*, 474. 10.1038/s41589-018-0026-3.

31. Wang, J., Liang, J., Chen, L., Zhang, W., Kong, L., Peng, C., Su, C., Tang, Y., Deng, Z., and Wang, Z. (2021). Structural basis for the biosynthesis of lovastatin. Nat Commun *12*, 867. 10.1038/s41467-021-21174-8.

32. 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. PNAS *103*, 11124–11129. 10.1073/pnas.0601924103.

33. Tang, Y., Chen, A.Y., Kim, C.-Y., Cane, D.E., and Khosla, C. (2007). Structural and mechanistic analysis of protein interactions in module 3 of the 6-deoxyerythronolide B synthase. Chem Biol *14*, 931–943. 10.1016/j.chembiol.2007.07.012.

34. Maier, T., Leibundgut, M., and Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. Science *321*, 1315–1322. 10.1126/science.1161269.

35. Bretschneider, T., Heim, J.B., Heine, D., Winkler, R., Busch, B., Kusebauch, B., Stehle, T., Zocher, G., and Hertweck, C. (2013). Vinylogous chain branching catalysed by a dedicated polyketide synthase module. Nature *502*, 124–128. 10.1038/nature12588.

36. Davison, J., Dorival, J., Rabeharindranto, H., Mazon, H., Chagot, B., Gruez, A., and Weissman, K.J. (2014). Insights into the function of trans-acyl transferase polyketide synthases from the SAXS structure of a complete module. Chem. Sci. *5*, 3081–3095. 10.1039/C3SC53511H.

37. Miyanaga, A., Kudo, F., and Eguchi, T. (2018). Protein–protein interactions in polyketide synthase–nonribosomal peptide synthetase hybrid assembly lines. Nat. Prod. Rep. *35*, 1185–1209. 10.1039/C8NP00022K.

38. Yun, C.-S., Nishimoto, K., Motoyama, T., Shimizu, T., Hino, T., Dohmae, N., Nagano, S., and Osada, H. (2020). Unique features of the ketosynthase domain in a nonribosomal peptide synthetase-polyketide synthase hybrid enzyme, tenuazonic acid synthetase 1. J Biol Chem *295*, 11602–11612. 10.1074/jbc.RA120.013105.

39. Brachmann, A.O., Garcie, C., Wu, V., Martin, P., Ueoka, R., Oswald, E., and Piel, J. (2015). Colibactin biosynthesis and biological activity depend on the rare aminomalonyl polyketide precursor. Chem Commun (Camb) *51*, 13138–13141. 10.1039/c5cc02718g.

40. 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. 10.1021/acschembio.6b00014.

41. Trautman, E.P., Healy, A.R., Shine, E.E., Herzon, S.B., and Crawford, J.M. (2017). Domain-Targeted Metabolomics Delineates the Heterocycle Assembly Steps of Colibactin Biosynthesis. J Am Chem Soc *139*, 4195–4201. 10.1021/jacs.7b00659.

42. 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. 10.1038/s41586-021-03819-2.

43. 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 U S A *112*, 12693–12698. 10.1073/pnas.1515460112.

44. Tsai, S.-C.S., and Ames, B.D. (2009). Structural enzymology of polyketide synthases. Methods Enzymol *459*, 17–47. 10.1016/S0076-6879(09)04602-3.

45. Kohlhaas, C., Jenner, M., Kampa, A., Briggs, G.S., Afonso, J.P., Piel, J., and Oldham, N.J. (2013). Amino acid-accepting ketosynthase domain from a trans-AT polyketide synthase exhibits high selectivity for predicted intermediate. Chem. Sci. *4*, 3212–3217. 10.1039/C3SC50540E.

46. Gay, D.C., Wagner, D.T., Meinke, J.L., Zogzas, C.E., Gay, G.R., and Keatinge-Clay, A.T. (2016). The LINKS motif zippers trans-acyltransferase polyketide synthase assembly lines into a biosynthetic megacomplex. Journal of Structural Biology *193*, 196–205. 10.1016/j.jsb.2015.12.011.

47. Moretto, L., Heylen, R., Holroyd, N., Vance, S., and Broadhurst, R.W. (2019). Modular type I polyketide synthase acyl carrier protein domains share a common N-terminally extended fold. Sci Rep *9*, 2325. 10.1038/s41598-019-38747-9.

48. Alekseyev, V.Y., Liu, C.W., Cane, D.E., Puglisi, J.D., and Khosla, C. (2007). Solution structure and proposed domain–domain recognition interface of an acyl carrier protein domain from a modular polyketide synthase. Protein Sci *16*, 2093–2107. 10.1110/ps.073011407.

49. Weber, T., Baumgartner, R., Renner, C., Marahiel, M.A., and Holak, T.A. (2000). Solution structure of PCP, a prototype for the peptidyl carrier domains of modular peptide synthetases. Structure *8*, 407–418. 10.1016/s0969-2126(00)00120-9.

50. 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. Chem Biol *10*, 723–731. 10.1016/s1074-5521(03)00156-x.

51. Manalastas-Cantos, K., Konarev, P.V., Hajizadeh, N.R., Kikhney, A.G., Petoukhov, M.V., Molodenskiy, D.S., Panjkovich, A., Mertens, H.D.T., Gruzinov, A., Borges, C., et al. (2021). ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. J Appl Crystallogr *54*, 343–355. 10.1107/S1600576720013412.

52. Faïs, T., Delmas, J., Barnich, N., Bonnet, R., and Dalmasso, G. (2018). Colibactin: More Than a New Bacterial Toxin. Toxins (Basel) *10*. 10.3390/toxins10040151.

53. 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. 10.1038/s41586-022-04444-3.

54. Tang, J.-W., Liu, X., Ye, W., Li, Z.-R., and Qian, P.-Y. (2022). Biosynthesis and bioactivities of microbial genotoxin colibactins. Nat Prod Rep. 10.1039/d1np00050k.

55. Vizcaino, M.I., Engel, P., Trautman, E., and Crawford, J.M. (2014). Comparative Metabolomics and Structural Characterizations Illuminate Colibactin Pathway-Dependent Small Molecules. J. Am. Chem. Soc. *136*, 9244–9247. 10.1021/ja503450q.

56. Chan, Y.A., and Thomas, M.G. (2010). Recognition of (2S)-aminomalonyl-acyl carrier protein (ACP) and (2R)-hydroxymalonyl-ACP by acyltransferases in zwittermicin A biosynthesis. Biochemistry *49*, 3667–3677. 10.1021/bi100141n.

57. Holmes, T.C., May, A.E., Zaleta-Rivera, K., Ruby, J.G., Skewes-Cox, P., Fischbach, M.A., DeRisi, J.L., Iwatsuki, M., Ōmura, S., and Khosla, C. (2012). Molecular insights into the biosynthesis of guadinomine: a type III secretion system inhibitor. J Am Chem Soc *134*, 17797–17806. 10.1021/ja308622d.

58. Tang, G.-L., Cheng, Y.-Q., and Shen, B. (2004). Leinamycin biosynthesis revealing unprecedented architectural complexity for a hybrid polyketide synthase and nonribosomal peptide synthetase. Chem Biol *11*, 33–45. 10.1016/j.chembiol.2003.12.014.

59. Straight, P.D., Fischbach, M.A., Walsh, C.T., Rudner, D.Z., and Kolter, R. (2007). A singular enzymatic megacomplex from Bacillus subtilis. Proc Natl Acad Sci U S A *104*, 305–310. 10.1073/pnas.0609073103.

60. Aron, Z.D., Fortin, P.D., Calderone, C.T., and Walsh, C.T. (2007). FenF: Servicing the Mycosubtilin Synthetase Assembly Line in trans. ChemBioChem *8*, 613–616. 10.1002/cbic.200600575.

61. Ye, Z., Musiol, E.M., Weber, T., and Williams, G.J. (2014). Reprogramming Acyl Carrier Protein Interactions of an Acyl-CoA Promiscuous trans-Acyltransferase. Chemistry & Biology *21*, 636–646. 10.1016/j.chembiol.2014.02.019.

62. Musiol, E.M., Härtner, T., Kulik, A., Moldenhauer, J., Piel, J., Wohlleben, W., and Weber, T. (2011). Supramolecular Templating in Kirromycin Biosynthesis: The Acyltransferase KirCII Loads Ethylmalonyl-CoA Extender onto a Specific ACP of the trans-AT PKS. Chemistry & Biology *18*, 438–444. 10.1016/j.chembiol.2011.02.007.

63. Huang, Y., Tang, G.-L., Pan, G., Chang, C.-Y., and Shen, B. (2016). Characterization of the Ketosynthase and Acyl Carrier Protein Domains at the LnmI Nonribosomal Peptide Synthetase-Polyketide Synthase Interface for Leinamycin Biosynthesis. Org Lett *18*, 4288–4291. 10.1021/acs.orglett.6b02033.

64. Vander Wood, D.A., and Keatinge-Clay, A.T. (2018). The modules of trans-acyltransferase assembly lines redefined with a central acyl carrier protein. Proteins *86*, 664–675. 10.1002/prot.25493.

65. Sieber, S.A., Linne, U., Hillson, N.J., Roche, E., Walsh, C.T., and Marahiel, M.A. (2002). Evidence for a monomeric structure of nonribosomal Peptide synthetases. Chem Biol *9*, 997–1008. 10.1016/s1074-5521(02)00214-4.

66. Schneider, T.L., Shen, B., and Walsh, C.T. (2003). Oxidase Domains in Epothilone and Bleomycin Biosynthesis: Thiazoline to Thiazole Oxidation during Chain Elongation. Biochemistry *42*, 9722–9730. 10.1021/bi034792w.

67. Nicolaou, K.C., Roschangar, F., and Vourloumis, D. (1998). Chemical Biology of Epothilones. Angew Chem Int Ed Engl *37*, 2014–2045. 10.1002/(SICI)1521-3773(19980817)37:15<2014::AID-ANIE2014>3.0.CO;2-2.

68. Bailly, C., Colson, P., Houssier, C., Houssin, R., Mrani, D., Gosselin, G., Imbach, J.L., Waring, M.J., Lown, J.W., and Hénichart, J.P. (1992). Binding properties and DNA sequence-specific recognition of two bithiazole-linked netropsin hybrid molecules. Biochemistry *31*, 8349–8362. 10.1021/bi00150a032.

69. Pang, B., Chen, Y., Gan, F., Yan, C., Jin, L., Gin, J.W., Petzold, C.J., and Keasling, J.D. (2020). Investigation of Indigoidine Synthetase Reveals a Conserved Active-Site Base Residue of Nonribosomal Peptide Synthetase Oxidases. J Am Chem Soc *142*, 10931–10935. 10.1021/jacs.0c04328.

70. Labby, K.J., Watsula, S.G., and Garneau-Tsodikova, S. (2015). Interrupted adenylation domains: unique bifunctional enzymes involved in nonribosomal peptide biosynthesis. Nat Prod Rep *32*, 641–653. 10.1039/c4np00120f.

71. Ghilarov, D., Stevenson, C.E.M., Travin, D.Y., Piskunova, J., Serebryakova, M., Maxwell, A., Lawson, D.M., and Severinov, K. (2019). Architecture of Microcin B17 Synthetase: An Octameric Protein Complex Converting a Ribosomally Synthesized Peptide into a DNA Gyrase Poison. Mol Cell *73*, 749-762.e5. 10.1016/j.molcel.2018.11.032.

72. Bent, A.F., Mann, G., Houssen, W.E., Mykhaylyk, V., Duman, R., Thomas, L., Jaspars, M., Wagner, A., and Naismith, J.H. (2016). Structure of the cyanobactin oxidase ThcOx from Cyanothece sp. PCC 7425, the first structure to be solved at Diamond Light Source beamline I23 by means of S-SAD. Acta Cryst D *72*, 1174–1180. 10.1107/S2059798316015850.

73. Pfeifer, B.A., Admiraal, S.J., Gramajo, H., Cane, D.E., and Khosla, C. (2001). Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. Science *291*, 1790–1792. 10.1126/science.1058092.

74. Kabsch, W. (2010). XDS. Acta Crystallogr. D Biol. Crystallogr. *66*, 125–132. 10.1107/S0907444909047337.

75. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658–674. 10.1107/S0021889807021206.

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

77. Cowtan, K. (2010). Recent developments in classical density modification. Acta Crystallogr. D Biol. Crystallogr. *66*, 470–478. 10.1107/S090744490903947X.

78. Cowtan, K. (2006). The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr. D Biol. Crystallogr. *62*, 1002–1011. 10.1107/S0907444906022116.

79. Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G.W., McCoy, A., et al. (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr D Biol Crystallogr *67*, 235–242. 10.1107/S0907444910045749.

80. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., and Vagin, A.A. (2011). REFMAC5 for the refinement of macromolecular crystal structures. Acta Crystallogr. D Biol. Crystallogr. *67*, 355–367. 10.1107/S0907444911001314.

81. Tickle, I.J., Flensburg, C., Keller, P., Paciorek, W., Sharff, A., Vonrhein, C., and Bricogne, G. STARANISO.

82. Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320–W324. 10.1093/nar/gku316.

Figure 1

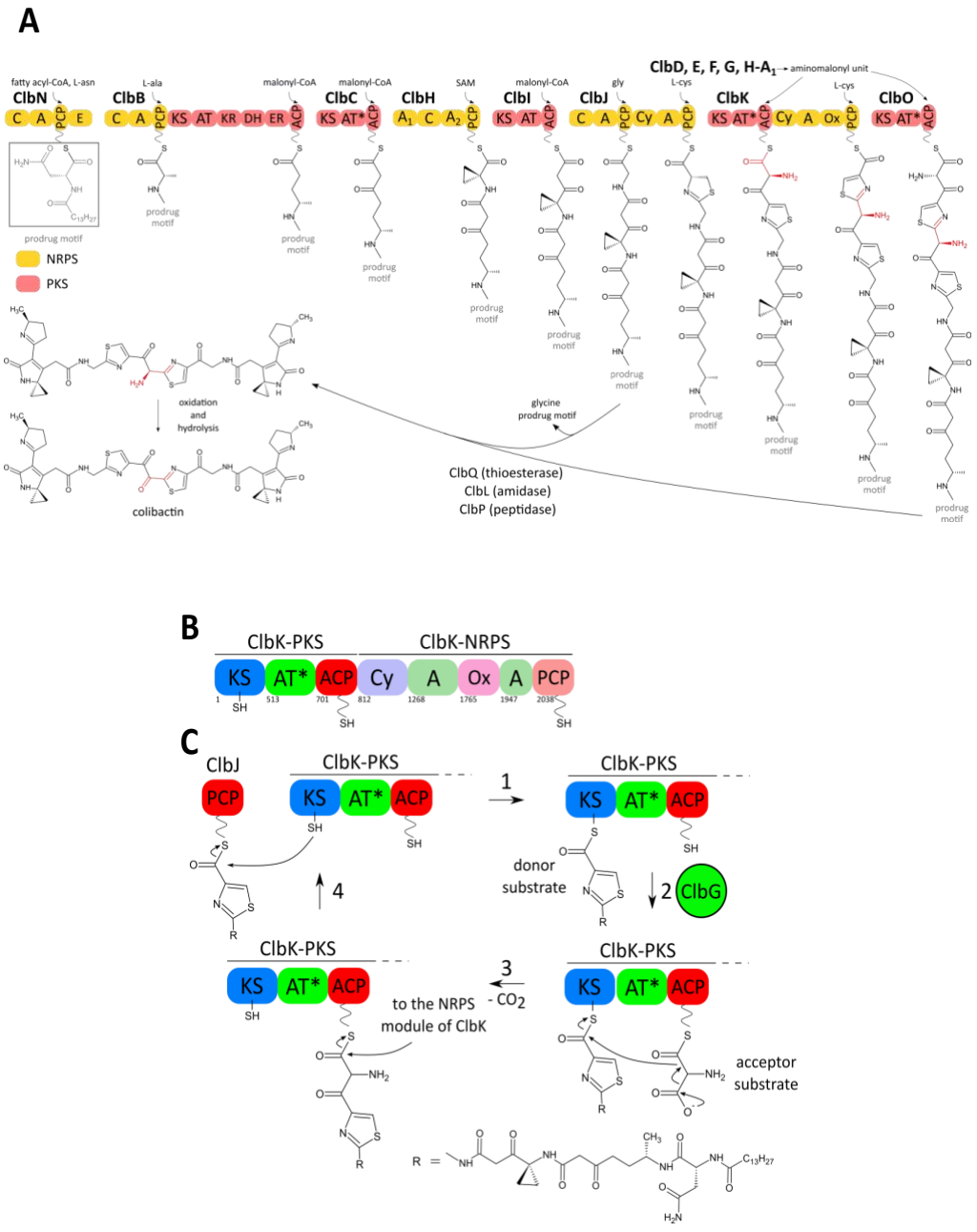

A
ClbN
ClbB
ClbC
ClbH
ClbI
ClbJ
ClbK
ClbO
ClbD, E, F, G, H-A1
aminomalonyl unit
fatty acyl-CoA, L-asn
L-ala
malonyl-CoA
SAM
gly
L-cys
C A PCP E
C A PCP KS AT KR DH ER ACP
KS AT* ACP
A1 C A2 PCP
KS AT ACP
C A PCP Cy A PCP
KS AT* ACP Cy A Ox PCP
KS AT* ACP
prodrug motif
NRPS
PKS
oxidation and hydrolysis
colibactin
glycine prodrug motif
ClbQ (thioesterase)
ClbL (amidase)
ClbP (peptidase)
B
ClbK-PKS
ClbK-NRPS
KS AT* ACP Cy A Ox A PCP
1 SH 513 701 SH 812 1268 1765 1947 2038 SH
C
ClbJ
ClbK-PKS
PCP
KS AT* ACP
1
2 ClbG
donor substrate
3
- CO2
4
to the NRPS module of ClbK
acceptor substrate
R =

Figure 2

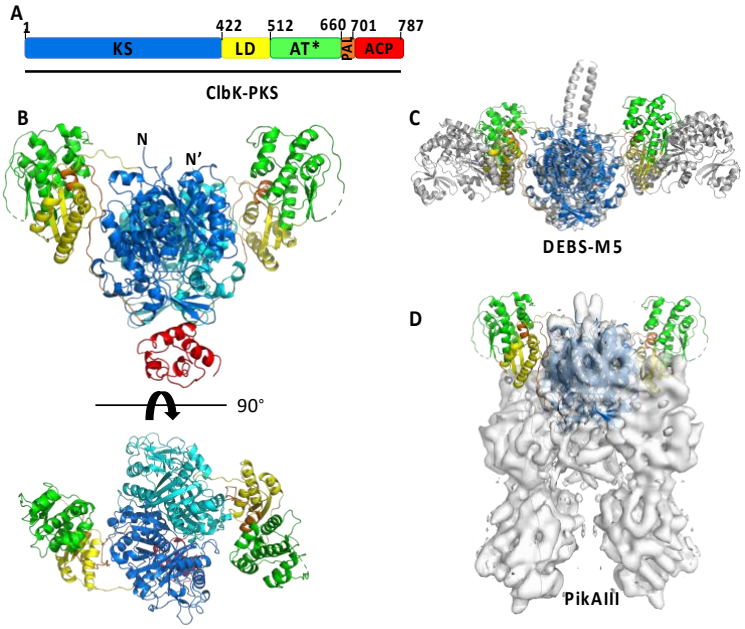

A
1
422
512
660
701
787
KS
LD
AT*
PAL
ACP
ClbK-PKS
B
N
N'
90°
C
DEBS-M5
D
PikAIII

Figure 3

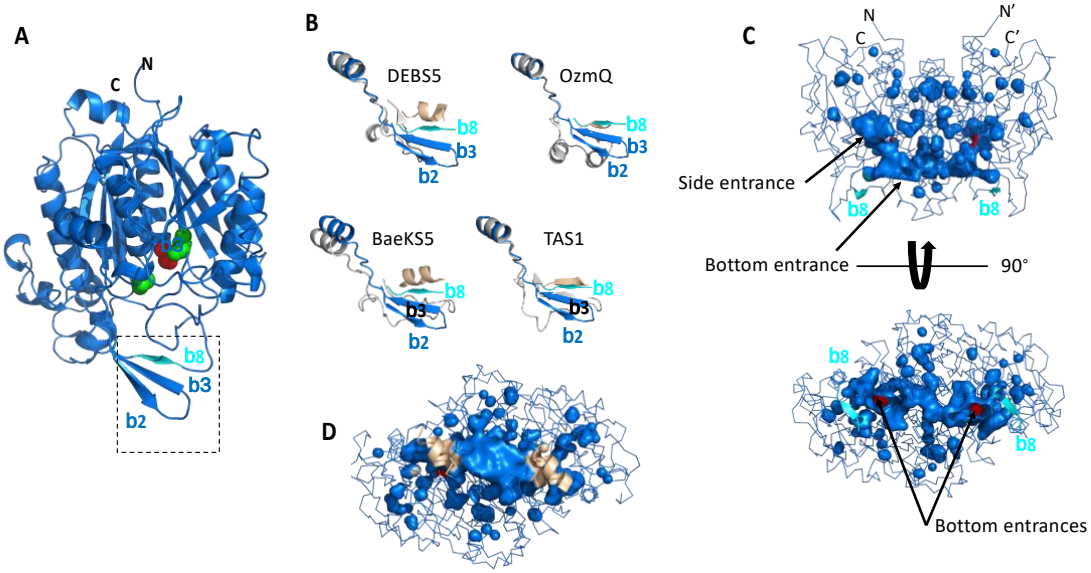

A
N
C
b8
b3
b2
B
DEBS5
OzmQ
BaeKS5
TAS1
b8
b3
b2
D
C
N
C
N′
C′
Side entrance
Bottom entrance
90°
b8
Bottom entrances

Figure 4

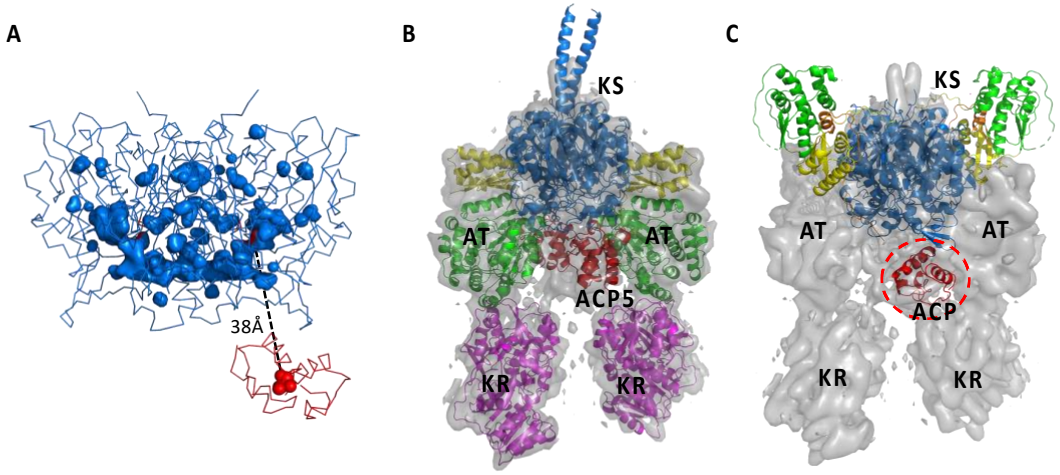
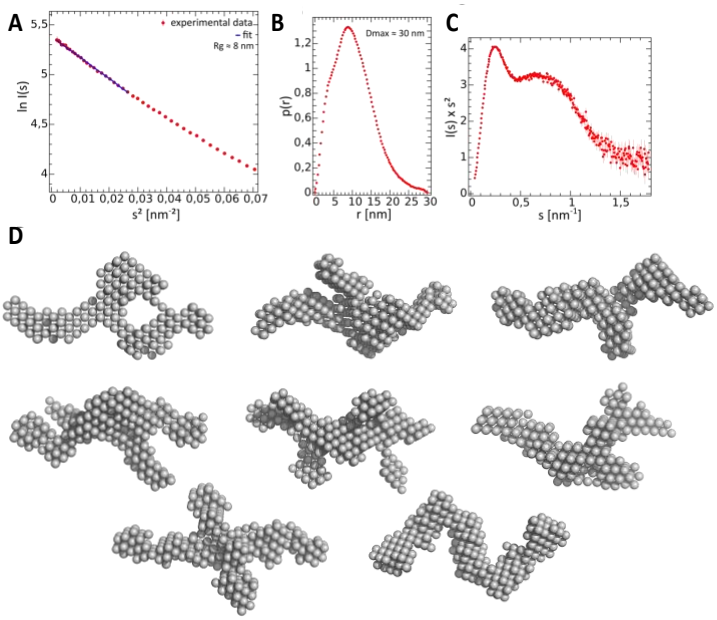


Figure 5


A
experimental data
fit
Rg = 8 nm
ln I(s)
s² [nm⁻²]
B
Dmax = 30 nm
p(r)
r [nm]
C
I(s) x s²
s [nm⁻¹]
D

Supplemental Information

Architecture of a PKS-NRPS hybrid megaenzyme involved in the biosynthesis of the genotoxin colibactin

Sarah Bonhomme, Carlos Contreras-Martel, Andréa Dessen, Pauline Macheboeuf

Email: pauline.macheboeuf@ibs.fr

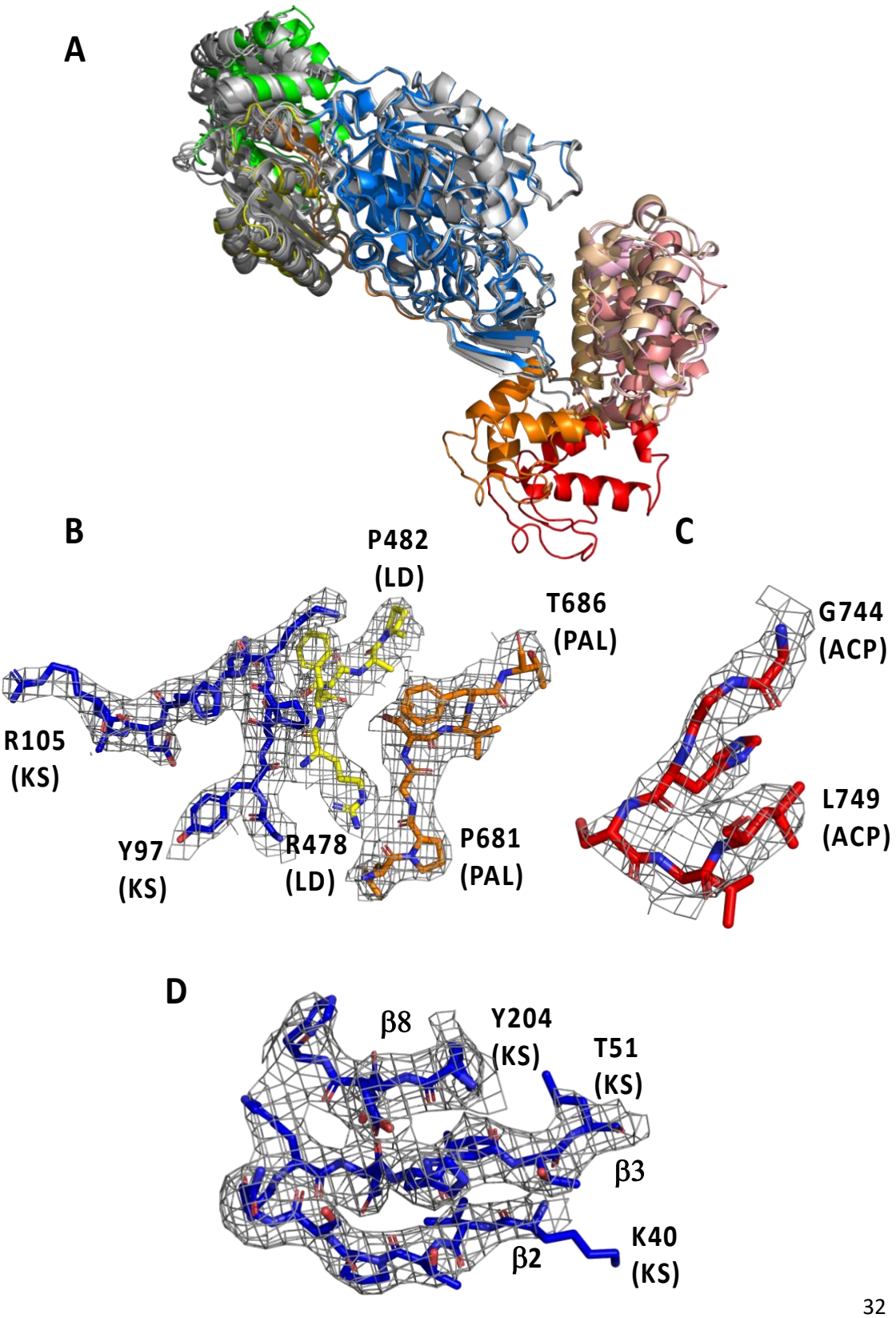

A
B
C
P482
(LD)
T686
(PAL)
R105
(KS)
Y97
(KS)
R478
(LD)
P681
(PAL)
G744
(ACP)
L749
(ACP)
D
β8
Y204
(KS)
T51
(KS)
β3
K40
(KS)
β2

Fig. S1. AlphaFold2 model predictions and solved structure of ClbK-PKS related to Figure 2. **(A)** Three models generated by the AlphaFold2 algorithm (colored in grey with ACP domains colored in different shades of pink) were aligned using Pymol and were compared to the solved structure of ClbK-PKS (ketosynthase (KS) domain in blue, linker domain (LD) in yellow, acyltransferase (AT*) domain in green, post-AT linker (PAL) in orange, ACP domain in red and orange). The KS-AT* core domains were perfectly superimposed between the models and with the solved structure. However, the ACP domains adopt different positions upon comparison of AlphaFold2 models (position 1 to 3 in different shades of pink) and in the ClbK-PKS structure (ACP domain in red and orange)**. (B, C and D)** 2Fo-Fc map contoured at 1.5 sigma around residues at the interface showing the KS in blue, the LD in yellow, the PAL in orange **(B)** and the ACP in red **(C).** The KS active site cap and clasping loop regions are also shown **(D)**. The map is represented within 2 Å of selected atoms.

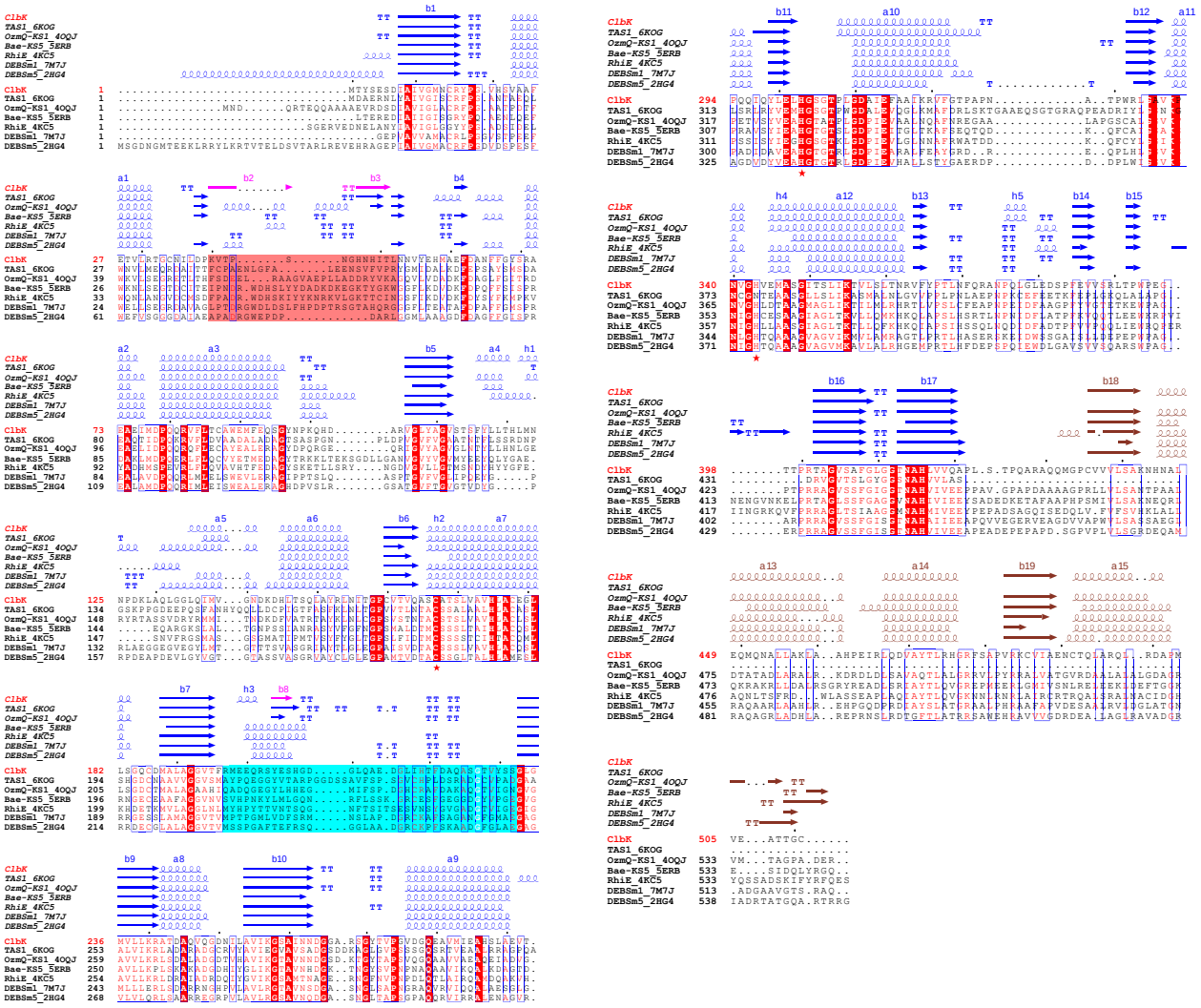


Fig. S2. Sequence alignment of KS and LD related to Figure 2**.** Sequences of several *trans*-AT and *cis*-AT PKSs were aligned and their KS and LD are shown here. The active cysteine and histidines are highlighted with a red star and strictly conserved residues are boxed in red. A column is framed in blue if more than 70% of its residues are similar according to physico-chemical properties. The salmon box highlights the clasping loop and the cyan box highlights the active site cap at the bottom of the KS. Secondary structures are colored in blue for the KS domain and in brown for the LD except for the ClbK-PKS β2, β3, β8 forming the clasping loop and active site cap that are colored in pink. The figure was prepared using ESPript [82].

Figure S3

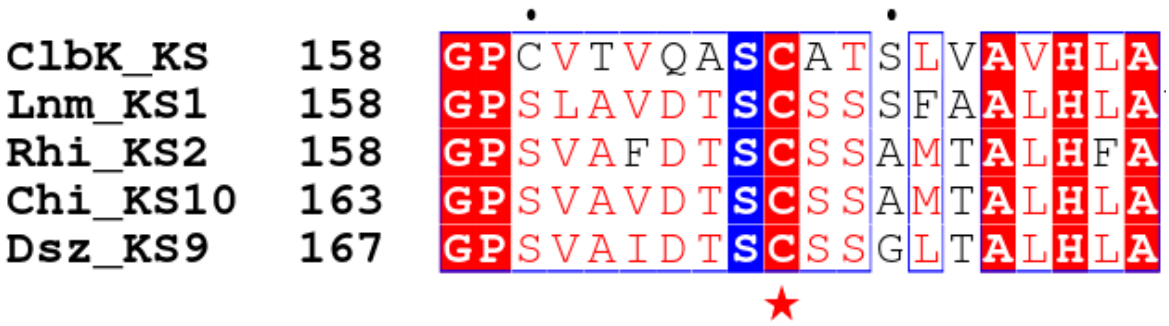


Fig. S3. Sequence alignment of the active cysteine region related to Figure 3**.** The PKS sequences are from proteins that accept donor substrates containing thiazole (ClbK, Lnm_KS1) or oxazole rings (Rhi_KS2, Chi_KS10, Dsz_KS9) These five KS display a serine residue (blue highlight) at the position before the active cysteine (highlighted with a red star). The other conserved residues are boxed in red. The blue box highlights a column with 70% of conserved. Lnm = Leinamycin, Chi = Chivosazol, Rhi = Rhizoxin, Dsz = Disorazol. The figure was prepared using ESPript.

Figure S4

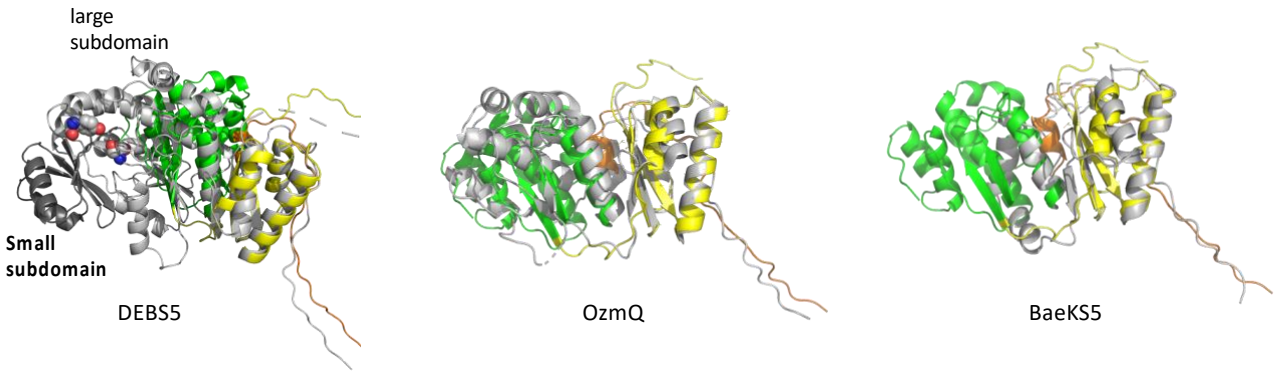
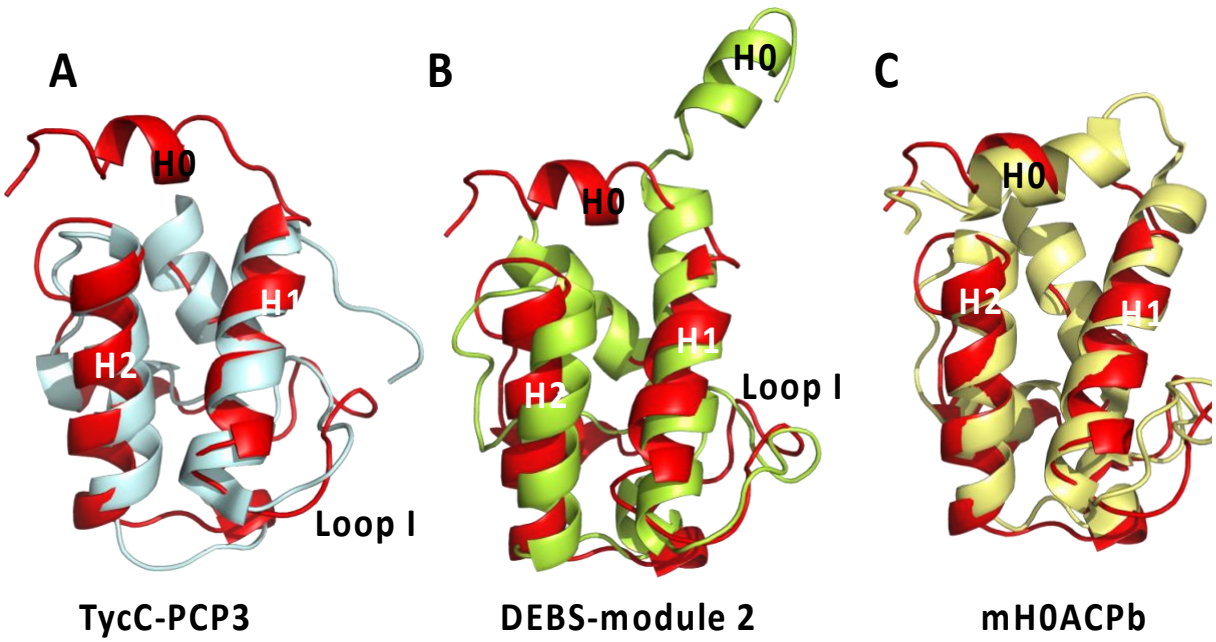


Fig. S4. Structural alignment of LD and AT domains related to Figure 2**.** ClbK-PKS is colored with the same colors as in Fig.2 (LD in yellow, AT* in green and PAL in orange) and the other structures are colored in grey (PDB: 2HG4, 4OQJ, 5ERB). The AT domain of DEBS module 5 possesses a small subdomain (dark grey) and a large subdomain (light grey). Its catalytic residues are shown as spheres.

Figure S5



D

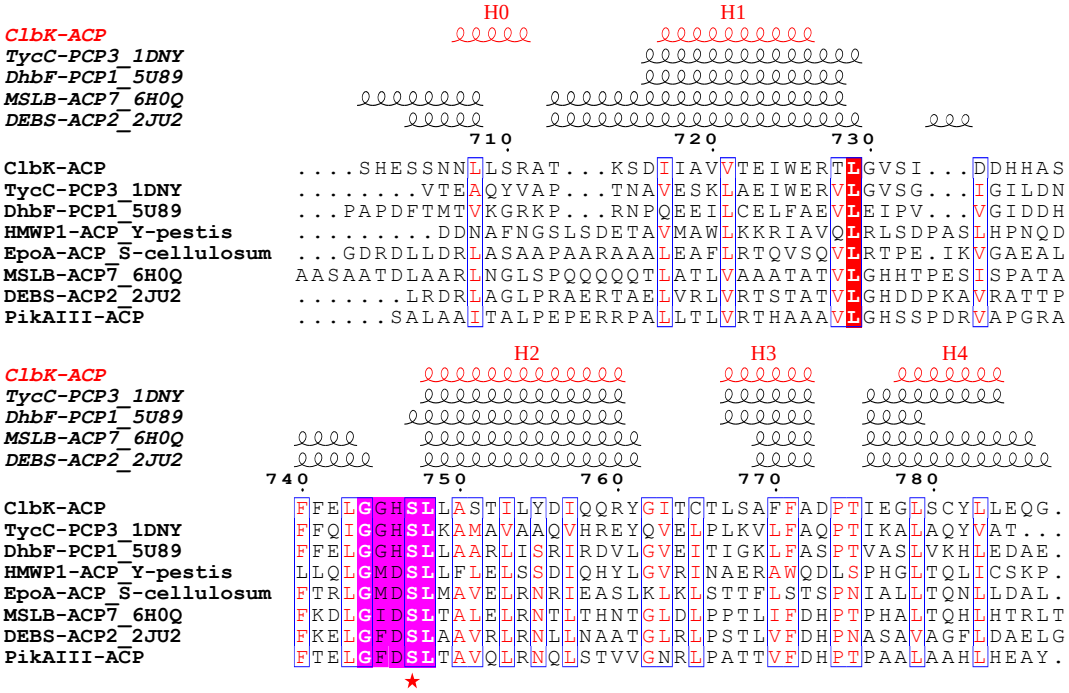


Fig. S5. Structural and sequence alignments of ACP/PCP domains related to Figure 2**. (A)** Superposition of ClbK-ACP (red) with TycC-PCP3 in light cyan (PDB 1DNY), **(B)** DEBS module

2 ACP in green (PDB 2JU2) and **(C)** mH0ACPb in pale yellow (PDB 6H0Q). ClbK-ACP possesses a mixture of ACP and PCP features: it starts with an H0 helix, very common in ACP from *cis*-AT PKS but displays a short loop I, similar to what is found in PCPs. **(D)** The sequence alignment shows the GXXSL motif boxed in pink and highlights the presence of a GGHS motif, typical of PCP domains, in ClbK ACP. The other ACPs possess a GXDS motif. The conserved serine of the motif is highlighted with a red star. Secondary structures are colored in red for ClbK and, when available, in black for the other PCPs/ACPs. The figure was prepared using ESPript.

Figure S6

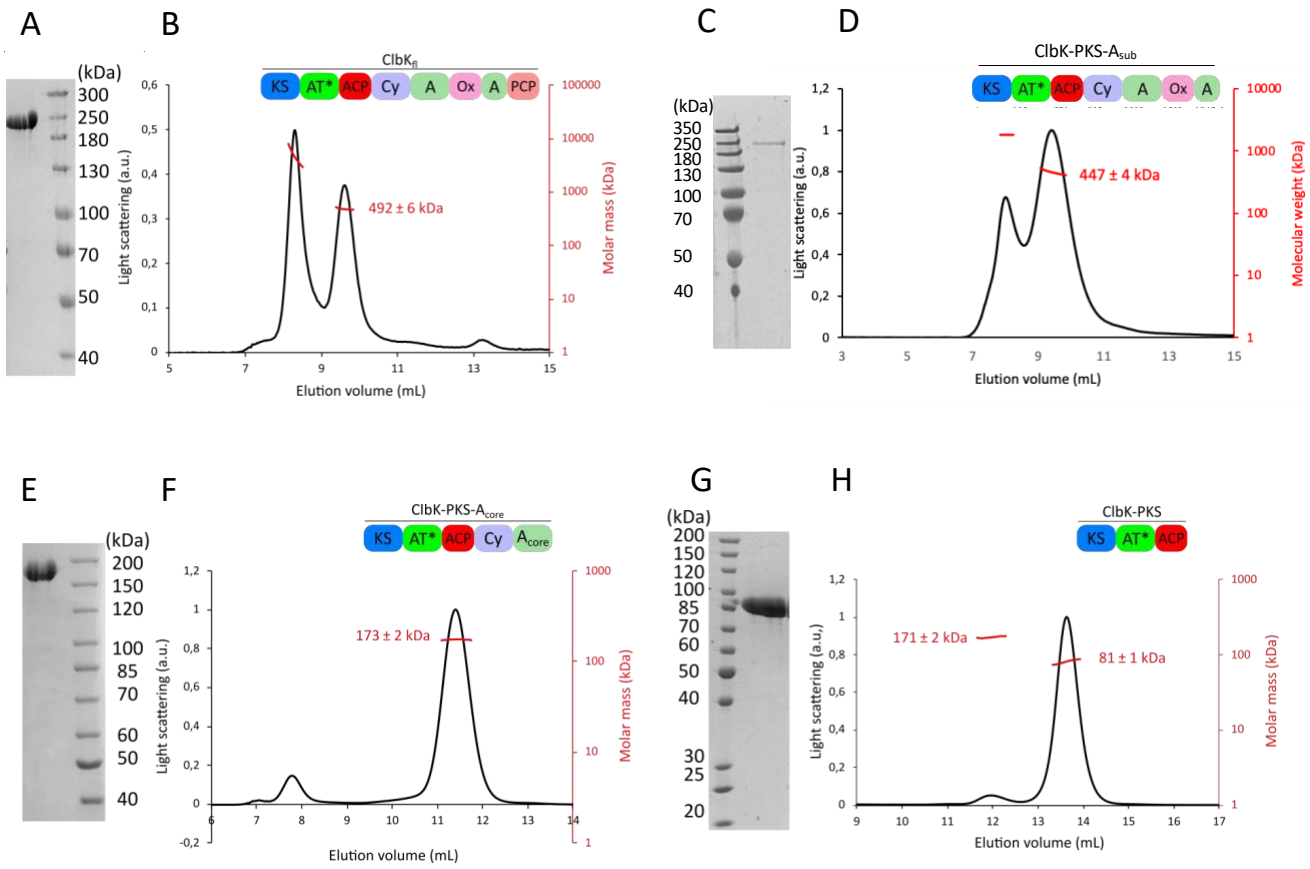


Fig. S6. Estimation of purity and SEC-MALLS analysis of ClbK fragments related to Figure 2**.** SDS-PAGE gels indicate that ClbK$_{fl}$ **(A), ClbK-PKS-A$_{sub}$ (C),** ClbK-PKS-A$_{core}$ **(E)** and ClbK-PKS **(G)** are pure. **(B, D, F, H)** SEC-MALLS analysis of ClbK$_{fl}$ **(B)**, ClbK-PKS-A$_{sub}$ (**D**), ClbK-PKS-A$_{core}$ **(F)** and ClbK-PKS **(H)**. **(B)** In solution, ClbK$_{fl}$ forms a polydispersed (aggregate) and a dimeric species (theoretical monomeric molar mass MM: 240 kDa). **(D)** ClbK-PKS-A$_{sub}$ shows a peak of aggregate and a peak of dimer (theoretical monomeric molar mass MM: 223 kDa). **(F)** ClbK-PKS-A$_{core\ i}$S purely monomeric in solution (theoretical monomeric MM: 187 kDa). The peak at 7.5 mL contains a small amount of aggregates **(F)** ClbK-PKS forms mostly monomers and a small amount of dimers in solution (theoretical monomeric MM: 89 kDa).

References

82. Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res *42*, W320–W324. 10.1093/nar/gku316.