

ClbP Is a Prototype of a Peptidase Subgroup Involved in Biosynthesis of Nonribosomal Peptides*[§]

Received for publication, February 4, 2011, and in revised form, June 17, 2011. Published, JBC Papers in Press, July 27, 2011, DOI 10.1074/jbc.M111.221960

Damien Dubois^{†§}, Olivier Baron^{¶***§§†}, Antony Cougnoux[§], Julien Delmas^{†§}, Nathalie Pradel[§], Michèle Boury^{¶***††}, Bernadette Bouchon^{¶¶}, Marie-Agnès Bringer[§], Jean-Philippe Nougayrède^{¶¶***††}, Eric Oswald^{¶¶***§§1}, and Richard Bonnet^{†§2}

From the [†]Laboratoire de Bactériologie Clinique, Centre Hospitalier de Clermont-Ferrand, Clermont-Ferrand F-63003, France, [§]Institut National de Recherche Agronomique, Clermont Université, JE 2526 Université d'Auvergne, Unité Sous Contrat 2018, Clermont-Ferrand F-63001, France, ^{¶¶}INSERM U990, Clermont-Ferrand, France, ^{¶¶}INSERM U1043, Toulouse F-31076, France, ^{¶¶}Centre National de Recherche Scientifique, U5282, Toulouse F-31076, France, ^{**}Institut National de Recherche Agronomique, Unité Sous Contrat, Toulouse F-31076, France, ^{††}Université Paul Sabatier-Toulouse III, Centre de Physiopathologie de Toulouse Purpan, Toulouse F-31076, France, and ^{§§}Centre Hospitalier Universitaire de Toulouse, Hôpital Purpan, Service de Bactériologie-Hygiène, Toulouse F-31059, France

The *pks* genomic island of *Escherichia coli* encodes polyketide (PK) and nonribosomal peptide (NRP) synthases that allow assembly of a putative hybrid PK-NRP compound named colibactin that induces DNA double-strand breaks in eukaryotic cells. The *pks*-encoded machinery harbors an atypical essential protein, ClbP. ClbP crystal structure and mutagenesis experiments revealed a serine-active site and original structural features compatible with peptidase activity, which was detected by biochemical assays. Ten ClbP homologs were identified *in silico* in NRP genomic islands of closely and distantly related bacterial species. All tested ClbP homologs were able to complement a *clbP*-deficient *E. coli* mutant. ClbP is therefore a prototype of a new subfamily of extracytoplasmic peptidases probably involved in the maturation of NRP compounds. Such peptidases will be powerful tools for the manipulation of NRP biosynthetic pathways.

Polyketide (PK),³ nonribosomal peptide (NRP), and hybrid NRP-PK natural products are very interesting compounds in drug development strategies (1). Their biosynthesis occurs on multimodular enzymatic assembly lines called megasynthases (2). The conformations that establish biological activity are attained by tailoring and editing enzymes such as acylases, alky-

lases, glycosylases, and oxidoreductases during chain growth and after thioesterase-mediated release. The dedicated tailoring enzymes are encoded by genes clustered with the assembly line genes for coordinated regulation.

Recently, Nougayrède *et al.* (3, 4) identified in *Escherichia coli* strains and other Enterobacteriaceae a large (54 kb) genomic island named *pks*, which encodes a typical cluster of NRP synthases (NRPS), PK synthases (PKS), and hybrid NRPS/PKS. The *pks* island was detected in *E. coli* strains involved in extraintestinal infections such as urinary tract infection and septicemia but also in commensal strains from healthy people (3, 5, 6). The genetic and functional analyses of the *pks* island indicate that it encodes for synthesis of a PK-NRP hybrid compound, named colibactin. *E. coli* strains expressing this gene cluster cause DNA double-strand breaks in human eukaryotic cells during infection (3). This DNA damage induces phosphorylation of the H2AX histones (γ H2AX) and activates the ATM-Chk2 signaling pathway that leads to transient G₂/M cell cycle arrest and cell swelling (megalocytosis). Exposed cells exhibit signs of incomplete DNA repair, which leads to chromosomal instability, an increase in gene mutation frequency and anchorage-independent colony formation, demonstrating the mutagenic and transforming potentials of colibactin (7).

In addition to eight NRP and PK megasynthases, the *pks* island encodes nine accessory, tailoring and editing enzymes (3). Eight of these proteins, including the protein ClbP, are required to induce DNA damage in infected cells. No protein similar to ClbP has been reported so far as an accessory or tailoring enzyme. Here, we determined the crystal structure of the ClbP major domain, its biochemical activity, and we characterized distant ClbP homologous proteins that form a new subgroup of serine-reactive peptidases associated with NRP assembly lines.

EXPERIMENTAL PROCEDURES

Strains, Plasmids, and Media—The strains and plasmids used in this study are listed in supplemental Tables 1 and 2, respectively. *E. coli* strains were routinely grown at 37 °C in lysogeny broth (LB) or on LB agar plates. Ampicillin (100 μ g/ml), kanamycin (50 μ g/ml), chloramphenicol (25 μ g/ml), or

* This work was supported by le Ministère Français de l'Éducation Nationale, de la Recherche, et de la Technologie Grant JE2526 and l'Institut National de la Recherche Agronomique Grant USC-2018. Molecular graphics images were produced using the University of California, San Francisco Chimera package (National Institutes of Health Grant P41 RR-01081).

This article is dedicated to the memory of our recently deceased friend and colleague Dr. Olivier Baron.

The atomic coordinates and structure factors (code 3O3V) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (<http://www.rcsb.org/>).

[§] The on-line version of this article (available at <http://www.jbc.org/>) contains supplemental Tables 1–6, Figs. 1–4, and additional references.

[†] In memoriam.

¹ To whom correspondence may be addressed. Tel.: 33-5-61-19-39-91; Fax: 33-5-61-19-39-75; E-mail: e.oswald@envt.fr.

² To whom correspondence may be addressed. Tel.: 33-4-73-75-49-20; Fax: 33-4-73-75-49-22; E-mail: rbonnet@chu-clermontferrand.fr.

³ The abbreviations used are: PK, polyketide; NRP, nonribosomal peptide; NRPS, NRP synthases; PKS, PK synthases; ZmA, zwittermicin A.

tetracycline (200 ng/ml) were added as required to the medium. Non-*E. coli* strains were grown as recommended by the Deutsche Sammlung von Mikroorganismen und Zellkulturen or Collection de l'Institut Pasteur collections or by the givers.

DNA Cloning for ClbP Investigations—The DNA sequences of *clbP* were amplified from the genome of *E. coli* strain IHE3034 (3). The DNA fragments encoding the entire ClbP and the catalytic domain of ClbP designated ClbPpep were amplified by PCR using primer pairs clbPMK-F and clbPBN-R, and clbP-NdeI2-F and clbP-NotI2-R (supplemental Table 3), respectively. The PCR reactions for cloning experiments were performed using the High Fidelity Platinum® Taq polymerase or the Expand High Fidelity PCR System according to the manufacturer's instructions (Invitrogen). The amplified fragments were purified, double-digested by MfeI/BamHI for cloning at the EcoRI/BamHI sites of pASK-IBA33plus to obtain ClbP-RGS6His (ClbP + C-terminal in-frame RGS6His tag), or double-digested by and cloned at the NdeI/NotI sites of pET9am to obtain pClbPpep (ClbP catalytic domain). In other experiments, ClbP-6His was subcloned from pMB702 (3) into pBRSK vector (8) to obtain pOB901. The constructions were checked by double-stranded DNA sequencing (Genome Express). Plasmids were subsequently electroporated into competent *E. coli* strains: p33ClbP and pOB901 into DH10B/pBACpksΔ*clbP* strain for HeLa cellular challenges, and pClbPpep into BL21(DE3) strain for ClbPpep and derivative overexpression (original strains are listed in supplemental Table 1).

Site-directed Mutagenesis—The site-directed mutagenesis experiments were performed using the Quik-Change™ site-directed mutagenesis kit according to the manufacturer's instructions (Stratagene). The substitutions E159K, S188A, H257A, G328S, F316R/G328S, N331A, and C337A were individually introduced into ClbP-RGS6His cloned into the p33ClbP plasmid and the substitution Y186E was introduced into ClbPpep cloned into the pClbPpep plasmid. In other experiments, ClbP-6His S95A, K98T, and Y186G mutants were generated by "reverse PCR" using pOB901 as template. Reverse PCR products were purified, ligated and transformed into DH5α. Successful mutations were confirmed by double DNA sequencing and the final constructs were used to transform *E. coli* DH10B/pBACpksΔ*clbP* for p33clbP and pOB901 derivative vectors, or *E. coli* BL21(DE3) for pClbPpep derivative vector. The mutagenic primers are listed in supplemental Table 4.

ClbPpep and ClbPpepY186E Production and Purification—The pClbPpep plasmid encodes the peptidase domain of ClbP without the first 102 nucleotides and the last 390 nucleotides of the *clbP* gene. These sequences encode the signal peptide and three putative anchor hydrophobic transmembrane helices that were removed to facilitate cytoplasmic expression and protein purification. The *E. coli* BL21(DE3) transformed with pClbPpep plasmid was cultured in Terrific broth medium supplemented with 50 μg/ml kanamycin, 0.4 M D-sorbitol, and 2.5 mM β-betaine. The cell culture was grown at 37 °C with shaking to an $A_{600\text{ nm}} \sim 0.7$. Overexpression of the *clbPpep*-encoding gene was induced with 0.2 mM isopropyl β-D-thiogalactopyranoside overnight at 25 °C with shaking. Cells were harvested by centrifugation (10,000 × *g* for 10 min at 4 °C), resuspended in 20 mM Tris-HCl buffer (pH 9.0) (4 ml/g of pellet). After addition

of 5 μg of lysozyme (2 h at 4 °C), cells were disrupted by ultrasonic treatment (four times for 30 s, each time at 20 watts). The extract was clarified by centrifugation at 14,000 × *g* for 60 min at 4 °C. After addition of 2 mg of DNase I (15 min at room temperature; Roche), the supernatant (10,000 × *g* for 60 min at 4 °C) was dialyzed overnight against 20 mM Tris-HCl buffer (pH 9.0) and filtrated (0.22 μm filter; Millex-GS, Millipore, Carrigtwohill, Ireland). Purification was carried out by ion-exchange chromatography onto a HiTrap™ Q Sepharose™ High Performance column (10 ml; Amersham Biosciences) equilibrated with 20 mM Tris-HCl (pH 7.0) and eluted with a linear NaCl gradient (0 to 500 mM). The ClbPpep-containing elution peak was then purified by gel filtration chromatography onto a Superose™ 12 column (3.2 × 30 cm; Amersham Biosciences), which had been equilibrated and eluted with 100 mM NaCl and 5 mM phosphate buffer (pH 8.0). The ClbPpep-containing elution peak was extensively dialyzed against 50 mM NaCl, 5 mM phosphate buffer (pH 8.0), concentrated by ultrafiltration to 10 mg/ml for crystallization. The enzyme was >95% homogeneous as determined by Coomassie Blue staining after SDS-PAGE. The N-terminal sequencing of the isolated protein was achieved by sequential Edman reaction (Plate-forme de Protéomique Analytique et Fonctionnelle, Institut National de Recherche Agronomique, Nouzilly, France). Production and purification of ClbPpepY186E were performed in the same way using the plasmid pClbPpepY186E.

Crystallization and Structure Determination—ClbP crystals were grown in hanging drops over a well solution of 0.8 M monosodium dipotassium phosphate buffer (pH 7.0). Crystals reached a size of 40 μm × 40 μm × 70 μm within 4 weeks. Before data collection, crystals were immersed for ~30 s in a cryoprotectant solution (20% sucrose, 1.0 M monosodium dipotassium phosphate, pH 7.0) and were flash-cooled in liquid nitrogen. Data were collected using a Q315r ADSC-CCD detector on ESRF beamline 14-4 at the European Synchrotron Radiation Facility (Grenoble, France). Reflexions were indexed, integrated, and scaled using the CCP4 package (9). The initial model was obtained by molecular replacement with the program PHASER (10) and the x-ray structures (Protein Data Bank codes 1E15, 1RGY, and 1FR1) as a search model. The structure was automatically and manually refined with REFMAC5 and COOT programs, respectively (11, 12). Cross-validation was used throughout, and 5% of the data were used for the R_{free} calculation. The stereochemical quality of the models was monitored with the PROCHECK program (13). Ramachandran plots were calculated by RAMPAGE (14). Processing and crystallographic refinement statistics for ClbPpep crystal structure are listed in supplemental Table 5.

Enzyme Assays—D-Amino peptidase activity was detected by monitoring the formation of *p*-nitroaniline from Gly-, D-Ala-, L-Ala-, and L-Ala-L-Ala-*p*-nitroanilide (Bachem) at 405 nm in a 100 mM Tris, 75 mM NaCl buffer (pH 7.5) at 25 °C. β-Lactamase activity was tested by monitoring the hydrolysis of nitrocefin (Calbiochem) at 482 nm in a 100 mM Tris, 75 mM NaCl buffer (pH 7.5) at 25 °C.

Mass Spectrometry—ClbPpep (20 μM) was incubated at 20 °C for 24 h in a 100 mM Tris, 75 mM NaCl buffer (pH 7.5) alone, or with 20 mM imipenem (Sigma), in a 1-ml final volume. The

Structure and Functionality of ClbP Peptidase

reaction mixtures were desalted by dialysis against a 5 mM ammonium bicarbonate buffer, before analysis by MALDI-TOF mass spectrometry. The sample was spotted onto the stainless steel target (Applied Biosystems) and an equal volume of matrix solution (14 mg/ml sinapinic acid in 50% acetonitrile in water containing 0.1% trifluoroacetic acid) was deposited into the sample drop. MALDI-TOF mass spectra were acquired on a Voyager DE-PRO mass spectrometer (Applied Biosystems) in a linear positive mode, using an accelerating voltage of 25 kV, a grid voltage of 93%, a pulse delay extraction of 700 ns, in the 15,000–50,000 m/z mass range. Calibration was performed in a close external mode using a protein calibration mixture (LaserBio Labs).

Western Blot Experiments—Protein samples (20–25 μg for each sample) were resolved by SDS-PAGE separation electrophoresis on 12.5% polyacrylamide gels, immobilized onto PVDF membranes, and analyzed by immunoblot using the ECL blotting system (Amersham Biosciences), and monoclonal anti-RGS6His or anti-5His antisera (Qiagen, France), according to the manufacturer's instructions.

DNA Cloning of ClbP Homolog-encoding Genes—The DNA fragments encoding ClbP homologs (the N-terminal signal sequence, the large extracytoplasmic and catalytic domain, and the three C-terminal transmembrane helices, without the putative “half-size” ABC-type exporter domain if present) of *Hahella chejuensis* KCTC 2396^T (locus tag *HCH_03446*), *Bacillus cereus* UW85 (locus tag *zmaM*), *Bacillus weihenstephanensis* KBAB4 (locus tag *BcerKBAB4_1783*), *Clostridium cellulolyticum* H10 (locus tag *Ccel_2389*), *Bacillus mycoides* DSM 2048 (locus tag *bmyco0001_51990*), *Bacillus pseudomycooides* DSM 12442 (locus tag *bpmx0001_2150*), and *Staphylococcus aureus* Mu50 (locus tag *SAV2441*) were amplified by PCR using primers listed in [supplemental Table 3](#), purified, double-digested by MfeI or EcoRI (according to primer name) and BamHI or XhoI for cloning at the EcoRI/BamHI or XhoI sites of pASK-IBA33plus (C-terminal in-frame with RGS6His tag) resulting in plasmids listed in [supplemental Table 2](#). The constructions were checked by double-stranded DNA sequencing (Genome Express). Plasmids were electroporated into competent DH10B/pBACpks Δ clbP.

Eukaryotic Cell Culture, Bacterial Infections, γ H2AX Staining, and Cell Cycle Analysis—Cell assays were performed as described previously with minor modifications (3). For bacterial infections, overnight LB-ampicillin cultures of bacteria were diluted in interaction medium (DMEM, 5% FCS, 25 mM HEPES, 100 $\mu\text{g}/\text{ml}$ ampicillin). For experiments with ClbP homologs, tetracycline was added to overnight cultures and interacting medium to induce their expression. Approximately 50% confluent HeLa cell cultures (ATCC CCL2) were then infected with a multiplicity of infection (number of bacteria per eukaryotic cell) of 100. Cells were washed 4 h after inoculation and incubated in DMEM 10% FCS, 200 $\mu\text{g}/\text{ml}$ gentamicin until analysis. Cell morphology was examined by Giemsa staining at 72 h after infection. For γ H2AX staining, cells were fixed at 4 or 20 h after infection in 4% formaldehyde and incubated with anti-phospho-H2AX (Ser-139) antibodies (JBW301, Upstate) followed by FITC- or rhodamine-conjugated secondary antibodies. DNA was stained with DAPI or TO-PRO-3 (Invitrogen)

and images were acquired with a Zeiss LSM 510 Meta or Olympus IX70 laser scanning confocal microscope, and the confocal aperture was set to achieve a z optical thickness of $\sim 0.5 \mu\text{m}$. For cell cycle analysis, cellular suspensions were fixed and permeabilized with 70% ethanol at 48 h after infection and then treated with 50 $\mu\text{g}/\text{ml}$ propidium iodide and 250 $\mu\text{g}/\text{ml}$ RNase. DNA content data were acquired with an SLR II or FACSCalibur flow cytometer (Becton Dickinson) and analyzed with FlowJo software (Tree Star).

Bioinformatics—The search for ClbP homologous proteins and analysis of their topology were performed using BLAST (15) and the InterProScan server. Multiple sequence alignment of peptidase domains and phylogenetic tree were performed using COBALT and TOPALi software (version 2.5) (16, 17). Analysis of the genetic context of ClbP- and FmtA-like-encoding genes (more or less 30 kb of both sides of the genes) was performed manually from the NCBI genome database. The analysis of PK and NRP megasynthase domains and their substrate specificity was performed using the SEARCH NRPS-PKS program.

RESULTS

In Silico Analysis of ClbP—The *clbP* open reading frame was predicted to encode a 504-amino acid periplasmic inner membrane-anchored protein comprising three parts: (i) a C-terminal domain of three transmembrane helices (residues 390–412, 433–455, and 465–485), (ii) an N-terminal signal sequence (ALA-30-QE-31), and (iii) a large periplasmic domain (positions 31–375). According to primary structure comparisons, the last domain exhibits the two conserved motifs (⁹⁵SMS⁹⁸K versus SxxK and ¹⁸⁶YAS versus YxN) that characterize the catalytic center of the MEROPS S12 enzyme family (18) and is therefore designated hereafter as ClbPpep. The S12 family is a heterogeneous group of active site serine enzymes, of which the substrates are peptides or closely related compounds such as β -lactams. The activity and biological function of a few of these enzymes have been described, such as class C β -lactamases, R61 D-Ala-D-Ala carboxypeptidase B, DmpB aminopeptidase, Pab87 peptidase, D-amino acid amidase, and EstB esterase (19–24). However, the activity and the function of numerous enzymes of the MEROPS S12 family remain unassigned, notably the FmtA-like proteins (25), which have been previously reported to be related to ClbP (3).

Structure Determination of ClbP Peptidase Domain—To investigate the function and activity of ClbP, the structure of the ClbPpep domain was determined by x-ray diffraction. The structure was refined against diffraction data extending to 2.4 Å resolution. The electron density map ([supplemental Fig. 1](#)) showed well defined density throughout most of the structure, with the exception of the six N-terminal residues. Diffuse density indicated disorder in the loop harboring residues 241–249. The stereochemical parameters of the model were satisfactory ([supplemental Table 5](#)); a Ramachandran plot showed no residues in disallowed regions of φ/ψ space. The final model included three ClbPpep monomers and 277 water molecules.

The three-dimensional structure of ClbPpep consists of two structural regions, all- α and α/β (Fig. 1A). The α/β region is formed by residues 41–112 and 206–375, and the domain folds

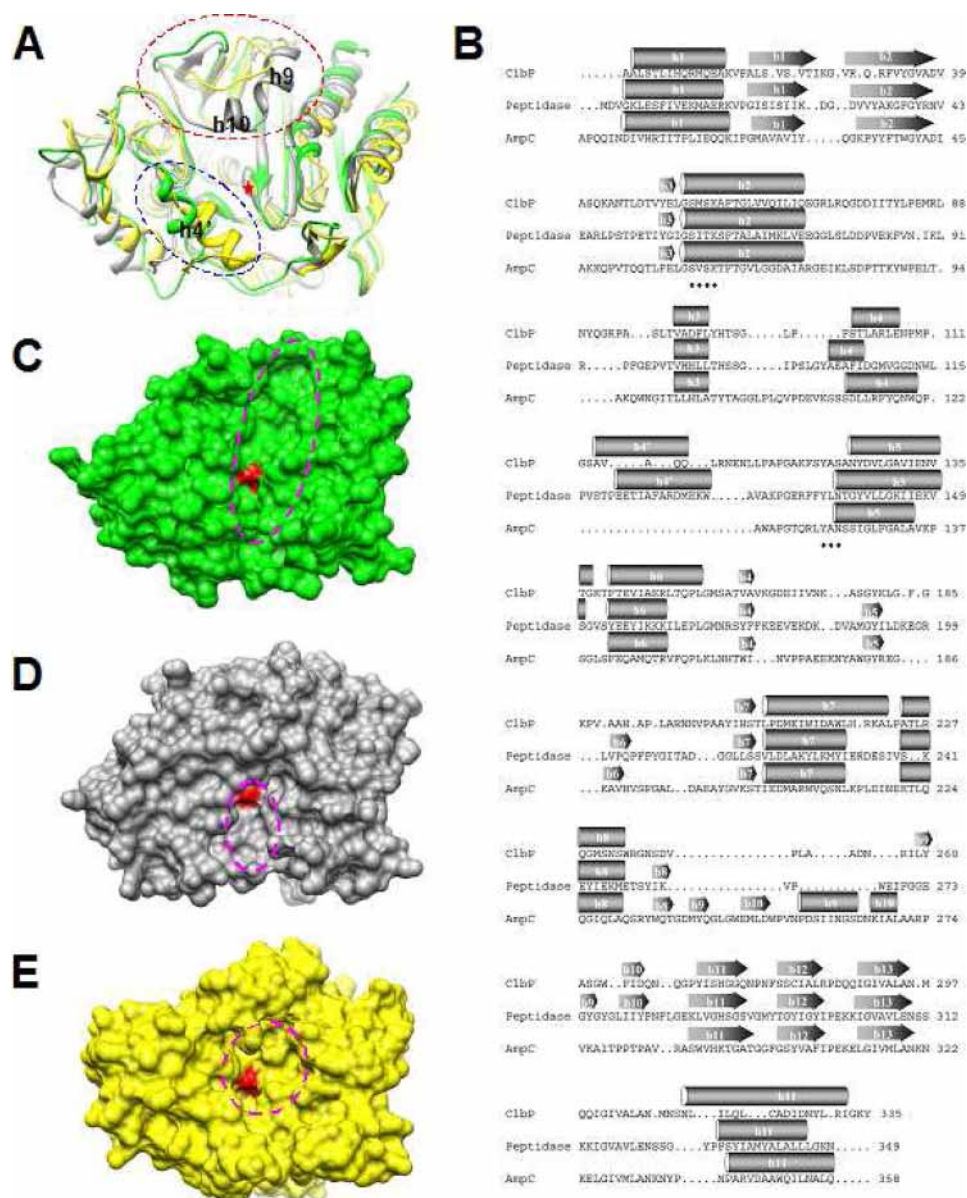


FIGURE 1. Overall structure of ClbPpep, AmpC β -lactamase (Protein Data Bank code 1KVL) and Pab87 peptidase (Protein Data Bank code 2QMI). A, superimposition of ClbPpep (green), AmpC (gray) (19), and Pab87 peptidase (yellow) (21) presented as ribbons. The loops including helices h9 and h10 are circled in red. Helices h4' are circled in blue. The active serine is indicated by a red star. ClbPpep (B), AmpC (C), and Pab87 peptidase (D) surfaces. The catalytic grooves are circled in pink, and the active serine surfaces are shown in red. E, sequence alignments with secondary structures of ClbPpep with Pab87 peptidase and *E. coli* AmpC β -lactamase. The alignment was performed based on the superposed structures. The motifs SxxK and YxN, which defined the MEROPS S12 peptidase family, are indicated by a diamond.

as a seven-stranded anti-parallel β -sheet (b1, b2, b9, b10, b11, b12, and b13) with six α -helices and three β -strands (b3, b4, and b7) packed on both faces of the sheet. The all-helical region (residues 113–205) contains four helices and loops. The conserved motifs ⁹⁵SMSK and ¹⁸⁶YAS, which correspond to S12 enzyme active site, are located in a large groove, between the two structural domains near the N terminus of the first helix of the all- α domain.

As expected, the most pronounced homology using the DALI program was obtained with the structure of S12 family members (supplemental Table 6), especially *E. coli* AmpC β -lactamase and Pab87 peptidase (26). Although their overall fold was clearly similar, there are significant differences affecting the accessibility of the active site from the surrounding sol-

vent. In comparison with AmpC β -lactamases, the upper part of the catalytic pocket was largely open, because of the absence of two helices (h9 and h10) (Fig. 1, A and E), as observed in Pab87 peptidase and R61 carboxypeptidase (21, 27). Moreover, the remaining structural element of ClbP in this area, the loop between residues 295–315, b9, and b10 strands (residues 310 to 311 and 316–318, respectively), moved away from the catalytic pocket, resulting in a widening of the catalytic groove, which was unusually larger than the known structures of S12 enzymes. In the bottom part of the catalytic pocket, ClbPpep and Pab87 peptidases harbored the additional helix h4' (residues 166–170) in contrast with AmpC enzymes (Fig. 1, A and E). This helix slightly obstructed accessibility to the ClbPpep catalytic pocket. However, the entrance of the binding site was more

Structure and Functionality of ClbP Peptidase

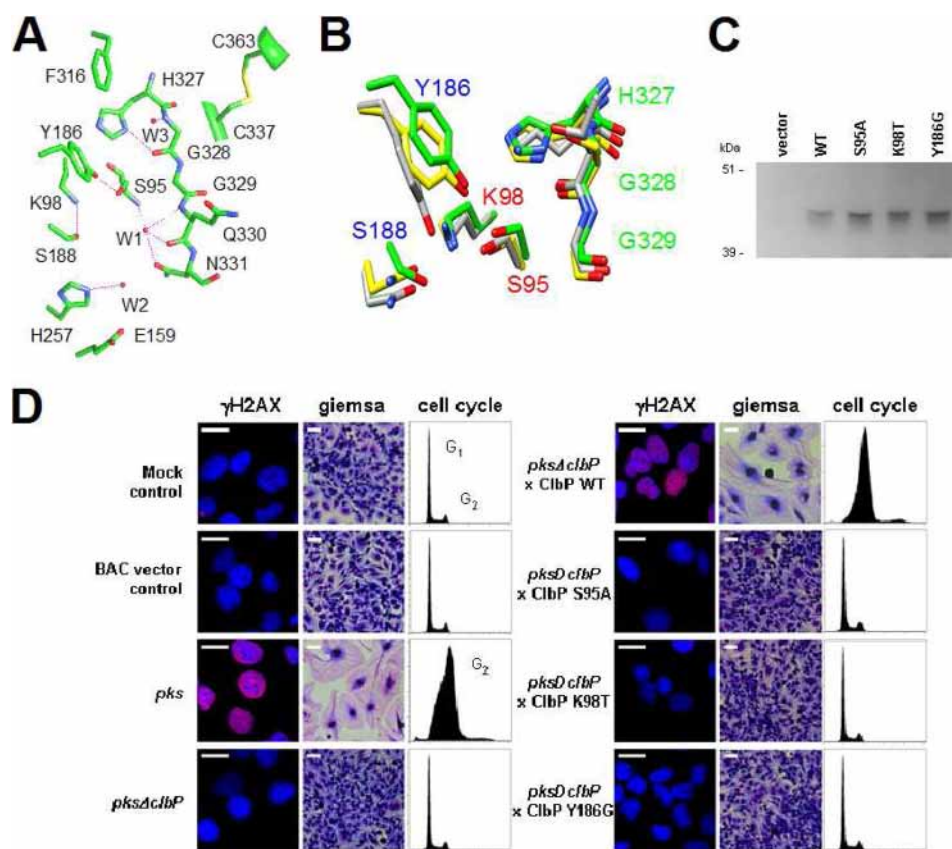


FIGURE 2. Key residues in the ClbP active site. A, residues of ClbP located in close vicinity of the binding site. Hydrogen bonds are indicated by pink dashed lines, and the water molecules are indicated by W. Carbon atoms are colored green, oxygen atoms are colored red, and nitrogen atoms are colored blue. B, overlay of penicillin-recognizing motifs of ClbP, Pab87 peptidase, and AmpC β -lactamase. Carbon atoms of ClbP, Pab87 peptidase, and AmpC β -lactamase are in green, yellow, and gray, respectively. The motifs SxxK, Yx(N/S/T), and (K/H)(S/T/G)G are numbered in red, blue, and green, respectively, and residue numbering is according to the ClbP sequence. C and D, phenotypic analysis of *pks* island activity in *E. coli* DH10B/*pBACpks* deleted or not for the *clbP* gene and trans-complemented with ClbP WT protein or the S95A, K98T, or Y186G active site ClbP mutants. HeLa cells were infected 4 h with 100 *E. coli* per cell, and then the cells were washed and incubated with gentamicin for 4–72 h. C, anti-5His Western blot analysis of the expression of His-tagged WT and active site ClbP mutants in *E. coli* DH10B/*pBACpks* Δ *clbP* after infection. D, host histone H2AX Ser-139 phosphorylation (γ H2AX) indicative of DNA double-strand breaks was assayed by confocal immunofluorescence 4 h after infection. DNA and γ H2AX are pseudocolored in blue and red, respectively (bars, 20 μ m). Cell swelling (megalocytosis) was observed following Giemsa staining 72 h after infection (bars, 50 μ m). G_2 cell cycle arrest following DNA damage was assessed by flow cytometric DNA content analysis, 48 h after infection.

open in ClbPpep than in S12 peptidases, which harbor one, two, or several bulky helices in this area. Of note, between Cys-337 and Cys-367, ClbP harbored a disulfide bond, which links b12 strand (residues 335–341) with the C-terminal helix h11. To our knowledge, this disulfide bond is not observed in other known structures of S12 enzymes. Overall, the ClbPpep structure was closely similar to that of S12 peptidases. However, ClbPpep had noticeable conformational differences in the active site, especially in the two extremities of the catalytic groove, which was unusually large (Fig. 1, B–D). This atypical shape suggests an adaptation of ClbP catalytic pocket to a specific substrate.

Catalytic Residues and Implications for Catalysis—Striking structural similarities were observed between the active site of ClbPpep, class C β -lactamases, and other S12 enzymes such as Pab87 peptidase (21, 28, 29), in particular, for the motif ⁹⁵SxxK and the Tyr-186 residue. The positioning of this last residue and the presence of His-327 make the ClbPpep active site very similar to that of Pab87 peptidase (Fig. 2, A and B). In addition, the location and geometry of the residues Ser-95, Lys-98, and Tyr-186 suggest a direct role in the catalytic mechanism mediated by ClbP (31, 32).

To confirm the importance of these residues in ClbP, substitutions S95A, K98T, and Y186G were introduced by site-directed mutagenesis residues into His-tagged ClbP. The resulting mutants were assessed for their ability to restore the cytopathic activity of the *pks*-positive *E. coli* Δ *clbP* isogenic mutant. Infected HeLa cells were analyzed for megalocytosis, G_2 /M cell cycle arrest, and phosphorylation of histone H2AX (3). None of these mutants were able to restore the cytopathic activity of the *pks*-positive *E. coli* Δ *clbP* isogenic mutant in contrast with the wild-type ClbP protein (Fig. 2, C and D). These results confirm the identification of the active site and the importance of ClbP main domain for the functionality of *pks* island. They also suggest that these enzymes and other S12 serine active-site enzymes share a catalytic mechanism.

Residues Surrounding Catalytic Residues and Substrate Binding—The motif YxN, which is observed in most S12 enzymes, was replaced by the ¹⁸⁶YAS motif in ClbP (Fig. 2B). Ser-188 in ClbP was hydrogen bonded to the Lys-98 residue of motif SxxK (Fig. 2A). In class C β -lactamases, the structural analog of Ser-188 is Asn-152, which interacts with both the substrates and Lys of the motif SxxK (19). His-257 of ClbP, which points at the vicinity of Ser-188, may also play a role in

substrate binding. Opposite the ¹⁸⁶YAS motif, the catalytic pocket of ClbP exhibits the HGG motif. The alignment of ClbP with class C β -lactamases revealed that HGG motif is a structural analog of KTG box (Fig. 2B), which is involved in substrate binding because of Lys and Thr (29). In β -lactamases, Lys can be replaced by the closely related residues His or Arg, and Thr may be replaced by Ser. A positively charged side chain followed by one bearing a hydroxyl group has therefore been suggested to be universally conserved. The second residue in ClbP motif is an atypical residue Gly-328 instead of the canonical residues Ser or Thr, which are usually directly involved in substrate recognition via a hydrogen bond (29). However, in the vicinity of Gly-328 (Figs. 1E and 2A), positions 330 and 331 of ClbP harbor the unusual residues Gln and Asn respectively, which point toward the catalytic pocket and may establish hydrogen bonds with substrates.

The residues Glu-159 and Phe-316 of ClbPpep are structural analogs of positions involved in the binding of the N-terminal and C-terminal parts of peptides in S12 amino- and carboxypeptidases, respectively. These positions determine the carboxypeptidase or aminopeptidase activity of S12 peptidases (30). Previous studies suggest that the residues Glu-159 and Phe-316 of ClbP (Fig. 2A) should favor aminopeptidase activity because the negatively charged residue Glu-159 may accommodate the positively charged N-terminal part of peptides, and the absence of positively charged residue in position 316 may alter the accommodation of the negatively charged C-terminal part of peptides.

Overall, the six positions 159, 188, 257, 316, 328, and 331 can therefore participate in substrate binding in ClbP because the side chain of these residues pointed toward the catalytic pocket (Fig. 2A). The hydrogen bond donors Ser-188 and His-257 of ClbP-His were replaced by Ala, which is not able to accept or give hydrogen bonds. The substitutions E159K and F316R were introduced into ClbP-His to modify the charges in the critical areas of the catalytic pocket. We introduced the substitution G328S to check whether a small residue is required in the second position of the motif HGG to accommodate ClbP substrate. The substitution C337A was also introduced into ClbP-His to check the importance of the disulfide bond Cys-337–Cys-367 for stability or folding. No substitution altered the trans-complementing activity of these ClbP variants in the *pks*-positive *E. coli* Δ *clbP* isogenic mutant. ClbP activity may have been only weakly modified by these substitutions. Unfortunately, efficient biochemical tests are not yet available to monitor ClbP activity. However, cellular tests are a good representation of the physiological context, and the results are evidence of substrate tolerance of ClbP.

ClbP Is a Peptidase and NRP-recognizing Protein—As ClbP presents a S12 enzyme scaffold and catalytic residue positioning close to Pab87 peptidase, peptidase activity was investigated using the reporters Gly-, D-Ala-, L-Ala-, and L-Ala-L-Ala-*p*-nitroanilides. The catalytic efficiency k_{cat}/K_m of ClbPpep against these substrates were 0.005, 0.018, 0.020, and 0.040 $\mu\text{M}^{-1} \text{min}^{-1}$, respectively. This peptidase activity is weak in comparison with that of enzymes really adapted to these substrates such as S12 aminopeptidases (30). However, in a ClbPpep mutant harboring the substitution Y186E, there was no activity,

which confirms the involvement of the ClbP binding site in peptidase activity.

Imipenem is a NRP-type compound belonging to the chemical group of β -lactams, which efficiently bind members of the S12 enzyme family, such as carboxypeptidase and AmpC β -lactamases (33, 34). ClbPpep incubated with imipenem was analyzed by mass spectrometry (supplemental Fig. 2). Imipenem induced a mass shift consistent with the formation of imipenem acyl-enzyme complex of ClbPpep, as observed for β -lactam-binding proteins (35). Despite a high concentration of imipenem (ratio imipenem/ClbPpep > 1/1000), part of ClbPpep exhibited a mass corresponding to the native protein. The results show that ClbP is able to bind NRP-type compounds, has a moderate β -lactam-recognizing activity, and is devoid of significant β -lactamase activity. The k_{cat} value ($<0.001 \mu\text{M}^{-1} \text{s}^{-1}$) was ~ 1 million-fold lower than typical values of AmpC enzymes (36). Consequently ClbP, in contrast to AmpC, was not able to increase the resistance level of *E. coli* to β -lactam antibiotics.

Identification of ClbP Homologs across Prokaryotic Lineages—We therefore screened public protein and genomic databases to identify proteins homologous to ClbP peptidase domain. 77 homologs were obtained including the protein ZmaM, which may be involved in the synthesis of the PK-NRP antibiotic zwittermicin A (ZmA) (37). These proteins only shared 29.3% (ZP_04059476.1) to 42.7% (ZP_04215168.1) identity and all belonged to the MEROPS S12 enzyme family. The amino acid sequences and the predicted topology of both ClbP and the homologs were compared with those of representative members of the S12 family that have been assigned to a biological function and/or a biochemical activity. The resulting phylogenetic tree and the topologies revealed that the proteins clustered in three major branches supported by high bootstrap values ($\geq 64\%$) (supplemental Fig. 3). One branch harbored the known representative members of the S12 family, which were subclustered according to their biochemical activity (β -lactamases, peptidases, esterases, and amidases). The second harbored the FmtA-like proteins ($n = 17$) of the *Staphylococcus* species (25) that were phylogenetically closer to ClbP than the other members of the S12 family and that exhibited a predicted topology identical to that of ClbP in contrast with the other S12 family members. The third branch exhibited the longest phylogenetic distance, which clearly defined a subgroup containing ClbP and most closely related proteins ($n = 60$). These proteins, designated ClbP-like, harbored a predicted topology identical to that of ClbP. Some ($n = 14$), such as ZmaM, possessed in addition a putative half-size ABC-type exporter domain in the C terminus. Except ClbP and its closest homolog (ClbP-like of *H. chejuensis*) produced by different γ -proteobacteria, all of the other ClbP-like were identified in taxons of the Firmicute phylum (*Bacillus* and *Clostridium* geni), which is the main bacterial source of PK, NRP, and hybrid PK-NRP compounds with the Actinobacteria phylum (38).

ClbP-like Proteins Display Functional Promiscuity—No clues for the presence of megasynthase catalytic unit-encoding genes were revealed in the vicinity of any *fntA*-like and 17 *clbP*-like genes. In contrast, putative NRPS- and PKS-encoding genes were observed in the vicinity of 18 *clbP*-like genes

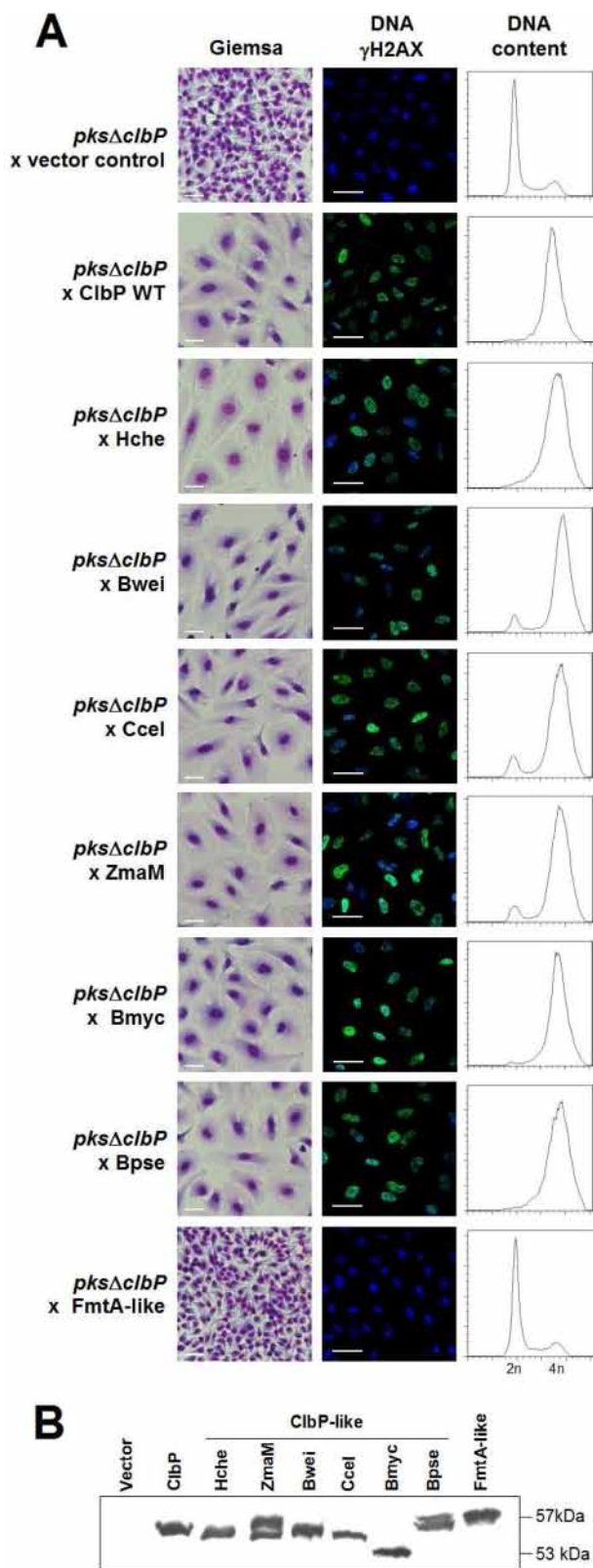


FIGURE 3. Functional promiscuity displayed by ClbP-like proteins. *A*, phenotypic analysis of *pks* island activity in *E. coli* strain DH10B/pBAC*pks* Δ *clbP* complemented with empty vector, ClbP WT, four ClbP-like proteins genetically linked to NRPS gene clusters from *H. chejuensis* (*Hche*), *B. weihenstephanensis* (*Bwei*), *C. cellulolyticum* (*Ccel*), and *B. cereus* (*ZmaM*), two ClbP-like proteins genetically unlinked to NRPS gene clusters from *B. mycoides* (*Bmyc*), and *B. pseudomycoides* (*Bpse*), or one FmtA-like protein from *S. aureus*. ~50% confluent HeLa cells were infected 4 h with 100 *E. coli* per cell. Left panels, the cell

(supplemental Fig. 4). Moreover, putative NRPS-encoding genes were present in all of these genomic islands unlike PKS-encoding genes, suggesting that the cognate products contain peptide moiety that may be related to the peptidase activity of ClbP-like proteins. To test whether these enzymes are functionally equivalent, we analyzed the ability of six representative ClbP-like proteins (originating from different phyla, different geni, and encoded by genes inserted or not in NRP biosynthesis clusters) and one FmtA-like to xeno-complement the cytopathic activity of the *pks*-positive *E. coli* Δ *clbP* isogenic mutant. Megalocytosis, G₂/M cell cycle arrest, and histone H2AX phosphorylation were restored by all ClbP-like proteins, whose gene is located in a NRPS-encoding genomic island ($n = 4$) or is not ($n = 2$) (Fig. 3). In contrast, the FmtA-like protein did not restore cytopathic activity despite its full expression (Fig. 3). The efficient xeno-complementations thus support the functional promiscuity of ClbP-like proteins. Moreover, these results are evidence of substrate tolerance of NRP-associated peptidases and confirm data obtained in experiments that have investigated residues of the ClbP catalytic pocket potentially involved in substrate binding.

DISCUSSION

The aim of this study was to identify the structure and function of the ClbP protein, which is essential to the genotoxic activity of the *pks* island. The crystal structure of the ClbP enzymatic domain together with mutagenesis experiments and biochemical assays revealed the same α/β fold observed in the members of the MEROPS S12 family and a serine-reactive active site conferring NRP-binding and peptidase activities necessary to colibactin bioactivity. ClbP presented an unusually wide and deep catalytic groove and an active site with original structural features. We extended investigations to ClbP-like proteins, which are the closest related proteins of ClbP, and are clustered on a distinct phylogenetic branch in the MEROPS S12 enzyme family. ClbP and some ClbP-like-encoding genes belong to NRPS-encoding gene clusters. The ClbP-like proteins displayed functional promiscuity and shared a particular cellular localization and topology according to *in silico* analyses, namely a large N-terminal extracytoplasmic peptidase domain anchored to the inner membrane by three C-terminal transmembrane helices. Hence, ClbP and the closest related ClbP-like proteins form a functional peptidase subgroup of S12 serine enzymes, which is probably involved in NRP or PK-NRP biosynthesis. We propose to designate this subgroup the “NRP-associated peptidases.”

Bioinformatics and biochemical analyses of the ZmA biosynthesis enzymes performed by Kevany *et al.* (37) suggest that the compound is initially biosynthesized as part of a large metabo-

morphology was observed following Giemsa staining at 72 h after infection (bars, 50 μ m). Middle panels, histone H2AX Ser-139 phosphorylation (γ H2AX) indicative of DNA double-strand breaks was assayed by confocal immunofluorescence at 20 h after infection. DNA and γ H2AX are pseudocolored in blue and green, respectively (bars, 50 μ m). Right panels, G₂ cell cycle arrest following DNA damage was assayed by flow cytometry 48 h after infection. *B*, anti-RGS6His Western blot analysis of the expression of RGS6His-tagged ClbP wild-type protein, ClbP-like proteins, and FmtA-like protein in *E. coli* DH10B/pBAC*pks* Δ *clbP*.

lite that is processed by a peptidase, resulting in the formation of ZmA. The only candidate enzyme encoded by the ZmA biosynthesis gene cluster to catalyze such a peptidase activity is ZmaM, a ClbP-like protein. This observation and our results suggest that NRP-associated peptidases can process NRP or PK-NRP compounds. ClbP encoded by the *pks* island and the other NRP-associated peptidases may process a procolibactin and other NRP compounds into biologically active molecules by cleaving a nonribosomal peptide bond. The ClbP-like peptidase domain of ZmaM had been suggested to cleave a “pre-ZmA” metabolite between the D-Asn and D-Ser, releasing ZmA and a fatty acyl D-Asn (37). Interestingly, *in silico* analysis of NRPS substrate specificity showed that all gene clusters producing NRP-associated peptidases encode NRPS able to insert the close related residues Asn and/or Gln into putative NRP compounds, suggesting that these residues may be involved in substrate recognition by NRP-associated peptidases. Of note, ClbP-like proteins were the only proteins encoded by these islands that harbored a predicted translocation signal sequence and a predicted periplasmic catalytic domain, in contrast to other enzymes involved in biosynthesis of PK, NRP, and mixed compounds that are cytoplasmic proteins. Accordingly, NRP-associated peptidases may act outside of the cytoplasm.

Several NRP-associated peptidases harbor in the C terminus a typical half-size ABC exporter domain such as ZmaM (supplemental Figs. 3 and 4). These ABC export systems are involved in the extrusion of noxious substances and the export of extracellular toxins (39). It has already been suggested that these ABC exporters transport NRP compounds (39). ClbP and other NRP-associated peptidases did not harbor such an export domain, but their cognate gene clusters encode a Multidrug and Toxic Compound Extrusion transporter (such as ClbM in the *pks* island) or other efflux pumps. After the translocation of an inactive NRP compound from cytoplasm to periplasm via efflux pumps, the periplasmic location of the NRP-associated peptidase domain may allow the maturation of the fully active toxin away from a potential cytoplasmic target of its producer cell. NRP-associated peptidases and efflux pumps can be seen therefore as a two-component resistance mechanism against the possible toxic activity generated by the cognate NRPS-encoding gene cluster.

NRP-associated peptidases genetically linked to NRPS gene clusters display a functional promiscuity with those genetically unlinked to NRPS gene clusters. These latter peptidases may also participate in the synthesis of a bioactive NRP compound because we found one to three putative NRP biosynthetic clusters in 15 of 16 genomes possessing a NRP-associated peptidase genetically unlinked to a NRPS gene cluster. It has already been suggested that enzymes required for NRP and PK biosynthesis such as phosphopantetheinyl transferases are encoded in a distinct region of the main biosynthetic gene cluster in bacterial genomes (40).

In conclusion, ClbP is a prototype of a new subfamily of extracytoplasmic serine-reactive peptidases probably acting as maturing enzymes for NRP compounds. The essential role of ClbP in the production of the biological effect make this protein a key target for the control of the bioactivity of the PK-NRP-producing *pks* gene cluster, which may affect commensalism

and/or pathogenicity and constitute a predisposing factor for the development of colorectal cancer (7). The engineering of NRP compounds should take into account these new catalytic units, which could be interesting tools for modifying the structures and bioactivity of NRP and PK-NRP compounds and be a target for identifying bioactive molecules.

Acknowledgments—We thank R. Perroux, M. Jan, and C. Habouzit for technical assistance, Alexei Sorokin (Institut National de Recherche Agronomique, Jouy en Josas, France), Michael G. Thomas (University of Wisconsin, Madison, WI), and Jihyun F. Kim (Korea Research Institute of Bioscience and Biotechnology, Daejeon, Republic of Korea) for providing the *B. weihenstephanensis* KBAB4 strain, the *B. cereus* UW85 strain, and a fosmid clone containing the *hch_03446* locus tag of *H. chejuensis* KCTC 2396^T, respectively. We also thank C. Blavignac and C. Sczepaniak (Centre d’Imagerie Cellulaire Sante, Université d’Auvergne, France) for technical assistance in flux cytometry. We also gratefully acknowledge the use of the experiment facilities at the European Synchrotron Radiation Facility (ESRF).

REFERENCES

- Li, J. W., and Vederas, J. C. (2009) *Science* **325**, 161–165
- Sattely, E. S., Fischbach, M. A., and Walsh, C. T. (2008) *Nat. Prod. Rep.* **25**, 757–793
- Nougayrède, J. P., Homburg, S., Taieb, F., Boury, M., Brzuszkiewicz, E., Gottschalk, G., Buchrieser, C., Hacker, J., Dobrindt, U., and Oswald, E. (2006) *Science* **313**, 848–851
- Putze, J., Hennequin, C., Nougayrède, J. P., Zhang, W., Homburg, S., Karch, H., Bringer, M. A., Fayolle, C., Carniel, E., Rabsch, W., Oelschlaeger, T. A., Oswald, E., Forestier, C., Hacker, J., and Dobrindt, U. (2009) *Infect. Immun.* **77**, 4696–4703
- Dubois, D., Delmas, J., Cady, A., Robin, F., Sivignon, A., Oswald, E., and Bonnet, R. (2010) *J. Clin. Microbiol.* **48**, 2122–2129
- Johnson, J. R., Johnston, B., Kuskowski, M. A., Nougayrède, J. P., and Oswald, E. (2008) *J. Clin. Microbiol.* **46**, 3906–3911
- Cuevas-Ramos, G., Petit, C. R., Marcq, I., Boury, M., Oswald, E., and Nougayrède, J. P. (2010) *Proc. Natl. Acad. Sci. U.S.A.* **107**, 11537–11542
- Schmitt, C. K., Darnell, S. C., Tesh, V. L., Stocker, B. A., and O’Brien, A. D. (1994) *J. Bacteriol.* **176**, 368–377
- Potterton, E., McNicholas, S., Krissinel, E., Cowtan, K., and Noble, M. (2002) *Acta Crystallogr. D. Biol. Crystallogr.* **58**, 1955–1957
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) *J. Appl. Crystallogr.* **40**, 658–674
- Steiner, R. A., Lebedev, A. A., and Murshudov, G. N. (2003) *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 2114–2124
- Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. D. Biol. Crystallogr.* **60**, 2126–2132
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–291
- Lovell, S. C., Davis, I. W., Arendall, W. B., 3rd, de Bakker, P. I., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) *Proteins* **50**, 437–450
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25**, 3389–3402
- Papadopoulos, J. S., and Agarwala, R. (2007) *Bioinformatics* **23**, 1073–1079
- Milne, I., Lindner, D., Bayer, M., Husmeier, D., McGuire, G., Marshall, D. F., and Wright, F. (2009) *Bioinformatics* **25**, 126–127
- Rawlings, N. D., Morton, F. R., Kok, C. Y., Kong, J., and Barrett, A. J. (2008) *Nucleic Acids Res.* **36**, D320–325
- Beadle, B. M., Trehan, I., Focia, P. J., and Shoichet, B. K. (2002) *Structure* **10**, 413–424
- Pratt, R. F. (2008) *Cell Mol. Life Sci.* **65**, 2138–2155
- Delfosse, V., Girard, E., Birck, C., Delmarcelle, M., Delarue, M., Poch, O., Schultz, P., and Mayer, C. (2009) *PLoS One* **4**, e4712

Structure and Functionality of ClbP Peptidase

22. Okazaki, S., Suzuki, A., Komeda, H., Asano, Y., and Yamane, T. (2008) *J. Synchrotron Radiat.* **15**, 250–253
23. Wagner, U. G., Petersen, E. I., Schwab, H., and Kratky, C. (2002) *Protein Sci.* **11**, 467–478
24. Bompard-Gilles, C., Remaut, H., Villeret, V., Prangé, T., Fanuel, L., Delmarcelle, M., Joris, B., Frère, J., and Van Beeumen, J. (2000) *Structure* **8**, 971–980
25. Komatsuzawa, H., Choi, G. H., Fujiwara, T., Huang, Y., Ohta, K., Sugai, M., and Suginaka, H. (2000) *FEMS Microbiol. Lett.* **188**, 35–39
26. Holm, L., and Sander, C. (1996) *Science* **273**, 595–603
27. Meroueh, S. O., Minasov, G., Lee, W., Shoichet, B. K., and Mobashery, S. (2003) *J. Am. Chem. Soc.* **125**, 9612–9618
28. Dodson, G., and Wlodawer, A. (1998) *Trends Biochem. Sci.* **23**, 347–352
29. Matagne, A., Lamotte-Brasseur, J., and Frère, J. M. (1998) *Biochem. J.* **330**, 581–598
30. Delmarcelle, M., Boursoit, M. C., Filée, P., Baurin, S. L., Frère, J. M., and Joris, B. (2005) *Protein Sci.* **14**, 2296–2303
31. Chen, Y., Minasov, G., Roth, T. A., Prati, F., and Shoichet, B. K. (2006) *J. Am. Chem. Soc.* **128**, 2970–2976
32. Galleni, M., Lamotte-Brasseur, J., Raquet, X., Dubus, A., Monnaie, D., Knox, J. R., and Frère, J. M. (1995) *Biochem. Pharmacol.* **49**, 1171–1178
33. Matagne, A., Ghuysen, M. F., and Frère, J. M. (1993) *Biochem. J.* **295**, 705–711
34. Beadle, B. M., and Shoichet, B. K. (2002) *Antimicrob. Agents Chemother.* **46**, 3978–3980
35. Drawz, S. M., Babic, M., Bethel, C. R., Taracila, M., Distler, A. M., Ori, C., Caselli, E., Prati, F., and Bonomo, R. A. (2010) *Biochemistry* **49**, 329–340
36. Bauvois, C., Ibuka, A. S., Celso, A., Alba, J., Ishii, Y., Frère, J. M., and Galleni, M. (2005) *Antimicrob. Agents Chemother.* **49**, 4240–4246
37. Kevany, B. M., Rasko, D. A., and Thomas, M. G. (2009) *Appl. Environ. Microbiol.* **75**, 1144–1155
38. Donadio, S., Monciardini, P., and Sosio, M. (2007) *Nat. Prod. Rep.* **24**, 1073–1109
39. Saurin, W., Hofnung, M., and Dassa, E. (1999) *J. Mol. Evol.* **48**, 22–41
40. Lu, Y. W., San Roman, A. K., and Gehring, A. M. (2008) *J. Bacteriol.* **190**, 6903–6908

Microbiology:

CibP Is a Prototype of a Peptidase Subgroup Involved in Biosynthesis of Nonribosomal Peptides

Damien Dubois, Olivier Baron, Antony Cougnoux, Julien Delmas, Nathalie Pradel, Michèle Boury, Bernadette Bouchon, Marie-Agnès Bringer, Jean-Philippe Nougayrède, Eric Oswald and Richard Bonnet
J. Biol. Chem. 2011, 286:35562-35570.
doi: 10.1074/jbc.M111.221960 originally published online July 27, 2011

MICROBIOLOGY

PROTEIN STRUCTURE
AND FOLDING

Access the most updated version of this article at doi: [10.1074/jbc.M111.221960](https://doi.org/10.1074/jbc.M111.221960)

Find articles, minireviews, Reflections and Classics on similar topics on the [JBC Affinity Sites](#).

Alerts:

- [When this article is cited](#)
- [When a correction for this article is posted](#)

[Click here](#) to choose from all of JBC's e-mail alerts

Supplemental material:

<http://www.jbc.org/content/suppl/2011/07/27/M111.221960.DC1.html>

This article cites 40 references, 17 of which can be accessed free at <http://www.jbc.org/content/286/41/35562.full.html#ref-list-1>