Cpl-7, a Lysozyme Encoded by a Pneumococcal Bacteriophage with a Novel Cell Wall-Binding Motif

Noemí Bustamante‡§, Nuria E. Campillo¶, Ernesto García§, Cristina Gallego‡§, Benet Pera‖, Gregory P. Diakun**, José Luis Sáiz‡§, Pedro García§‖, J. Fernando Díaz‖, and Margarita Menéndez‡§1

From the ‡ Instituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006, Madrid, Spain, the § CIBER de Enfermedades Respiratorias (CIBERES), the ¶ Instituto de Química Médica, Consejo Superior de Investigaciones Científicas, Juan de la Cierva 3, 28006, Madrid, Spain, the ‖ Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Ramiro de Maeztu 9, 28040, Madrid, Spain, and the ** CCLRC Daresbury Laboratory, Warrington WA4 4AD, United Kingdom

Running head: Cpl-7 pneumococcal phage endolysin

1 To whom correspondence should be addressed: Instituto de Química-Física Rocasolano, Consejo Superior de Investigaciones Científicas, Serrano 119, 28006, Madrid, Spain. Tel.: 3491-561-9400; Fax: 3491-564-2431; E-mail: mmenendez@iqfr.csic.es.

Bacteriophage endolysins comprise a group of new antibacterials refractory to development of resistance. We present here the first structural study of the Cpl-7 endolysin, encoded by pneumococcal bacteriophage Cp-7. It contains an N-terminal catalytic module (CM) belonging to the GH25 family of glycosyl hydrolases, and a C-terminal region comprising three identical repeats of 42 amino acids (CW_7 repeats), unrelated to choline-targeting motifs present in other cell wall hydrolases produced by Streptococcus pneumoniae and its bacteriophages, that account for cell wall attachment. By combining different biophysical techniques and molecular modeling, a three-dimensional model of the overall protein structure is proposed, consistent with circular dichroism and sequence-based secondary structure prediction, small angle X-ray scattering data, and Cpl-7 hydrodynamic behavior. Cpl-7 is an approximately 115 Å-long molecule with two well differentiated regions, corresponding to the CM and the cell wall-binding region (CWBR), arranged in a lateral disposition. The CM displays the (βα)5β barrel topology characteristic of the GH25 family, and the impact of sequence differences with the CM of the Cpl-1 lysozyme in substrate binding is discussed. The CWBR is organized in three tandemly-assembled three-helical bundles whose dispositions remind a super-helical structure. Its approximate dimensions are 60 × 20 × 20 Å and presents a concave face that might constitute the functional region involved in bacterial surface recognition. The distribution of CW_7 repeats in the sequences deposited in the Entrez Database has been examined, and the results drastically expanded the antimicrobial potential of the Cpl-7 endolysin.

Bacteria and bacteriophages produce a variety of enzymes that cleave the peptidoglycan, the cell wall hydrolases (CWHs)1, either to lyse host cells or to re-model the cell wall during growth and division. Many of these enzymes have a modular structure and are composed of catalytic units linked to a number of other modules. The Cpl-7 lysozyme, the lytic enzyme (endolysin) encoded by the pneumococcal bacteriophage Cp-7, consists of a catalytic N-terminal domain fused to a C-terminal region containing 3 identical tandem repeats of 48 amino acids (the third motif has the last six residues missing), which is essential for activity and unique among CWHs encoded by the pneumococcus and its bacteriophages (1, 2). However, the same type of repeats (termed CW_7 hereafter) has been later identified in other proteins putatively associated with
bacterial cell wall degradation activity (3). The catalytic module (CM) of Cpl-7 belongs to the GH25 family of glycosyl hydrolases and shows 85.6% sequence identity with the CM of the Cpl-1 lysozyme, the endolysin encoded by the pneumococcal bacteriophage Cp-1 (4). However, the cell wall-binding region (CWBR) of Cpl-1, made of six tandem copies of \( \overline{20} \) amino-acid residues plus a terminal tail, belongs to the family of pneumococcal choline-binding modules (5). Its lytic activity is thus conditioned by the presence of choline moieties in pneumococcal teichoic and lipoteichoic acids and becomes inhibited at high choline concentrations (IC\(_{50} = 2\) mM) (1, 6). In contrast, Cpl-7 degrades pneumococcal cell walls containing either choline or ethanolamine and retains full activity even in presence of high concentrations of these aminoalcohols (7). This strongly indicates that the C-terminal modules of Cpl-7 and Cpl-1 are responsible for cell wall targeting but recognize different structural elements of the bacterial envelope. Indeed, the loss of the CWBR by controlled proteolytic digestion reduces four orders of magnitude the activity of Cpl-7 CM against pneumococcal cell walls containing either choline or ethanolamine (8). The reduced activity of this truncated form is comparable to those reported for the isolated CM of Cpl-1 against both types of substrates and for the full-length form of Cpl-1 when assayed against ethanolamine-containing cell walls, and reflects the intrinsic activity of Cpl-7 and Cpl-1 CMs without cooperation of their cell wall binding motifs. The relationship between the CW_7 repeats and the substrate specificity displayed by the Cpl-7 lysozyme was clearly demonstrated by construction of a chimeric protein, LC7, comprising the CM of the LytA amidase (the major pneumococcal autolysin and a choline-dependent enzyme) fused to the C-terminal region of Cpl-7 (7). This chimeric enzyme exhibited an amidase activity capable of degrading ethanolamine-containing cell walls not inhibited by choline. Furthermore, the CWBR also determined the optimal pH for the catalytic activity of LC7 and the parental enzymes. Interestingly, the acquisition of one or two CW_7-like motifs by the endopeptidase module of the \( \lambda \)Sa2 endolysin, a bifunctional enzyme containing endopeptidase and \( \beta \)-D-N-acetylglucosaminidase activities, not only increased the activity on several streptococcal strains but were shown to be essential for its activity on staphylococci (3).

The crystallographic structure of Cpl-1, both in its free-state and bound to muropeptide analogues, has been reported (9, 10), and its solution behavior characterized both in the absence and in the presence of choline (11, 12). In contrast, although the CM of Cpl-7 has been recently crystallized (13), there is no available information about the overall structure of the enzyme or its CWBR.

Because of their unique ability to cleave peptidoglycan in a generally species-specific manner, endolysins represent a novel class of antibacterial agents (enzybiotics) and provide a mean of selective and rapid killing of pathogenic bacteria, refractory to resistance development (14, 15). Indeed, the endolysins encoded by the lytic pneumococcal bacteriophages Dp-1 and Cp-1 have been shown to efficiently eradicate pneumococci without affecting the normal microbiota (16–22), a feature related with their targeting of choline moieties.

The presence in the Cpl-7 endolysin of a cell wall-binding motif that may recognize a distinct element of the cell wall opens new possibilities to the potential use of both the wild-type enzyme and chimeric constructions bearing CW_7 repeats as efficient enzybiotics against \( S.\ pneumoniae \) and other Gram-positive bacterial pathogens. This paper reports the first structural characterization of the Cpl-7 lysozyme using low resolution experimental techniques, i.e., circular dichroism (CD), analytical ultracentrifugation and small angle X-ray scattering (SAXS), and three-dimensional modeling. Cpl-7 behaves as a monomer in solution and its two modules are arranged forming an elongated particle of \( \approx 115\) Å long. The CM has the \((\beta_3)\beta_3\) barrel characteristic of the GH25 family whereas the CWBR would comprise three tandemly-assembled, three-helical bundles whose disposition reminds a super-helical structure. The high resolution models proposed here for the isolated modules together with the SAXS-derived model for the overall protein structure provide clues to facilitate the identification of residues and surfaces important for Cpl-7 attachment to the cell wall. Finally, the three-helical bundle formed by the CW_7 repeat seems to be conserved in the related motifs that have been identified, in variable number and position, linked to a variety of catalytic units and cell wall-binding motifs in many putative proteins potentially involved in cell-wall metabolism.

**EXPERIMENTAL PROCEDURES**

*Materials—Plasmid pCP700 overexpressing*
the Cpl-7 lysozyme (7) was constructed by cloning a 2.5 kb BamHI–BclI fragment of pCP70 containing the cpl-7 gene (1) into BamHI-digested pN-III([pp]P)-A3 (23), and Cpl-7 lysozyme (38,438 Da; Entrez code, AAA72844) was purified from Escherichia coli RB791 cells harboring pCP70. E. coli cultures were grown at 37 °C in LB medium containing ampicillin (100 µg ml⁻¹) to an optical density at 600 nm (OD₆₀₀) of ≥ 0.6. Then, lactose was added (2% w/v final concentration) and incubation continued overnight. Cells were harvested at 4 °C by centrifugation in an SLA-300 rotor (Sorvall) for 20 min at 8,000 rpm, resuspended in 20 mM sodium phosphate buffer, pH 6.9 (Pi buffer), and disrupted in a French Pressure Cell Press. The insoluble fraction was separated by centrifugation (13,000 rpm in an SS-34 rotor (Sorvall) for 20 min), and the supernatant was treated with streptomycin sulfate (0.8% w/v final concentration) to precipitate nucleic acids. The soluble fraction was separated by centrifugation and subjected at a fractioning at increasing concentrations of ammonium sulfate. Proteins precipitated between 20% and 35% (w/v) were resuspended and exhaustively dialyzed against Pi buffer (24 hours, 5 °C) to remove minor contaminations by nucleic acids. The protein was eluted (0.5 ml min⁻¹) iso-

cratically with Pi buffer and the collected fraction was separated by centrifugation (13,000 rpm in an SS-34 rotor (Sorvall) for 20 min at 8,000 rpm, resuspended in 20 mM sodium phosphate buffer, pH 6.9 (Pi buffer), and disrupted in a French Pressure Cell Press. The insoluble fraction was separated by centrifugation (13,000 rpm in an SS-34 rotor (Sorvall) for 20 min), and the supernatant was treated with streptomycin sulfate (0.8% w/v final concentration) to precipitate nucleic acids. The soluble fraction was separated by centrifugation and subjected to a fractioning at increasing concentrations of ammonium sulfate. Proteins precipitated between 20% and 35% (w/v) of salt saturation were resuspended and exhaustively dialyzed against Pi buffer (24 hours, 5 × 500 ml), and then loaded onto a DEAE-cellulose column (15 cm × 2 cm) at 0.5 ml min⁻¹ flow rate. The column was washed with Pi buffer until the initial baseline (OD₂₈₀) was recovered. A discontinuous NaCl gradient was then applied using a Pharmacia LKB fast-protein liquid chromatograph. After a first gradient from 0 to 0.3 M of salt (60 min), followed by washing with 0.3 M NaCl until base-line recovering, the protein was eluted with a gradient from 0.3 to 0.5 M NaCl (60 min) followed by 30 min washing with 1 M NaCl. Cpl-7 eluted around 0.46 M salt and the fractions collected showed a single band in SDS-PAGE (12% acrylamide/bis-acrylamide) whose mobility agreed with the protein theoretical mass. Fractions with OD₂₈₀/OD₂₆₀ ratios ≥ 1.8 were pooled, concentrated, and dialyzed against Pi buffer before being stored at −20°C. Those with lower absorbance ratios were subjected to an additional gel-filtration chromatography step in a Superose™ 12 column (Pharmacia Biotech; 30 × 1.5 cm) to remove minor contaminations by nucleic acids. The protein was eluted (0.5 ml min⁻¹) isocratically with Pi buffer and the collected fractions were stored as above. All purification steps were performed at 4 °C and reagents (Sigma Chemical Co.) were of analytical grade. Protein activity was tested using [methyl-²H]choline-

labeled pneumococcal cell walls (7).

Cpl-7 concentration was measured spectrophotometrically using a molar absorption coefficient at 280 nm of 65,700 ± 700 M⁻¹ cm⁻¹, determined according to Gill and von Hippel (24). The N-terminal of Cpl-7 was sequenced using an automatic sequencer (model 477A from Applied Biosystems) as described (25).

Circular dichroism—CD spectra (average of 4 scans) were recorded in the far-UV region using a JASCO-J810 spectropolarimeter (Jasco Corp.) equipped with a Peltier type cell holder. Measurements were performed using a scan rate of 20 nm min⁻¹, a response time of 4 s, a bandwidth of 1 nm and protein concentrations of 5.2 µM and 26 µM (1-mm and 0.2-mm path length cells, respectively). Buffer contribution was subtracted from the experimental data and the corrected ellipticity converted to mean residue ellipticity using an average molecular mass per residue of 112.39. Secondary structure (SS) content was estimated by deconvolution of the experimental curves with CONTIN (26), SELCON (27) and CDNN (28) programs using reference data sets with seventeen (CONTIN and SELCOM) and thirteen (CDNN) protein spectra. They evaluate four conformations (α-helix, β-sheet, β-turn and remainder) but CONTIN does not discriminate between parallel and antiparallel β-sheet.

Secondary-structure prediction—Prediction of SS was performed using Prof (29), PSIPRED (30) and Jpred 3 (31), all of them accessible via the ExPASy website (http://www.expasy.ch/). Final prediction data include only those residues whose secondary structure was predicted with reliability above 86% by at least two of the three methods.

Analytical ultracentrifugation—Sedimentation velocity measurements were carried out using an Optima XL-A (absorption optics) and an Optima XL-I (interference) analytical ultracentrifuges (Beckman Coulter) for detection of concentration gradients. The experiments were performed in Pi buffer at 45,000 rpm, with 400 µl of protein solution in each cell, using double sector Epon-charcoal centerpieces. Differential sedimentation coefficients, c(s), were calculated by least-squares boundary modeling of sedimentation velocity data using the program SEDFIT (32). The SEDNTERP program (33) was used to calculate standard sedimentation coefficients, s₂₀,w, from the experimental values, the effective Stockes radius (Rₑ), and the frictional ratio (f/f₀) assuming an specific hydration grade of 0.404 g of water per gram of protein. A partial specific volume
of 0.722 ml g⁻¹ at 20 °C was calculated from the amino acid composition (33). Equilibrium sedimentation experiments were performed at 20 °C by centrifugation of 80 μl samples at 10,000, 12,000 and 17,000 rpm in the Optima XL-A ultracentrifuge. Conservation of mass in the cell was checked in all the experiments. Data were analyzed following previously reported protocols (34). The theoretical value of the sedimentation coefficient of ab initio and high resolution models proposed for the overall structure were calculated using HYDRO and HYDROPRO (35, 36).

Three-dimensional modeling of Cpl-7 modules—A search for homologues of Cpl-7 modules in the non-redundant protein sequence database at NCBI using PSI-BLAST (37) returned the catalytic module of the Cpl-1 lysozyme (PDB code: 1h09) as the best matching structure for the CM of Cpl-7. However no sequences related to the CWBR with known structure were found. Protein-fold recognition methods were, therefore, applied using the FUGUE (38), 3D-PSSM (39), and SP4 (40) servers, taking as searching target either the sequence of the whole module or a single CW_7 repeat. Only the sequences identified as the best hits by at least two searching methods showing some relevant residue conservation and a good correlation with the SS of the CWBR were considered as possible templates for the structural model.

Sequence alignments were performed using CLUSTAL_X (41) and corrected, when necessary, by manual adjustment with SEAVIEW (42) using as a guide the predicted SS of the target. The structural models were built automatically using the SWISS-MODEL server (http://swissmodel.expasy.org) in the alignment mode (43) and their quality was evaluated with PROCHECK (44), VERIFY3D (45), QMEAN (46), and ERRAT (47). Several cycles of model building, realignment and validation were repeated for the CWBR until no further optimization of the structure was achieved.

The best structural models of each module were inspected visually to identify atomic clashes that were relieved by manually twisting the amino acid side-chains using the Sybyl software (Sybyl7.2, Tripos Inc.). After addition of all hydrogen atoms, the energy of the optimized structures was minimized with the AMBER force field implemented in the Sybyl package, using 3000 steps of the conjugate gradient minimization, and the quality of the final models was evaluated again as indicated above.

SAXS measurements—Data collection was performed at the 2.1 station of Daresbury’s synchrotron. The camera was set to cover ranges of the scattering vector (defined as reciprocal Bragg spacing, i.e., 2 sin(θ)/λ) from, approximately, 0.001 to 0.060 Å⁻¹ and data collected ranged from 0.045 to 0.05 Å⁻¹. Absolute values of the scattering vector were obtained by reference to the orders of the 67 nm repeat in wet rat tail collagen. The temperature of the samples was set at 4 °C and the X-ray scattering profiles were recorded in 60 time frames of 30 s. Prior to SAXS data collection, Cpl-7 samples were centrifuged for 20 min at 50,000 rpm and 4 °C in a Beckman TL-100 (TL100.3 rotor) and the lowest part of the protein solution was discarded to remove possible aggregates.

Data processing was performed using the software package provided by the Collaborative Computational Project for Fiber Diffraction and Solution Scattering (http://www.ccp13.ac.uk/). Raw data were normalized by beam intensity and detector response before being processed. Time frames showing radiation damage were removed before averaging.

The Guinier plot of log(I(s)) against s², where I(s) is the scattered intensity, was used to check the quality of the scattering data and the presence or absence of aggregates or interparticle interferences in solution (48). A first estimation of Rg and the forward scattering intensity (Io) was obtained from the slope and the y-intercept in the range of s values where the Guinier plot fits to a straight line (48). The radius of gyration (Rg), Io, and the pair-distance distribution function (p(r)) were then calculated from the experimental scattering data using the package GNOM (49). The maximum dimension of the particle, Dmax, was determined empirically by examining the qualities of the fits of p(r) transform to the experimental data for a given range of Dmax values. Ab initio bead models of Cpl-7 were constructed with DALAI_GA (50) using SAXS data. Typically, 10 models were generated, superimposed and aligned in pairs with SUPCOMB (51). The models with lowest average spatial discrepancy were considered to be the most probable, while the most divergent ones were considered as outliers. The selected aligned structures were averaged (DAMAVER) and filtered (DAMFILT) using as cut-off in DAMFILT the 70% of its default value (52).

The molecular envelope of the final model was obtained using the Sitos package (53) and the structural models built for the isolated modules were manually docked in this envelope us-
ing the USCF-Chimera program (54). The scattering profile and the radius of gyration of Cpl-7 model based in the docked structures were calculated with CRYOSOL (55).

RESULTS

**CW_7 repeats in Nature**—The Cpl-7 lysozyme is the only member known of the pneumococcal CWH family whose activity does not require the presence of phosphocholine residues in teichoic and lipoteichoic acids associated to the bacterial cell wall (2). Instead of the choline-binding repeats that characterize pneumococcal CWHs, the CBWR of Cpl-7 comprises 3 identical repeats of 48 amino acids although in the third one the last six residues are missing (the CW_7 repeats), found for the first time in this lysozyme (1). However, a search of the Entrez Database repeats), found for the first time in this lysozyme one the last six residues are missing (the CW_7 motifs (PF08230) has significant similarity with 39-42 amino acid long segments present, in variable number (1 to 3) and position (7 major architectures), in 67 different proteins whose other modules suggest their implication in cell wall metabolism in most cases (last date accessed, 8 April 2010) (Fig. 1). Most bacterial species (30 out of 46) belong to the Firmicutes phylum and 21 of them to the order Clostridiales (supplemental Table S1). The majority of species represent normal inhabitants of the microbiota of human gastrointestinal tract (including 4 members of the phylum Bacteroidetes) (56-58), although some of them may cause different diseases including endocarditis (e.g., Corynebacterium jeikeium and Granulicatella elegans), and oral (e.g., Bifidobacterium dentium and Parvimonas micra) and skin infections (e.g., Propionibacterium acnes and Streptococcus dysgalactiae subsp. equisimilis). However, some beneficial bacterial species such as certain bifidobacteria also encode CW_7-containing proteins. Besides, ‘Dehalococcoides ethenogenes’ (phylum Chloroflexi) and Ethanoligenens harbinense (phylum Firmicutes) were isolated from contaminated waters (supplemental Table S1). Nine proteins containing CW_7 motifs are encoded by prophages (or phage remnants) (supplemental Table S1).

The CW_7 repeats combine with lysozyme modules (Glyco_hydro_25; PF01183), as in Cpl-7, as well as with other modules such as N-acetylmuramyl-L-alanine amidase (NAM-amidase)—Amidase_2 (PF01510) or Amidase_5 (PF05382)—N-acetylglucosaminidase (GlcNAcase) (Glucoaminidase; PF01832), transglycosylase (Transglycosylase; PF05036), or cysteine, histidine-dependent amidohydrolases/peptidases (CHAP; PF05257) domains. Eleven proteins contain different combinations of CMs (Fig. 1). Besides, the multiple alignment of a representative sample of the amino acid sequences of CW_7 repeats showed that sequence conservation expands the first 42 residues of the CW_7 motifs of Cpl-7 (Fig. 2). This indicates that the six residues missing in the last repeat of Cpl-7 correspond, in fact, to a linker connecting the CW_7 motifs. Another finding of potential interest was that there are two subfamilies of CW_7 motifs differing by 3 amino acid residues near the middle of the repeat (Fig. 2). With only three exceptions (i.e., EEA90700 from Collinsella stercoreis, EEG 89729 from Coprococcus comes, and EEI88064 from Mobiluncus curtisii), both types of repeats do not coexist in the same protein.

**Cpl-7 purification**—The Cpl-7 lysozyme (predicted molecular mass of 38,438 Da) was purified from extracts of *E. coli* RB791[pCP700] cells by ion-exchange chromatography in DEAE-cellulose of a crude extract of proteins precipitated between 20–35% ammonium sulfate, followed by size-exclusion chromatography on a dextran-agarose column (Superose™ 12, Pharmacia Biotech). Typically, up to 35 mg of electrothroetically pure Cpl-7 were obtained from 2 liters of culture. N-terminal sequence determination showed that Met1 was removed under the over-expression and purification conditions used. The specific activity of the purified protein on [3H]choline-labeled pneumococcal cell wall was 1.1×10^5 U mg^-1 under optimal conditions (25 mM sodium acetate, pH 5.5, 37 °C), which was slightly lower than the value reported for the Cpl-1 lysozyme at equivalent conditions (6).

**Secondary structure composition of the Cpl-7 lysozyme**—The SS of Cpl-7 was experimentally evaluated using CD. The far-UV region of the CD spectrum presents a maximum at 195 nm and two minima at 209 and 220 nm, characteristic of proteins with a high α-helical content (Fig. 3A). SS composition was calculated by deconvolution of the spectrum using different programs with
similar results. The average content was estimated to be 36.8 ± 0.4% α-helix, 17 ± 1% β-sheet (7% antiparallel and 10% parallel), 18 ± 1% % turns and 28 ± 3% non-periodic structure.

The SS of the catalytic module was independently estimated using the crystallographic structure of Cpl-1 (PDB code: 1h09) as the CMS of both lysozymes show 85.6% identity (Fig. 4 A). Its contribution to the whole protein structure was thus estimated to be 14% for β-strands and 16% for α-helices. According to these values and taking into account that the CD estimation for total β-strands was ± 17%, the CWBR would display an all-α-helical fold, contributing with around 21% to the total α-helical content of Cpl-7. This value agrees fairly well with the predictions obtained by computational methods. According to them, CWBR residues predicted to be in α-helical conformation represented around 25% of the total sequence, and each repeat would fold into a three α-helical motif containing two clearly amphipatic helices (Fig. 2 and supplemental Fig. S1).

Association state and hydrodynamic behavior—Cpl-7 was subjected to sedimentation velocity experiments to analyze the homogeneity of the protein solutions, its association state and the particle hydrodynamics. The distribution of sedimentation coefficients measured in Pi buffer at 20 °C, using protein concentrations from 2.5 to 16.8 μM, showed that Cpl-7 sediments as a single species with a $S_{20,w}^0$ value of 2.73 ± 0.02 S (Fig. 3B). The same behavior was observed at 4 °C (data not shown) using protein concentrations one order of magnitude higher. The molecular weight of the sedimenting species was determined by equilibrium sedimentation experiments. As expected, the equilibrium profiles can be fitted to a single species model (Fig. 3C) with an average molecular mass, $M_{w,app}$, of 37 ± 2 kDa (average of 10 experiments), which corresponds to the monomer molecular mass. The effective Stockes radius ($R_s$ = 34.6 ± 0.4 Å) and the translational frictional coefficient ($f/f_0 = 1.52 ± 0.2$) were calculated from the experimental data using the SEDNTERP program. The high deviation of $f/f_0$ from the value expected for globular particles ($f/f_0 ≈ 1$) indicated that Cpl-7 has a highly elongated shape.

Modeling of Cpl-7 catalytic module. Protein-substrate interactions—The catalytic barrel of Cpl-7 was modeled by homology using as template the CM of Cpl-1 lysozyme. Figure 4A shows the sequence alignment together with the distribution of SS elements in Cpl-1 structure. The final model (Fig. 4B) displays a good geometry as defined by PROCHECK (98.7% of the residues fall in the most-favored regions of the Ramachandran plot), and the energetic evaluation showed comparable scores for the template and the model (Table 1). According to them, the proposed model would represent an excellent approximation of the real structure, as also indicated the full superimposition of the catalytic residues and the rmsd values (0.07 Å for Cα atoms and 0.43 Å for all atoms). As shown in Figures 4A and C, non-conservative substitutions expand primarily along the region comprised between the C-terminus of the loop connecting strand-3 to helix-3 and the beginning of helix-4, including the loop containing the proton donor Glu94. The differences in amino acid composition modify the electrostatic potential of the module surface (data not shown), particularly at the barrel edge comprising helices 3 and 4 and also at a locus close to the catalytic cleft (positions 63 and 96) where the catalytic pair Asp92-Glu94 is located. Besides, the substitution of Arg63 by cysteine in Cpl-7 implies the loss of hydrogen bonding to the side chains of Asp37, Asp95 and His96 (changed to aspartic in Cpl-1), which could also modify the binding locus for the stem peptide of MurNAc unit bound at position −1 (10).

Figure 4C shows the structural superimposition of the CM model with the coordinates of the complex formed by Cpl-1 and the muropeptide (2S5P)$_3$ (10), an hexasaccharide tri-pentapeptide that expands from positions −2 to +3 of the active site (saccharide units flanking the scissile glycoside bond are assigned as positions −1 and +1). The positions of non-conserved residues suggest that sequence differences between the CMSs of Cpl-1 and Cpl-7 would affect substrate recognition by the active site. Major changes would arise from substitutions at positions 12 and 96. The change of Ser12 by alanine would prevent the H-bond formed between the serine oxygen of Cpl-1 and the lactil group of MurNAc at position −1 (10). Besides, the substitution of Asp96 by histidine also implies the loss of the H-bonds discerned in Cpl-1 between the carboxylate oxygens of Asp96 and the Lys moiety of the peptide bond to the same MurNAc. More importantly, it also introduces unfavorable electrostatic interactions with the Lys amino group at neutral and acidic pHs. Finally, the structural changes of the CM derived from the change in the pattern of hydrogen bonds mediated by the residue at position 63 might also affect substrate binding, as in-
peptidic chains cannot be completely ruled out. However, the possibility that this region could interact with adjacent glycopeptidic chains cannot be completely ruled out.

**Modeling of Cpl-7 cell wall-binding region**—The PSI-Blast search of databases for three-dimensional structures homologous to the CWBR did not identify any protein to be used as possible template. Protein fold recognition methods were therefore applied using the sequences of either the whole module or a single CW_7 repeat in the search for remotely related structures. The highest-scoring results returned as possible templates for the complete module by more than one method are shown in Table 2. RecX (PDB code 3c1d), the first hit of FUGUE and SP4 servers is a modular protein comprising 3 tandem repeats of a three-helix motif (≥ 50 amino acids each) that can be superimposed despite their low sequence similarity (59). The similarity between RecX and the CWBR was near 46% (≥ 14% sequence identity), and the alignments returned by FUGUE and SP4 servers showed a good correlation between the distribution of residues involved in α-helices in RecX structure and Cpl-7 SS prediction. In contrast, the structure of 1ng6, identified by the three servers with much lower scores, consists of two α-helical bundles comprising four- and three-helical motifs, respectively, in tandem disposition, and its compatibility with the secondary structure predicted for the N-terminal half of the CWBR was clearly worst (data not shown). On the other hand, 4 proteins were identified as possible templates by at least two servers among the ten top hits when the sequence of a single CW_7 repeat was used in the search for related structures (Table 2). All of them were α-helical proteins and the two first-ranked candidates (PDB codes 1ixs and 1gab) display a three-helical bundle topology. Since the confidence levels of the highest-scoring result (1ixs) were in the order of those obtained for the best full-length candidate (3c1d), and the use of a single repeat for modeling the full CWBR structure would entail the additional difficulty of finding the correct orientation of CW_7 repeats within the full structure, the crystal structure of 3c1d (RecX protein) was finally selected as the best possible template. After several steps of alignment-modeling-evaluation, two promising models were finally obtained using as inputs the alignments shown in Figure 5A, where residues conserved in the RecX family sequences (59) present in the CWBR were also indicated. The two models slightly differ in the length of the α-helices comprised in the last repeat (Fig. 5A) and their superimposition with the RecX structure is shown in Figure 5B. Their evaluation was also similar, showing good stereochemical and energetical values (Table 1) and rmsd for Cα atoms of 3.15 Å (model M1) and 4.83 Å (model M2). For these reasons and taking into account that modeling of the CWBR was based on threading methods, models M1 and M2 will be considered as a single one for discussion. According to them, the CWBR would display an elongated structure with approximated overall dimensions of 60 × 20 × 20 Å and a moderate degree of curvature along the long axis. The side chains of conserved hydrophobic residues (Fig. 2) are tightly packed into the bundle hydrophobic core (Fig. 6A), and their conservation seems to be related to the structural role played in the bundle folding. Besides, conserved glycines will facilitate the sharp turns between the helices. On the other hand, the fully conserved arginine (Fig. 2), deeply buried at the inside, might contribute to stabilize the bundle structure by making hydrogen bonds with other residues, as could also do conserved residues of glutamine and asparagines located at positions 15, 34 and 38 of each repeat (Fig. 2).

Inspection of the model surface showed clear patterns in polarity and charge distribution (Fig. 6B). The convex face is crossed by polar patches due to the solvent-exposed surface of the third helix of each bundle, where neutral and charged residues alternate, and amino acid conservation is rather low. In contrast, fully-conserved arginines extend, deep buried, along the central region of the concave face, whose edges are lined by conserved aromatic residues and acidic groups (Fig. 6B and C). Interestingly, exposed regions involved in functional interactions show a higher degree of amino acid conservation than otherwise expected for surface residues. In addition both aromatic and polar residues with planar side chains (Asp, Glu, Asn, Gln, Lys and Arg) are a common signature of glycan-binding sites in carbohydrate recognition motifs (60). Therefore, the presence of both features in the CWBR concave face suggests that it might represent the functional surface involved in the attachment of Cpl-7 to the cell wall.

**Modular organization: SAXS modeling of the overall Cpl-7 structure**—SAXS is a powerful methodology to investigate domain organization...
in modular proteins, particularly when structural models of the components can be used to analyze the scattering data and to assess the most probable solution (48, 61-62). The modular organization of Cpl-7 in solution was thus investigated using synchrotron X-ray scattering data (Fig. 7) using a protein concentration of 187 μM. First, the quality of the scattering data and the presence or absence of aggregates or interparticle interference was checked from the Guinier plot. The linearity of \( \log(I(s)) \) against \( s^2 \) at low \( s \) values (supplemental Fig. S2A) indicated the absence of significant aggregates or interparticle effects in solution. The values of \( R_g(29.8 \pm 0.9 \, \text{Å}) \) and \( I_0 \) were thus obtained using the Guinier approximation. The analysis of the experimental curves with GNOM, that uses the entire scattering profile and which therefore can be more precise, provided an estimation of 30.3 Å for \( R_{ex} \), in good agreement with the Guinier value, and a maximum intraparticle distance, \( D_{max} \), of 115 Å (Fig. 7A). A good correlation between the scattering profile and the Fourier transform of \( P(r) \) was also observed (supplemental Fig. S2C). These results confirm that Cpl-7 is indeed a rather elongated molecule, as suggested the sedimentation velocity results. Besides, the behavior of the \( P(r) \) function at extended \( D_{max} \) (supplemental Fig. S2B) further supported the quality of the scattering profile and the absence of severe interferences from aggregation or interparticle interference (48). The pair-distance distribution function presents three well defined maxima at 22, 54, 74 and 98 Å (Fig. 7A, inset), indicating the presence of distinct structural entities within the molecule, as could be expected from the models derived for the isolated modules. Low resolution bead-models of Cpl-7 were reconstructed ab initio using DALI GA and the most probable one is shown in Figure 7B. The theoretical scattering profile of the model overlaps the experimental one (Fig. 7A, green line) and its \( R_g = 32.7 \pm 0.2 \, \text{Å} \), agrees well with the experimental one (Table 4). Besides, its hydrodynamic features correlate well with the experimental data as shown in Table 4. The bead model shows a very elongated molecule (\( \approx 116 \, \text{Å} \) long) with two well differentiated regions (Fig. 7B). One of them resembles a flattened ellipsoid whose dimensions (ca. 46 \( \times \) 38 \( \times \) 30 Å) are in the order of those reported for the CM of Cpl-1 (9). The other one is elongated (ca. 70 \( \times \) 30 \( \times \) 20 Å) as the model proposed for the CWBR and shows a small curvature along the long axis. The assignment was further supported by manual fitting of the high resolution models obtained for the isolated modules into the bead-model surface (Fig. 7C). Two models with similar deviations between the experimental and the theoretical profiles of SAXS were finally selected whose major divergence lies in the relative position of the CM (Fig. 7C). In model \( a \) (Chi-square = 1.76) the catalytic cavity is oriented towards the same side than the convex face of the CWBR whereas it is rotated toward the side of the concave face in model \( b \) (Chi-square = 1.77). In both models the region of the \( \beta \)-barrel devoid of \( \alpha \)-helices (\( \beta_6 \) to \( \beta_8 \)) seems to be part of the interface with the CWBR. Table 4 summarizes the Chi-square values for the fit to the data of the theoretical scattering profiles together with the structural and hydrodynamic parameters estimated by CRYSOI or HYDROPRO for both models. The theoretical sedimentation coefficient is similar for both models and close to the experimental value, as it also occurred with the Stockes radius and the radius of gyration estimated by HYDROPRO (Table 4), reflecting the compatibility of the proposed models with the hydrodynamic features of Cpl-7.

**DISCUSSION**

Endolysins are bacteriophage-encoded enzymes produced during the late phase of gene expression in the lytic cycle to degrade peptidoglycan, thereby enabling the virion progeny to be liberated. Their unique ability to cleave peptidoglycan in a generally species-specific manner makes of endolysins promising antibacterial agents, capable of killing pathogenic bacteria without affecting the normal microbiota. Moreover, the use of near-species-specific endolysins might help to avoid the resistance that appears with the use of broad-range antimicrobials. Species specificity in endolysins produced by pneumococcal bacteriophages relies on the acquisition of additional modules that specifically target the cell wall, and increase the lytic activity by several orders of magnitude (2, 6). In the Cpl-7 endolysin, encoded by the pneumococcal bacteriophage Cp-7, the targeting region comprises a set of three perfect repeats – the CW_7 motifs – sequentially unrelated with the choline-binding motifs that characterize the cell wall-binding modules of all the other CWHs produced by the pneumococcus and its bacteriophages.

Although Cpl-7 was described in 1990, sequences similar to the CW_7 repeats were not reported until 2002 as being encoded by various
prophages of Streptococcus pyogenes and Streptococcus agalactiae (supplemental Table S1). During the last years, the availability of relatively inexpensive next-generation sequencing technologies has permitted the complete genomic sequences of hundreds of genomes of prokaryotic organisms, and today up to 67 different proteins comprising CW_7 motifs have been included in the Entrez Database (last date accessed, 8 April 2010). Although the majority of these proteins are annotated as ‘hypothetical’, they contain combinations of motifs that permit to classify them as probable CWHs. Only 10 proteins do not appear to contain any known motif in addition to CW_7 (Fig. 1). To our knowledge, apart from the Cpl-7 lysozyme, only the endolysin encoded by the λSa2 prophage of S. agalactiae strain 2603 V/R has been biochemically characterized so far, and showed to be a bifunctional enzyme with γ-D-glutamyl-L-lysine endopeptidase (Amidase_5; PF05382) and β-D-N-acetylglucosaminidase (Glucosaminidase; PF01832) activities (63). Cell wall-degrading activity has also been reported for the LySMP endolysin of the lytic phage SMP infecting Streptococcus suis type 2 strains (64), a close homologue (76% identity) of the λSa2 endolysin. It is worth noting that only in the cpl-7 gene, the long (144 nucleotides) tandem repeat units are identical (1), which suggests that the duplication event occurred very recently, in evolutionary terms.

It is interesting to notice that proteins containing CW_7 repeats appear to be present only in four bacterial phyla (out of 25): Actinobacteria, Bacteroidetes, Chloroflexi, and Firmicutes. Surprisingly, they appear to be absent from bacteria belonging to the Proteobacteria phylum, in spite of the fact that the NCBI genome database contains more than 800 proteobacterial genomes. Most of the bacterial species coding for CW_7-containing proteins belong to the Firmicutes phylum and nearly all form part of the human gut (fecal) microbiota where, in healthy adults, 80% of the identified fecal bacteria can be classified into three dominant phyla: Bacteroidetes, Firmicutes and Actinobacteria, although the ratio of Firmicutes to Bacteroidetes evolves along the different life stages (65). The fecal microbiota is a highly complex and diverse bacterial ecosystem. Prominent clusters include some of the most abundant gut species, such as members of the Bacteroidetes and Eubacterium/Ruminococcus groups and also bifidobacteria (Actinobacteria), Proteobacteria and streptococci/lactobacilli groups (58). The finding that ‘D. ethenogenes’ (phylum Chloroflexi) and E. harbinense (phylum Firmicutes), two species living in contaminated waters, also code for CW_7-containing proteins may be the result of horizontal gene transfer events. Interestingly, although the root of the tree of life (and the last universal common ancestor) has been suggested to be located within or next to the Chloroflexi (66), evidences suggesting that ‘D. ethenogenes’ has experimented horizontal transfer events have been provided (67).

As mentioned above, most of the proteins containing CW_7 motifs appear to be related with the presence of prophages (or parts of them). This hypothesis is consistent with the observation that some of these proteins are found only in certain strains of a particular species. For instance, the putative NAM-amidase EEW16061 (Corynebacterium jeikeium strain ATCC 43734) is absent in the genome of strain K411 (Accession number CR931997). Another interesting example is the protein EEJ91939, identified in strains SD2112 and CF48-3A (EEI65431) of Lactobacillus reuteri (supplemental Table S1), but missing in the genomic sequences of strains DSM 20016 (CP000705), MM2-3 (ACLB01000000), MM4-1A (ACGX01000000), JCM 1112 (AP007281), or 100-23 (ACGW01000000). Phages (either temperate or lytic) may have served as vehicles for horizontal gene transfer among different bacterial species; among other examples, this is strongly supported by the striking similarity found between the prophage φ3396 from S. dysgalactiae subsp. equisimilis and the S. pyogenes φ315.1 temperate phage (68). Furthermore, evidences suggesting that the virulent lactococcal phase 1706 may have originated from a phage infecting Rumino- cocccus torques and/or Clostridium leptum has been provided (69).

Here we have reported the first in-depth characterization of the Cpl-7 endolysin. The protein was purified to electrophoretic homogeneity by a simple procedure, including protein fractioning with ammonium sulfate and two chromatographies. Cpl-7 behaves in solution as a monomer and its hydrolytic activity on pneumococcal cell walls is slightly lower than that of the Cpl-1 endolysin, whose good profile as a potent antimicrobial agent against pneumococci has been already proved (17–19, 22). By combination of SAXS and molecular modeling, a threedimensional model of the overall protein structure, compatible with the SS inferred from CD and prediction methods, has been proposed.
SAXS experiments showed that Cpl-7 adopts in solution an extended configuration, as also indicated the velocity sedimentation results. Even though SAXS analysis is unable to identify unique structures because of rotational averaging, it is able to rule out incorrect structures, including more globular arrangements of Cpl-7 modules that will have displayed different radii of gyration and maximum inter-particle distances. Besides, the molecular envelope generated from SAXS data is clearly compatible with the structures derived for the isolated modules of Cpl-7 using molecular modeling, and the high resolution models proposed for the overall protein structure are also compatible with the SAXS spectrum and Cpl-7 hydrodynamic behavior. All this indicates that the proposed models represent a reasonable approximation to the real structure of the protein. Besides, the fitting of CWBR high-resolution models into the SAXS envelope might denote a certain mobility of the last repeat, probably due to the own modular structure displayed by the CWBR of Cpl-7.

The model of the CM displays a very good stereochemistry and the rmsd from the template was rather low, which indicates that it constitutes an excellent approximation to the actual module structure. Non-conservative substitutions are primarily localized at the region comprising from helix-3 to helix-4 (Fig. 4C). According to the model inferred for the complex formed by Cpl-1 and (2S5P)₃ muropeptide (10), substitutions at positions 12 and 96 would modify the interaction with the substrate by i) reducing the number of hydrogen bonds formed in the complex with Cpl-7, and ii) inverting the electrostatic interactions (unfavorable in Cpl-7) between the side chain of the residue at position 96 and the stem peptide of the MurNAc at position –1. Besides, in TIM barrels, the loops at the C-terminus of β-strands usually contribute to determine the specific configuration of the active site and the substrate specificity. Therefore, sequence differences in the loop containing the proton donor (residues 93 to 100) could also contribute to modulate the murolytic activities of Cpl-1 and Cpl-7. A similar effect might also have the drastic change in the protein hydrogen-bond network derived from substitution of Arg63 by Cys in Cpl-7 described above. The relevance of the substitutions accumulated on the edge of the CM along helices three and four that strongly modify the surface charge distribution remains unknown. Participation of this region in the CM/CWBR interface seems unlikely considering the best SAXS-based models for the overall Cpl-7 structure, and its distance to the glycance chain bound to the active site is too large to interact with it (Fig. 4C). However, the possibility that this part of the CM might interact with adjacent glycopeptidic chains, providing additional contacts with the cell wall cannot be presently discarded.

The stereochemical and energetical profiles of the CWBR model were fairly good. According to the proposed model, the CWBR comprises three tandemly-assembled three-helical bundles, whose disposition reminds a super-helical structure, connected by 6 amino acid-long linkers. Each bundle contains 42 residues and the inner core is stabilized by conserved hydrophobic residues, whereas sharp turns between helices are facilitated by the presence of glycine residues highly conserved in all the related sequences so far identified (Fig. 2). The lack of three residues in the middle of certain CW_7 repeats would probably shorten the bundle second-helix. Although the model may not reveal in detail the module structure, it provides some clues to aid identification of residues and surfaces important for the cell wall attachment. Residues located on the convex face of the CWBR are poorly conserved within the CW_7 family. On the contrary, conserved acidic and aromatic residues are present along the concave face (Fig. 6C), suggesting that this could be the region involved in cell wall attachment, since solvent-exposed areas involved in functional interactions with other macromolecules are subjected to strong evolutionary pressure and, therefore, show a higher degree of amino acid conservation than otherwise expected for surface residues. Moreover, both aromatic and polar residues with planar side chains, as those conserved in the concave face, are a common signature of glycan-binding sites in carbohydrate recognition motifs (60).

Among the great variety of cell wall-binding motifs acquired by CWHs only a few have been structurally characterized so far, including SH3.5 (PF08460), LysM (PF01476), SPOR (PF05036), PG_binding_1–3 (clan PGBD; CL0244), and CW_binding_1 (choline-binding repeats; PF01473) domains. They all comprise relatively short sequences forming different structural motifs. PG_binding motifs show a three-helical bundle fold but, in contrast to the CW_7 motif, the helices have a parallel disposition and are connected by rather long loops. This could explain why they were not identified as potential templates of the CWBR by fold recognition methods, and suggests that CW_7 repeats
might constitute a new cell wall-binding structural motif. The existence of protein sequences with a single CW_7 repeat evidences the ability of each bundle to fold and recognize its target, as demonstrated by the increase in activity observed upon fusing one CW_7 motif to the endopeptidase module of the λSa2 endolysin (3). However, the common presence (21 out of 67) of tandemly-arranged repeats suggests a mechanism to improve cell wall recognition, and the model proposed here for the CWBR of Cpl-7 represents a first approximation to the architecture that multiple CW_7 repeats would adopt. It is worth noting that the acquisition of a second CW_7 repeat by the endopeptidase module of the λSa2 endolysin not only enhanced the lytic activity of the truncated protein on different streptococcal species, but it was shown to be essential for its activity on staphylococci (3). Moreover, the acquisition of other cell wall-binding motifs, like LysM and SPOR, by so many CW_7-containing proteins (Fig. 1) could denote alternative ways for tight substrate binding, by directing muropeptides into the catalytic cavity, or even modifying the species (or the strain) specificity in certain CWBs.

The CM module is linked to the CWBR by an extremely polar segment made of 16 residues, including six acidic and two lysine residues, which probably adopts an extended conformation and introduces certain intermodular flexibility. Besides, the linker sensitivity to the proteolytic action of trypsin (8) suggests a solvent exposed disposition according to its polar character. Nevertheless, the mass density at the module’s interface in the SAXS-derived model strongly indicates the existence of a large interacting surface. According to it, the interface between Cpl-7 modules would comprise, as in Cpl-1 (9), the region of the catalytic barrel lacking α-helices, which is fully conserved in both lysozymes. Finally, if the linker would adopt an extended configuration its estimated length (≥ 64 Å) does not exclude the possibility of a tail-to-tail disposition of the CM and the CWBR in the protein overall structure. Whatever the case, the relative disposition of Cpl-7 modules within the protein structure diverges from the one adopted in Cpl-1 lysozyme (9, 11).

The effective lytic spectrum of endolysins is usually constrained by cell wall elements that serve as binding targets and are themselves distributed in species/strain specific manners. Nevertheless, endolysins often lyse a broader range of species than just the producing organism, and the substrate specificity inherent to the acquisition of the CW_7 motifs by the Cpl-7 endolysin opens appealing perspectives to the use of this lysozyme, either wild-type or in chimeric constructions, to control pathogenic, resistant bacteria. Interestingly, CW_7 repeats are also present in the CWBs encoded by a large variety of Gram-positive pathogens and certain phages of S. agalactiae (a leading cause of bacterial sepsis, pneumonia and meningitis in neonates) and S. pyogenes (a prevalent human pathogen that has re-emerged as a public health hazard). To that end, the great variety of architectures shown by the CW_7-containing proteins highlights the versatility of this motif to combine with different CMs and cell wall-binding domains in a single polypeptide chain to alter bond and/or substrate specificity while retaining activity.

REFERENCES

261–277

FOOTNOTES

This work was supported by grants BFU2006-10288, SAF2009-10824, BIO2007-61336 and BFU2009-10052 from Ministerio de Ciencia e Innovación. Additional funding was provided by
BIPPED-CM from the ‘Comunidad de Madrid’, the CIBER de Enfermedades Respiratorias (CIBERES), an initiative of the ISCIII, and the Glycodynamics network (FP6-UE, MCTN-CT-2005-019561). N. Bustamante was initially supported by a fellowship from MEC. We also acknowledge the support of the European Community, Beam time proposal number 48086.

The abbreviations used are: CD, circular dichroism; CM, catalytic module; CW_7, cell wall-binding repeat of Cpl-7; CWBR, cell wall-binding region; CWH, cell wall hydrolase; GlcNAcase, N-acetylglucosaminidase; GH, glycosyl hydrolase; NAM-amidase, N-acetylmuramoyl-L-alanine amidase; OD, optical density; PDB, protein data bank; rmsd, root mean square deviation; (S2P5)_3, [GlcNAc-MurNAc-(L-Ala-D-isoGln-L-Lys-D-Ala-D-Ala)]_3; SAXS, small angle X-ray scattering; SS, secondary structure.

FIGURE LEGENDS

FIGURE 1. Schematic representation of proteins containing the CW_7 motif. Proteins were ordered by increasing complexity. The accession number of each protein is shown. Comprised modules are color-coded and their designations according to the Pfam database are shown at the bottom. Additional data are shown in supplemental Table S1.

FIGURE 2. Sequence conservation in the CW_7 motifs. Alignment of representative sequences of the CW_7 motif from Pfam entry PF08230. The Entrez accession numbers are shown. The last amino acid positions are indicated on the right. Asterisks, colons, and dots correspond to strictly conserved residues, conservative, and semi-conservative substitutions (based in BLOSUM substitution matrices), respectively. Colored mark-up is: pink, glycine; yellow, proline; light blue, small and hydrophobic; dark blue, tyrosine and histidine; green, hydroxyl and neutral polar; red, acidic; and black, basic. Residues with percentages of solvent-accessible surface below 25%, 5% and 0%, as well as predicted α-helices (H) along the motif are also indicated.

FIGURE 3. Characterization of Cpl-7 secondary and quaternary structures. A, far-UV CD spectra; symbols are the experimental data and the continuous line the theoretical fit of CONTIN. B, distribution of sedimentation coefficients measured at 2.5 μM (open squares), 6.5 μM (circles), 11.7 μM (triangles) and 16.8 μM (black squares). C, sedimentation equilibrium profile at 17,000 rpm showing the fit (continuous line) of the experimental data (circles) to a single species with an average molecular mass of 40.2 ± 0.2 kDa; the residual plot is shown in panel D. All measurements were performed in Pi buffer at 20 °C.

FIGURE 4. Homology model of the Cpl-7 catalytic module. A, sequence alignments of Cpl-7 and Cpl-1 CMs. Color code as in Figure 2. The position of β-strands and α-helices in the CM of Cpl-1 is shown. Black triangles indicate the catalytic residues (Asp10, Asp92, Glu94 and Asp182). B, superimposition of the template (green) and model (blue) structures. Catalytic residues are in stick representation (Cpl-7, yellow; Cpl-1, green). C, Superimposition of the coordinates for Cpl-1/(2S5P)_3 complex (10) with the computational model of Cpl-7 CM. Non-conserved residues together with substitutions relevant for substrate binding are shown in stick representation (red, acidic; blue, basic; green, polar uncharged; and gray, hydrophobic). (2S5P)_3 stands for [GlcNAc-MurNAc-(L-Ala-D-isoGln-L-Lys-D-Ala-D-Ala)]_3; carbon atoms are in black (sugar chain) or orange (peptide stem).

FIGURE 5. Three-dimensional threading model of Cpl-7 CWBR. A, structure-based sequence alignments of the CWBR and the template (PDB code 3c1d, RecX protein). Residues in α-helical conformation according to RecX structure and Cpl-7 SS prediction are highlighted in gray. Strictly conserved residues are marked with asterisks; colons and dots indicate, respectively, conservative and semi-conservative changes. Residues conserved in RecX family sequences (59) are in red, as also are similar residues present in the CWBR sequence. Green traces under the alignments indicate α-helix positions in model 1 (top) and 2 (bottom). B, superimposition of template molecules a (cyan) and b (blue) present in 3c1d structure with the best CWBR structural models (M1, red; M2, green).
FIGURE 6. *Three-dimensional structure and surface properties of Cpl-7 CWBR.* A, three-helical bundle of the second repeat. Side chains of hydrophobic residues conserved in CW_7 family are shown in stick representation (V56 corresponds to V8 in the multiple alignment of Fig. 2A). B, molecular surface of the CWBR colored according to its electrostatic potential (top, concave face; bottom, convex face). C, distribution of conserved aromatic and charged residues along the concave face.

FIGURE 7. *SAXS-based modeling of Cpl-7 overall structure.* A, experimental SAXS profile of Cpl-7 (circles) and theoretical spectra calculated from the *ab initio* bead model (green line) in B, and the high resolution models in C. The inset shows the pair-distance distribution function, $P(r)$, generated by GNOM with a $D_{\text{max}}$ of 115 Å (data collected at 4 °C in Pi buffer). B, *ab initio* bead model of Cpl-7 derived from SAXS data using DALAI_GA (50). C, molecular envelope (grid representation) of Cpl-7 bead model with the best fits of the three-dimensional models built for the CM and the CWBR manually docked inside (the antiparallel $\beta 8$ strand of the CM is colored in red). Red and blue lines in A are the theoretical SAXS profile derived for models $a$ (left hand model) and $b$ (right hand model), respectively, using CRYSOL (54).
TABLE 1
Stereochemical and energetical evaluation of the models generated for the CM and the CWBR of Cpl-7

<table>
<thead>
<tr>
<th>Protein</th>
<th>PROCHECK summary&lt;sup&gt;a&lt;/sup&gt;</th>
<th>QMEAN&lt;sup&gt;b&lt;/sup&gt;</th>
<th>VERIFY3D&lt;sup&gt;c&lt;/sup&gt;</th>
<th>ERRAT&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Most favored %</td>
<td>Additional allowed %</td>
<td>Generously allowed %</td>
<td>Disallowed %</td>
</tr>
<tr>
<td>CM-Cpl-7</td>
<td>87.3</td>
<td>11.4</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>1h09&lt;sup&gt;e&lt;/sup&gt;</td>
<td>89.0</td>
<td>10.3</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>CWBR-M1</td>
<td>93.1</td>
<td>5.2</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>CWBR-M2</td>
<td>93.3</td>
<td>5.0</td>
<td>1.7</td>
<td>0.0</td>
</tr>
<tr>
<td>3cd1 (b)&lt;sup&gt;f&lt;/sup&gt;</td>
<td>97.0</td>
<td>3.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<sup>a</sup> Distribution of residues in the Ramachandran plot region.

<sup>b</sup> Pseudo-energy of the global structure (46).

<sup>c</sup> Minimum and maximum values of energy per residue (45).

<sup>d</sup> Percentage of residues whose structure is under the rejection limit (47).

<sup>e</sup> Evaluation of the CM structure.

<sup>f</sup> Protein subunit is indicated in parentheses.
<table>
<thead>
<tr>
<th>PDB</th>
<th>FUGUE</th>
<th>SP4</th>
<th>3DPSSM</th>
<th>Template structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ranking</td>
<td>Score</td>
<td>Ranking</td>
<td>Score</td>
</tr>
<tr>
<td>3c1d (b)</td>
<td>1</td>
<td>4.35</td>
<td>1</td>
<td>4.94</td>
</tr>
<tr>
<td>1ng6</td>
<td>9</td>
<td>1.95</td>
<td>5</td>
<td>3.79</td>
</tr>
</tbody>
</table>

*Only the top 10 hits of each method were considered.*

*Protein subunit is indicated in parentheses.*
<table>
<thead>
<tr>
<th>PDB&lt;sup&gt;b&lt;/sup&gt;</th>
<th>FUGUE</th>
<th>SP4</th>
<th>3D-PSSM</th>
<th>Template structure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ranking</td>
<td>Score</td>
<td>Ranking</td>
<td>Score</td>
</tr>
<tr>
<td>1ixs (a)</td>
<td>2</td>
<td>3.11</td>
<td>1</td>
<td>7.15</td>
</tr>
<tr>
<td>1gab</td>
<td>7</td>
<td>2.47</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>1ytf (b)</td>
<td>–</td>
<td>–</td>
<td>5</td>
<td>6.07</td>
</tr>
<tr>
<td>1fse (a)</td>
<td>6</td>
<td>2.52</td>
<td>10</td>
<td>5.54</td>
</tr>
</tbody>
</table>

<sup>a</sup> Only the top 10 hits of each method were considered.

<sup>b</sup> The protein subunit is indicated in parentheses.
### TABLE 4
Comparison of the structural and hydrodynamic parameters of SAXS-based model of Cpl-7 with the experimental values measured by SAXS or sedimentation velocity

<table>
<thead>
<tr>
<th>Parameter</th>
<th>SAXS</th>
<th>Ab initio model</th>
<th>High resolution models</th>
<th>Sedimentation velocity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Guinier</td>
<td>GNOM</td>
<td>DALAI GA</td>
<td>HYDRO</td>
</tr>
<tr>
<td>$R_g$ (Å⁻¹)</td>
<td>29.8</td>
<td>30.3</td>
<td>32.7</td>
<td>33.3</td>
</tr>
<tr>
<td>$D_{max}$ (Å⁻¹)</td>
<td>-</td>
<td>115</td>
<td>116</td>
<td>-</td>
</tr>
<tr>
<td>$R_s$ (Å⁻¹)</td>
<td>-</td>
<td>-</td>
<td>32.9</td>
<td>-</td>
</tr>
<tr>
<td>$s$ (S)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>2.87</td>
</tr>
<tr>
<td>Chi-square&lt;sup&gt;a&lt;/sup&gt;</td>
<td>-</td>
<td>4.79×10⁻⁵</td>
<td>7.18×10⁻⁵</td>
<td>-</td>
</tr>
</tbody>
</table>

<sup>a</sup>Chi-square deviation between the experimental and theoretical scattering profiles.

<sup>b</sup>Calculated from the Fourier transform of $P(r)$. 
Fig. 1. Bustamante et al.
Fig. 2. Bustamante et al.
Fig. 3. Bustamante et al.
Fig. 4. Bustamante et al.
Fig. 5. Bustamante et al.
Fig. 6. Bustamante et al.
Fig. 7. Bustamante et al.
Cpl-7, a lysozyme encoded by a pneumococcal bacteriophage with a novel cell wall-binding motif
Noemi Bustamante, Nuria E. Campillo, Ernesto Garcia, Cristina Gallego, Benet Pera, Gregory P. Diakun, Jose L. Saiz, Pedro Garcia, J. Fernando Diaz and Margarita Menendez

J. Biol. Chem. published online August 18, 2010

Access the most updated version of this article at doi: 10.1074/jbc.M110.154559

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/2010/09/01/M110.154559.DC1

This article cites 0 references, 0 of which can be accessed free at
http://www.jbc.org/content/early/2010/08/18/jbc.M110.154559.full.html#ref-list-1