Structure, Vol. 11, 1239–1249, October, 2003, ©2003 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.str.2003.09.005

Structural Basis for Selective Recognition of Pneumococcal Cell Wall by Modular Endolysin from Phage Cp-1

Juan A. Hermoso,[1,*] Begoña Monterroso,[2] Armando Albert,[1] Beatriz Galán,[3] Oussama Ahrazem,[3] Pedro García,[3] Martín Martínez-Ripoll,[1] José Luis García,[3] and Margarita Menéndez[2]

[1]Grupo de Cristalografía Macromolecular y Biología Estructural

[2]Departamento de Química-Física de Macromoléculas Biológicas

Instituto Química-Física Rocasolano
CSIC
Serrano 119
28006 Madrid
Spain

[3]Departamento de Microbiología Molecular
Centro de Investigaciones Biológicas
CSIC
Velázquez 144
28006 Madrid
Spain

Summary

Pneumococcal bacteriophage-encoded lysins are modular choline binding proteins that have been shown to act as enzymatic antimicrobial agents (enzybiotics) against streptococcal infections. Here we present the crystal structures of the free and choline bound states of the Cpl-1 lysin, encoded by the pneumococcal phage Cp-1. While the catalytic module displays an irregular $(\beta/\alpha)_5\beta_3$ barrel, the cell wall-anchoring module is formed by six similar choline binding repeats (ChBrs), arranged into two different structural regions: a left-handed superhelical domain configuring two choline binding sites, and a β sheet domain that contributes in bringing together the whole structure. Crystallographic and site-directed mutagenesis studies allow us to propose a general catalytic mechanism for the whole glycoside hydrolase family 25. Our work provides the first complete structure of a member of the large family of choline binding proteins and reveals that ChBrs are versatile elements able to tune the evolution and specificity of the pneumococcal surface proteins.

Introduction

Streptococcus pneumoniae is a common and important human pathogen associated with pneumonia, septicemia, meningitis, and otitis media. The high morbidity and mortality caused by pneumococcal diseases, particularly in infants, elderly, and immunocompromised patients, is exacerbated by the increasing prevalence of antibiotic-resistant strains and the suboptimal efficacy of available vaccines (Kristinsson, 1997; Pelton, 2000). Since the integrity of the peptidoglycan is essential for bacterial survival, bacteriophage-coded cell wall lysins may constitute effective antibacterial agents against their host. All pneumococcal bacteriophages known so far code for either an N-acetylmuramoyl-L-alanine amidase or a lysozyme that hydrolyzes the pneumococcal cell wall for progeny liberation (López et al., 1997). The capability of two amidases encoded by streptococcal phages to kill streptococci and to prevent and eliminate the pharyngeal colonization in mice has been recently proven (Loeffler et al., 2001; Nelson et al., 2001), strongly suggesting that other pneumococcal lysins, as phage-encoded lysozymes, can also constitute effective antimicrobial agents (enzybiotics) against *S. pneumoniae*. In this sense, recent results carried out using a mouse model for pneumococcal septicemia have established that Cpl-1, the lysozyme coded by the pneumococcal phage Cp-1, acts indeed as an enzybiotic (P.G. and J.L.G., unpublished data).

All the pneumococcal endolysins so far described display a modular structure such that, in addition to the catalytic module, most of them posses a choline binding module to facilitate their anchoring to the choline-containing teichoic acid of the pneumococcal cell wall (García et al., 1988). This choline binding module is formed by a repeat (ChBr) of about 20 amino acids found in multiple tandem copies (ranging from 4 to 18) in a large family of surface proteins (144 members identified by Pfam, http://www.sanger.ac.uk/Software/Pfam/) from gram-positive bacteria and their bacteriophages. These proteins, named choline binding proteins (ChBPs), play important physiological functions in pneumococcal virulence (López et al., 1997). Until now, only the structure of a fragment of the choline binding module from the major pneumococcal autolysin LytA has been reported (Fernández-Tornero et al., 2001, 2002). The Cpl-1 lysozyme (339 amino acids) is an archetype of the ChBP family. This enzyme has been extensively characterized: it cleaves the glycosidic N-acetylmuramoyl-(β1,4)-N-acetylglucosamine bonds of the pneumococcal glycan chain (García et al., 1988; Sanz et al., 1993). Based on amino acid sequence similarity, lysozymes are usually classified into four groups (GH-22, GH-23, GH-24, and GH-25) (Henrissat and Bairoch, 1993). The first three groups show common topological features in their three-dimensional structure (Monzingo et al., 1996) not shared by the GH-25 family. Cpl-1 belongs to the chalaropsis type (GH-25; Ch-type) for which only the structure of the muramidase from *Streptomyces coelicolor,* cellosyl, has been reported (Rau et al., 2001). In addition to the catalytic module, Cpl-1 carries a choline binding module at the C terminus that comprises 6 ChBrs (p1-p6 repeats) plus a short tail (García et al., 1988).

Here, we report the crystal structures of the free and choline bound states of the Cpl-1 endolysin. This is the first 3D structure for an intact ChBP, namely with the catalytic and the cell wall-anchoring modules, a pneumococcal cell wall lysin, and a modular lysozyme, so far reported, largely improving our knowledge of the

*Correspondence: xjuan@iqfr.csic.es

Table 1. Structure Determination and Statistics for Hg Derivative, Native Cpl-1, and Cpl-1:Choline Complex

	Hg λ1	Native	Cpl-1:Cho
Space group	$C222_1$	$C222_1$	$C222_1$
a (Å)	75.90	77.95	78.03
b (Å)	98.72	95.78	95.95
c (Å)	129.07	129.28	129.30
Data Collection			
Wavelength (Å)	1.004	1.004	1.5418
Resolution (Å)	54.0–2.8	39.0–2.1	24.0–2.4
Unique reflections	12,212	28,359	17,479
Redundancy	7.1	4.7	3.5
Completeness (%)	99.4	99.3	96.1
$I/\sigma(I)$	5.5	6.8	7.2
R_{merge}	11.6	7.3	9.8
Refinement Statistics			
Protein nonhydrogen atoms		2763	2741
Ligand nonhydrogen atoms		-	14
Solvent nonhydrogen atoms		326	257
Resolution range (Å)		15.0–2.1	24.0–2.4
R_{work} (%)		20.7	20.5
R_{free}* (%)[a]		25.9	26.4
Rmsd bonds (Å)		0.005	0.006
Rmsd angles (°)		1.30	1.30
Average B factor (Å^2)		34.6	28.0

[a] R calculated for 7% of data excluded from the refinement.

modular organization of ChBPs and the interactions between their modules. This valuable information will be very useful for understanding the mechanisms involved in degradation of the pneumococcal envelope and for developing more potent enzybiotics against pneumococci. In addition, it provides interesting clues on the evolution and functional adaptation of modular proteins such as the paradigmatic family of ChBPs.

Results and Discussion

Overall Structure of Cpl-1

The crystal structure of Cpl-1 was solved using the single-wavelength anomalous diffraction (SAD) method (see Experimental Procedures and Table 1). No significant differences are observed between the overall Cpl-1 structures in both the free state and in the complex with choline (root-mean-square deviation of 0.36 Å for 339 Cα atoms). Hence, hereafter and unless stated, all the structural results will refer to the Cpl-1 unbound state. The polypeptide chain consists of the catalytic and the choline binding modules joined by an acidic linker comprising residues 189–199 (Figure 1A). The catalytic module is formed by a single structural domain (residues 1–188) resembling a flattened ellipsoid of dimensions 45 × 35 × 25 Å^3 that folds into an irregular $(\beta/\alpha)_5$ β_3 barrel (Rau et al., 2001). As in regular TIM barrels, the first five β strands and α helices alternate, but the α5 helix is followed by strands β6 to β8, which are connected by loops lacking any helices (Figure 1B). All β strands are arranged in a parallel fashion, except β8, which runs antiparallel to the other strands. A small hairpin and a short helix (αA) cap the C-terminal side of the β sheet (Figure 1B). The well-defined structure of the linker unambiguously identifies the beginning of the choline binding module. Each repeat (p1–p6) comprises a symmetrical β hairpin followed by a loop and a coiled region (Figure 1C) and are arranged into two well-defined structural regions referred to as CI and CII (Figure 1A). The CI structural domain (residues 200–281) folds following a left-handed superhelical arrangement of the initial four repeats with the hairpins extending perpendicularly to the axis of the superhelix, each CI repeat being located at a 120° counterclockwise rotation. The CII structural domain (residues 282–339) folds completely differently as an almost antiparallalel six-stranded β sheet formed by the last two repeats (p5–p6) and the C-terminal tail. This region is responsible for the interaction between Cpl-1 modules. Overall, the choline binding module of Cpl-1 constitutes a novel and distinct arrangement of six conserved supersecondary repeats.

The Catalytic Module

The Cpl-1 Active Site

The active site of Cpl-1 is located at the C-terminal end of the catalytic barrel, where a long groove is found. This cleft displays shape and charge properties consistent with the peptidoglycan binding site. It also shows a deep hole of highly negative electrostatic potential that constitutes the active site. Lysozyme enzymatic hydrolysis of the glycosidic bonds takes place via a general acid/base catalysis that requires two critical acidic residues: a glutamate acting as a proton donor and an aspartate acting as a nucleophile/base. Cpl-1 lysozyme belongs to the GH-25 family, for which the positions and the nature of the catalytic acidic residues remain uncertain (http://afmb.cnrs-mrs.fr/CAZY/). Among the acidic residues lining the substrate binding site, two pairs of negatively charged residues (Asp10-Asp182 and Asp92-Glu94) are disposed facing each other in the cen-

tral hole (Figure 2A). Three of them (Asp10, Asp92, and Glu94) are strictly conserved, while Asp182 seems to be spatially conserved when the structures of Cpl-1 and cellosyl are compared.

Considering the Cpl-1 active site geometry, each residue from the two pairs of acidic residues mentioned above could be involved in the catalytic process. However, the residual activity found for Cpl-1 mutants D10N, D10E, and D10A (2.2%, 1.7%, and 0.2% of the wild-type, respectively) (Sanz et al., 1992a) pointed to Asp10 as a critical residue in the catalytic mechanism. A search for structural similarities (Holm and Sander, 1996) with the catalytic module of Cpl-1 shows that, apart from cellosyl that belongs to the same glycosil hydrolase family 25 (rmsd of 2.4 Å for 174 Cα atoms), the closest relatives are chitinases, chitobiases, and related enzymes from glycosyl hydrolase families 18 and 20 (rmsd of 3.3–3.5 Å for 147–163 Cα atoms) and glucanases from GH-1, GH-5, and GH-17 families (rmsd of 3.5–3.8 Å for 165–137 Cα atoms). Despite the functional differences among these enzymes and Cpl-1, their structural superposition shows that the proton donor residue is placed in an equivalent position at the C terminus of the fourth β strand of the barrel and matches the Glu94 residue of Cpl-1. In order to test the role of Asp92, Glu94, and Asp182 in the catalytic mechanism, the enzymatic activity of Cpl-1 was investigated by site-directed mutagenesis (Table 2). The activity practically vanishes when Glu94 is replaced either by glutamine or alanine, suggesting that, indeed, Glu94 plays a crucial role in the catalytic mechanism. The decrease in activity observed by mutating Asp92 and Asp182 can be understood in the light of the very short distances found for the H bonds between residues forming the couples Asp10-Asp182 and Asp92-Glu94 (2.42 Å for Asp10-Asp182 and 2.59 Å for Asp92-Glu94). This result suggests that these pairs of residues behave as low barrier hydrogen bonds (Cleland et al., 1998), allowing protons to move freely between them. Low barrier hydrogen bonds have been proposed to be involved in proton trafficking, thus ensuring regeneration of the protonated states of the catalytic residues in the κ-carrageenase (Michel et al., 2001).

On the other hand, the average distance found between the pair of carboxylic acids (Asp10 and Glu94), 9.5 Å, is that expected for an inverting enzyme (9.0 and 9.5 Å for inverting α- and β-glycosidases, respectively, and 4.8 and 5.3 Å for retaining α- and β-glycosidases, respectively) (Wang et al., 1994). All these facts suggest that hydrolysis occurs via a net inversion of the anomeric configuration, with Asp10 acting as the general base, helping to activate the nucleophilic water molecule, and Glu94 acting as the general acid, protonating the departing oxygen atom in a concerted fashion as the bond cleaves. As mentioned above, all the residues involved in the proposed catalytic mechanism are strictly conserved in the GH-25 family, suggesting its general applicability for all the glucosidases belonging to this family.

Pneumococcal Peptidoglycan Susceptibility to Ch-Type Lysozymes

Resistance of pneumococcal peptidoglycan to the hydrolytic action of lysozymes is due to its unusual high proportion of non-N-acetylated hexosamine units (over 80% of the glucosamine and 10% of the muramic acid residues) (Vollmer and Tomasz, 2000). In this sense, the identified gene *pgdA*, encoding for a peptidoglycan N-acetylglucosamine deacetilase A, may contribute to pneumococcal virulence by providing protection against host lysozymes (Vollmer and Tomasz, 2000).

Superimposition of Cpl-1 with the X-ray structures of several chitinolytic enzymes from families 18 and 20, complexed with substrates or inhibitors (Tews et al., 1997), indicates that the substrate polysaccharide chain can be placed along the groove containing the Cpl-1 active site (Figure 2B). According to this model, the C2 N-acetyl group of glucosamine (+1 position), pointing out toward the solvent, would not be recognized by Cpl-1. On the contrary, this N-acetyl group strongly interacts with polar or charged residues (Asn44 and Glu35; HEWL numbering) within the catalytic cleft of lysozymes from families 22, 23, and 24 due to their different fold. This difference in substrate recognition would explain why the unusual high proportion of non-N-acetylated glucosamine residues in pneumococcal peptidoglycan made it resistant to lysozymes from families GH22-24 but does not affect its degradation by the Ch-type enzymes.

The Choline Binding Module

Structural Comparison with LytA Choline Binding Module

Recently, the structure of a fragment of the choline binding module from the major pneumococcal autolysin LytA (hereafter C-LytA) in the presence of choline has been reported (Fernández-Tornero et al., 2001, 2002). The C-LytA structure was obtained from an incomplete choline binding module (the linker and half of the first repeat were deleted) and in the absence of the N-terminal amidase catalytic module. Despite the high sequence similitude between the choline binding modules of Cpl-1 and LytA (more than 50% identity), very important structural differences are observed (Figure 3). C-LytA presents a single regular fold formed by a superhelical arrangement of the ChBrs and the C-terminal tail, a fold that authors claimed is maintained by choline binding (Fernández-Tornero et al., 2001). In contrast, the cell wall-anchoring module of Cpl-1 displays two well-differentiated structural domains: CI, showing a superhelical fold similar to that of C-LytA, and CII, a new β sheet-folded region involved in intermodular interactions. On the other hand, the geometry of Cpl-1 choline binding sites is essentially identical in the presence or in the absence of choline (see Figure 4A), and the superhelical fold of the p1-p4 ChBrs is also maintained in the absence of the aminoalcohol. Therefore, the absolute requirement of choline for preserving such geometry is doubtful. This fact is in agreement with previous thermal denaturation studies showing that the choline-free choline binding modules of LytA and Cpl-1 denatured in a highly cooperative way (Varea et al., 2000; Sanz et al., 1993).

The structural differences observed between the C-LytA structure and the choline binding module of Cpl-1 lysozyme can be attributed either to the sequence differences between their ChBrs or to the interactions established between modules in the intact lysozyme. The high percentage of identity between them suggests that the

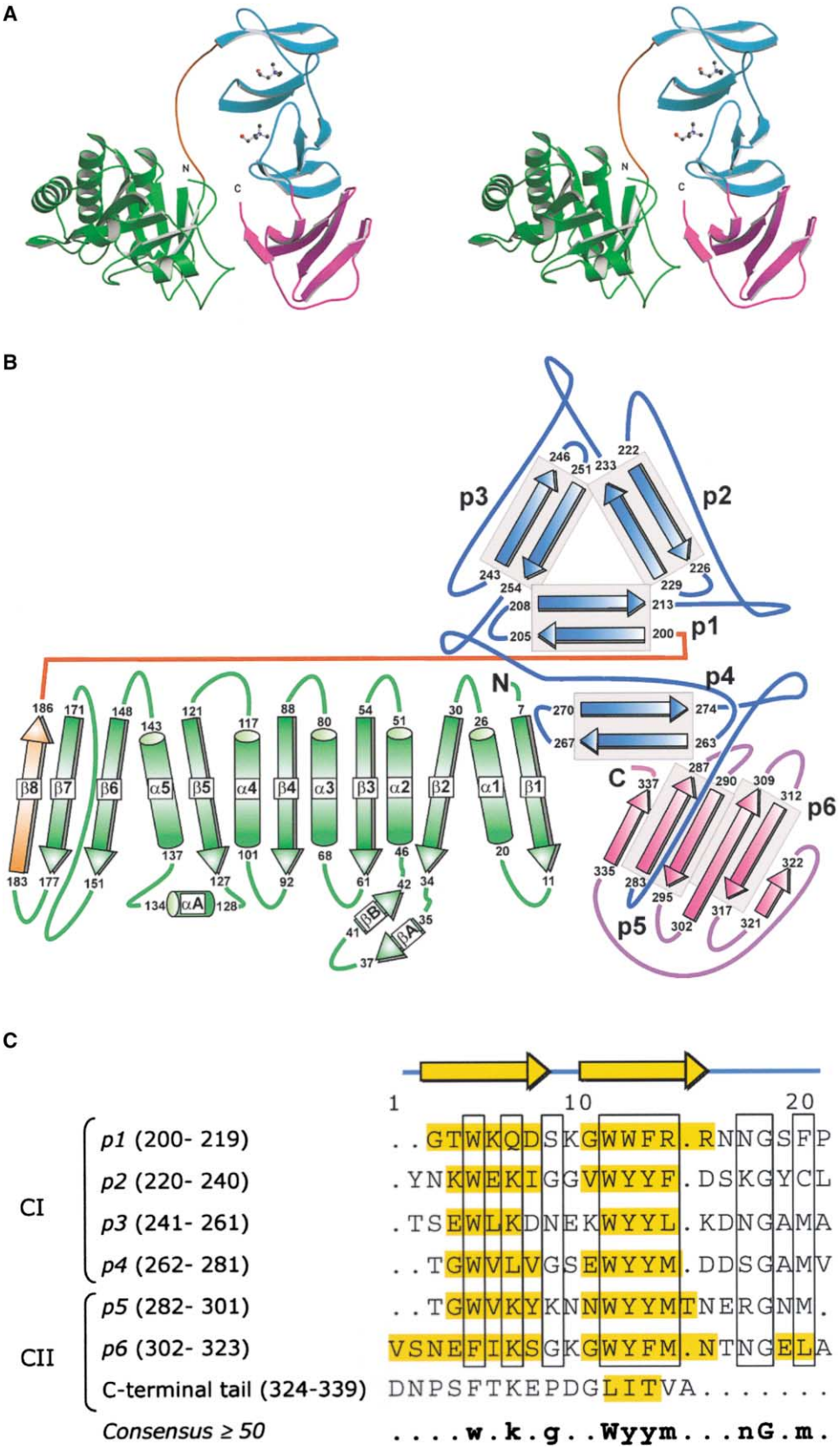

Figure 1. Structure of Modular Cpl-1 Endolysin

(A) Stereo representation of Cpl-1 structure with domains colored differently. Catalytic N-terminal, green; linker, orange; CI domain, cyan; CII domain, magenta. Choline molecules are drawn in a ball-and-stick representation.

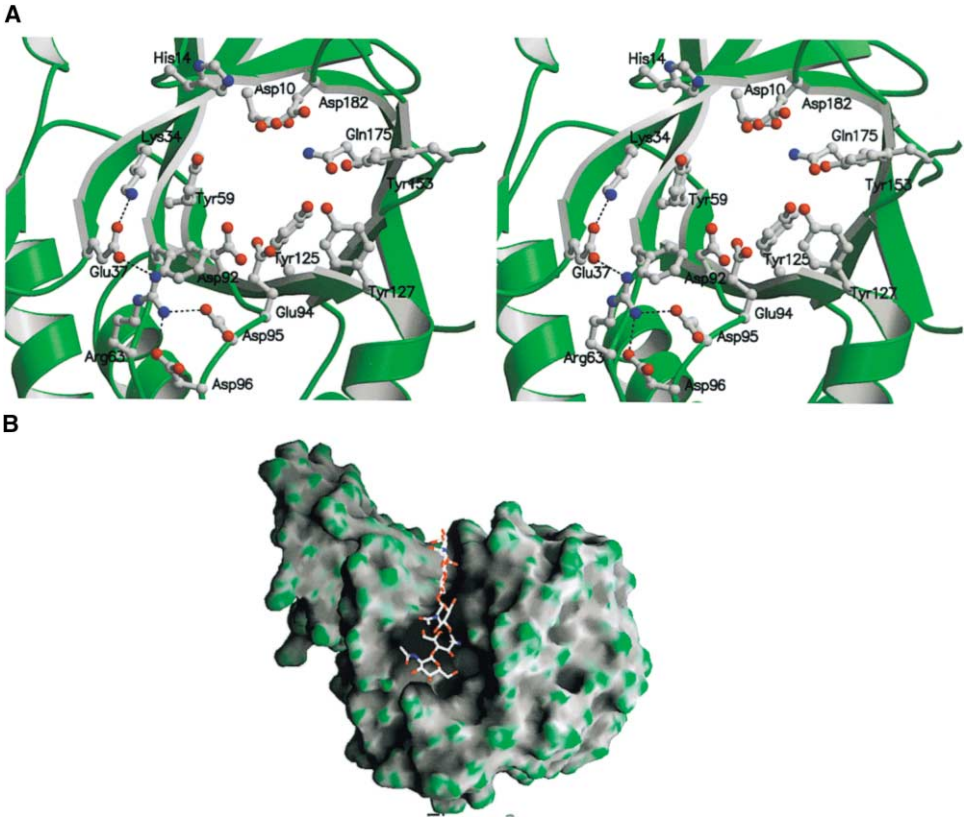

Figure 2. Active Site of Cpl-1

(A) Stereo diagram of the residues lining the central hole of the $(\beta/\alpha)_5$ $(\beta)_3$ TIM-like barrel are shown in a ball-and-stick representation and colored by atom type. The ionic network around the central hole is represented as dotted lines.

(B) Hypothetical Cpl-1:N-acetylglucosamine-pentamer complex as obtained by structural superposition with chitinase B from *Serratia marcescens* in complex with N-acetylglucosamine-pentamer (PDB code 1e6n). No additional fitting was performed on the position of the substrate. Cpl-1 surface curvature is colored from green (convex) to dark gray (concave) as drawn by GRASP (Nicholls and Honig, 1991).

interactions between their constituting modules may play a key role in determining the global folding of the repeats in the complete enzyme. Hence, it has to be seen if the superhelical fold of C-LytA was preserved in the full-length amidase.

Choline Recognition

Taking into account the structural requirements for biological recognition of choline derivatives (Dougherty and Stauffer, 1990) and considering the crystal structure of the shortened C-LytA fragment in complex with choline (Fernández-Tornero et al., 2001), four potential choline binding sites could be predicted in Cpl-1. However, the crystal structure of Cpl-1 complexed with choline clearly showed that only the first two sites (formed by ChBrs p1-p2 and p2-p3) were really functional (Figure 1A). A deeper insight on Cpl-1 choline binding module structure showed that disruption of the superhelical fold to form the CII β sheet places Met301 within the third possible site (repeats p3-p4), and site 4 (repeats p4-p5) becomes hindered by Tyr287. Choline is located at the interface of two consecutive repeats in such a way that three structurally conserved aromatic residues (two Trp and one Tyr) form a cavity in which choline methyl groups are placed (Figure 4A). In addition, a Lys residue caps the site and could stabilize the phosphate group of phosphoryl-choline through one H bond (Figure 4A). Reported changes in Cpl-1 CD spectra induced by choline (Sanz and García, 1990) are consistent with the environmental change undergone by the aromatic residues interacting with the ligand and do not suggest any strong conformational change that would support the structural reorganization of the CII domain upon choline binding in solution.

The Cpl-1 structure leads to a more accurate description of the ChBr that will comprise a β hairpin plus a coiled region (Figure 1C). Thus, each choline binding site only requires the contribution of residues from two consecutive repeats, instead of the three previously

(B) Topology diagram of Cpl-1. Domains are color-coded as in (A) with the antiparallel β8 strand of the catalytic module highlighted in orange. In the choline binding model, the different ChBrs (p1–p6) are labeled.

(C) Comparison of the amino acid sequence of the ChBrs and the C-terminal tail in Cpl-1. The structural domain, the repeat number, and their corresponding amino acids are shown on the left. Residues in β strand conformation are shadowed in yellow. Conserved amino acids among the repeats appear in black boxes. Its consensus sequence (≥50%) is shown at the bottom. Upper case indicates 100% conservation.

Table 2. Enzymatic Activity of Wild-Type and Mutant Cpl-1 Lysozymes

Enzyme	Specific Activity (U/mg)	% Activity
Wild-type	824,000	100
D92A	1,700	0.2
D182A	8,500	1.0
E94A	<1	$<10^{-4}$
E94Q	<1	$<10^{-4}$

suggested (Fernández-Tornero et al., 2001). This subtle discrepancy was clearly caused by a different assignation of the ChBr sequence due to the lack of half of the p1 repeat in the crystallized fragment of C-LytA (see Figure 3) and not to the absence of the catalytic module.

Modular Organization in Cpl-1

Intermodular Interactions

The relative position of the catalytic and the choline binding modules is restrained by both the tightness of the linker and the interactions between the catalytic barrel and the CII structural domain. The linker extends from approximately residue 189, the end of the last β-barrel strand, to residue 199. Despite its flexibility (average B factor of 64 Å^2 for all atoms versus 34.6 Å^2 for average), the electron density of the linker defines both the backbone and most of its component side chains. The first residues comprise the DDEEDD motif that probably determines its extended conformation at neutral or basic pH. The linker has no contacts with either the catalytic or the choline binding modules, and there is only a van der Waals interaction between Lys195 and the choline binding module of a symmetry-related molecule. On the other hand, the interface between Cpl-1 modules is built up by a hydrophobic cavity, comprised between strands β6 to β8 of the catalytic module (Ala160, Phe162, Phe165, Trp174, Tyr176, Ile185, and Leu187), where the C-terminal end of the choline binding module inserts, shielding the hydrophobic interface from the solvent (Figure 5A). It is worth noting that this hydrophobic cavity is not present in the structure of cellosyl (Rau et al., 2001), a nonmodular lysozyme from the GH-25 family. This finding suggests that this hydrophobic interaction is genuine of the modular organization of Cpl-1. Moreover, hydrophobic interactions between the C-terminal ends of the choline binding modules seem to be also the driving force for C-LytA oligomerization in the absence of the catalytic module (Fernández-Tornero et al., 2002). Intermodular interactions in Cpl-1 are reinforced with a bifurcated salt bridge between modules (Figure 5A). The preservation of the tight intermodular interactions upon choline binding to Cpl-1 is confirmed by the structure of the Cpl-1:choline complex and agrees with previous DSC (differential scanning calorimetry) studies performed at increasing choline concentrations (Sanz et al., 1993).

Modular Organization Enhances Catalytic Activity

The specific recognition of choline by the cell wall binding module of pneumococcal lysins seems to be related to the presence of this molecule in the pneumoccocal envelope, a characteristic shared only by other related gram-positive bacteria. Acquisition of the choline binding module increases the Cpl-1 activity by 3–4 orders of magnitude (Sanz et al., 1992b), a value that is significantly higher than those usually reported for other modular glycoside hydrolases (Din et al., 1994). This suggests that the effect of Cpl-1 choline binding module is not just to localize the catalytic domain near the substrate, increasing its local concentration on the cell wall, but more importantly to orient it with the polysaccharide substrate in an optimal way, through the molecular restraints between modules. This hypothesis is supported by the Cpl-1 organization, since the C-terminal tail of the choline binding module might act as a hinge between modules with the linker restraining their relative positions. The Cpl-1 structure reveals that the catalytic site and the two functional choline binding sites have a defined orientation with distances of ~47 Å and ~37 Å between the catalytic site and the two ligand sites (see Figure 5B). These values correspond to, approximately, three repeats of the teichoic pentasaccharide unit (Klein et al., 1996). These findings suggest that the choline binding module also enhances Cpl-1's catalytic activity

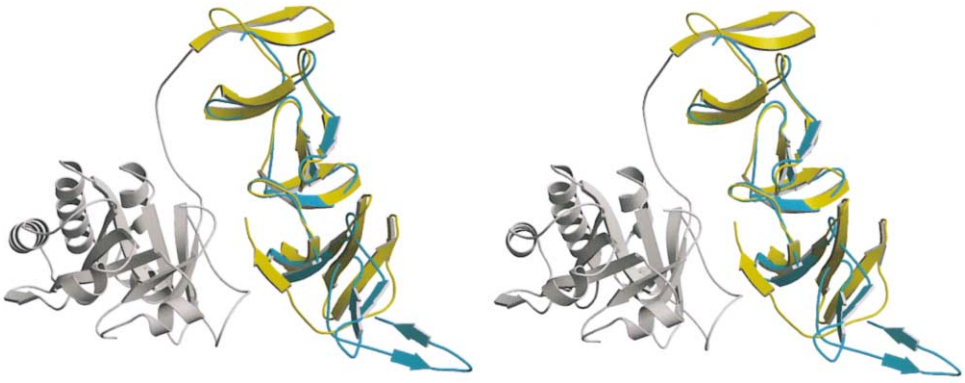

Figure 3. Stereo Representation of Structural Superposition of Cpl-1 and C-LytA

Only choline binding modules are colored: C-Cpl-1 is colored in yellow and C-LytA is colored in cyan. Note that structural superposition starts at p2 repeat, as p1 repeat was not present in the crystallized C-LytA structure. While there is a perfect superposition of the p2–p4 repeats, strong differences are observed in the p5–p6 and the C-terminal tail between the isolated choline binding module of LytA and the complete Cpl-1 structure.

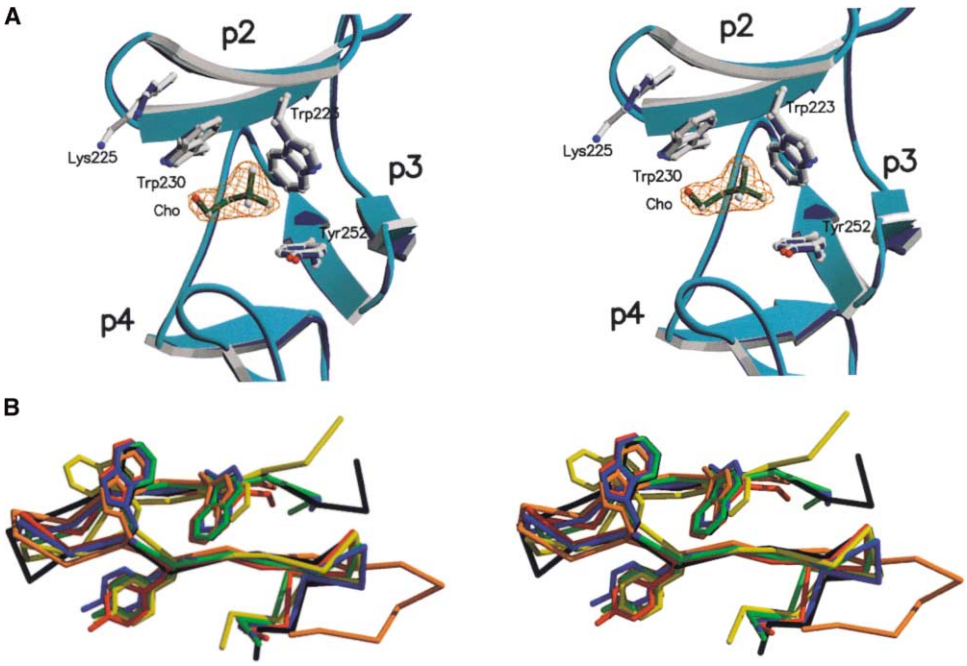

Figure 4. Choline Binding Site

(A) Stereo diagram of the second choline binding site showing the structural differences in the presence (cyan) and in the absence (dark blue) of choline; the choline molecule is highlighted in green. Residues that participate in phosphocholine stabilization are labeled (Met residue placed at the bottom of the cavity has been omitted for clarity). The $2F_o - F_c$ map (orange) of the choline molecule is contoured at 1.0σ.
(B) Backbone structural superposition of the six choline binding repeats. p1, black; p2, green; p3, blue, p4, red; p5, orange; and p6, yellow. Aromatic residues forming the choline binding site cavity are represented in a stick model.

by selecting those N-acetylmuramoyl-(β1,4),-glucosamine glycosid bonds that fulfill these geometrical constraints. As has been proposed for modular bacterial cellulases (Jervis et al., 1997), surface mobility of the absorbed Cpl-1 may arise because of the multiple binding contacts between the enzyme and the cell wall teichoic acids. In this sense, individual bonds are weak enough (K_d = 3.6 mM for choline binding; unpublished data) to permit diffusion across the surface, but the ensemble of bonds maintains the protein at the surface.

Versatility of the ChBrs

The analysis of the Cpl-1 structure shows that recognition between their modules implies the disruption of the superhelical fold in the choline binding module and the loss of three of the five choline binding sites affordable by six repeats. As discussed above, the selectivity of this endolysin to specifically target the pneumococcal cell wall together with the particular structure of the catalytic module might account for this reduction in choline binding ability. These facts suggest that the ChBrs are versatile supersecondary structures capable of playing different roles in the formation of choline binding sites or producing the interaction between modules, despite their similarity, both in sequence and supersecondary structure (Figures 1C and 4B). This versatility is further supported by the different structures of the choline binding modules of Cpl-1 and LytA (Figure 3). ChBr versatility seems to be crucial in the ChBP family, a group of surface proteins with 15 members identified in the genome of *S. pneumoniae* (Tettelin et al., 2001) and a large number of proteins from other gram-positive bacteria and their bacteriophages (http://www.Sanger.ac.uk/Software/Pfam/, accession number PF01473). Besides, pneumococcal choline binding proteins are involved in a wide range of different functions and present a great variability in the number of ChBrs (García et al., 1998). As observed in the Cpl-1 structure, ChBr versatility precludes the estimation of the number for choline binding sites without the determination of the complete modular structure. The structural variability observed in this ChBP with a relatively low number of ChBrs, only six, will probably increase in other members of this family with longer choline binding modules and different specific binding requirements.

Concluding Remarks

The elucidation of the 3D structure of Cpl-1 lysozyme, i.e., the first structure of a ChBP protein, paves the way to understanding how the proteins of this paradigmatic family that likely represents one of the best models to exemplify the concept of modular evolution are assembled. In light of the presented results, some important conclusions can be drawn.

(1) The 3D structure of Cpl-1 has indubitably established the postulated bimodular structure of ChBP enzymes that had been envisioned based only on biochemical and genetic analyses so far.

(2) After modular assembling, while the catalytic module keeps its independent robust shape, the structure of the choline binding module appears to suffer some significant rearrangements when compared to that of the isolated choline binding module of the LytA amidase.

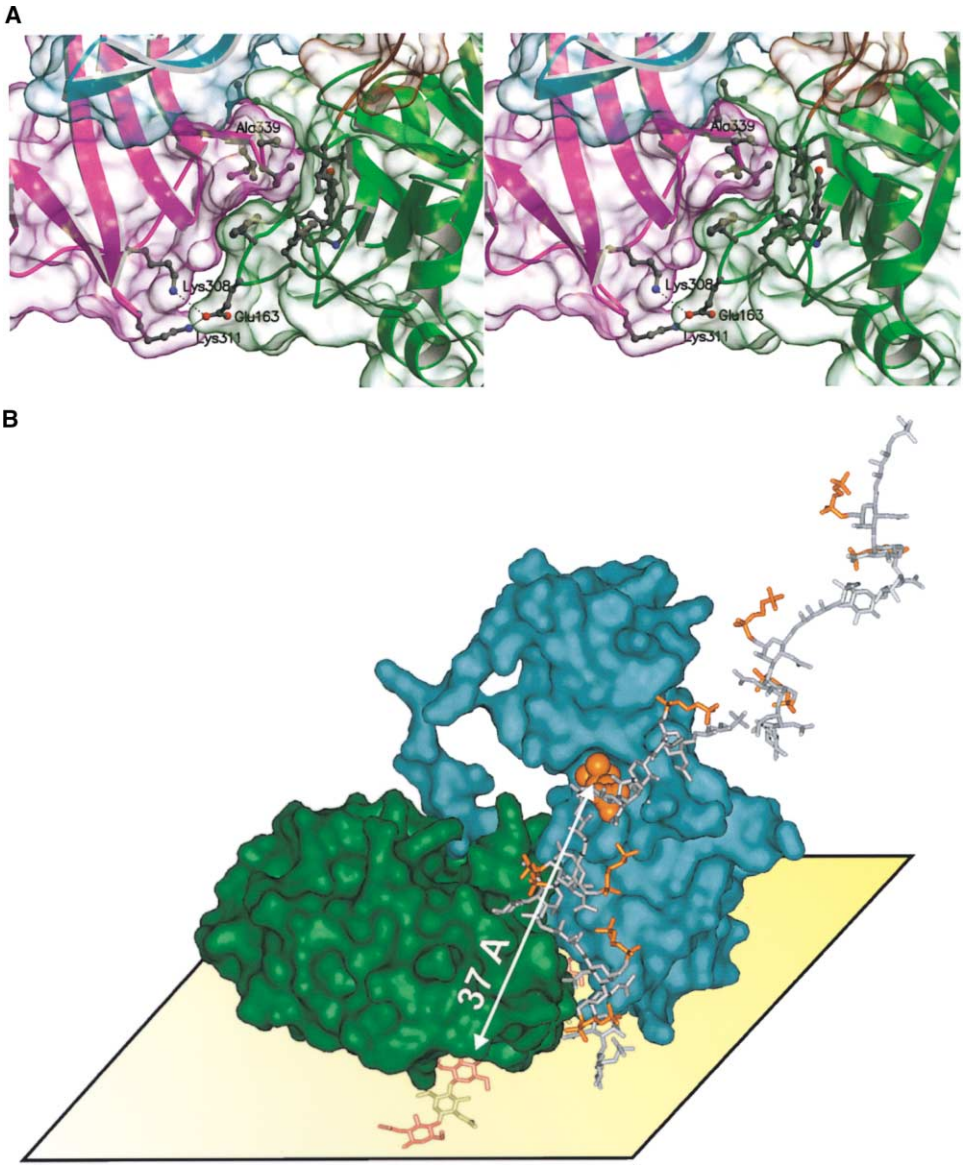

Figure 5. Intermodular Interactions and Cell Wall Recognition Mechanism

(A) Stereo diagram showing intermodular interactions in Cpl-1. Framework stabilization of Cpl-1 modules is produced by the creation of a hydrophobic core and a bifurcated salt bridge between the catalytic domain and the CII domain of the choline binding model. Residues that participate in these interactions are shown in a ball-and-stick representation. Domains and molecular surfaces are color-coded as in Figure 1.

(B) Proposed model of the selective recognition of pneumococcal cell wall by Cpl-1. The scheme shows Cpl-1 molecular surface (the catalytic module is colored in green, the linker and the choline binding module are colored in cyan) anchored on the peptidoglycan layer represented by a yellow square (only one glycan chain is depicted in the Cpl-1 active site). One pentameric teichoic acid chain (Klein et al., 1996) is represented (colored in gray with their phosphocholine moieties colored in orange), with one phosphocholine moiety (represented in cpk mode) being recognized by Cpl-1 through its choline binding site 2. The position of the teichoic acid chain was modeled manually as a rigid body onto the Cpl-1 structure.

These rearrangements are caused by its interaction with the catalytic module that significantly disturbs the shape of the most C-terminal region of Cpl-1, transforming the expected pattern of a left-handed superhelical structure into a new β sheet domain. Therefore, the fused choline binding module shows a more complex configuration than that observed in the isolated C-LytA module.

(3) Remarkably, the observed structural change of the choline binding module provokes a reduction of the available number of choline binding sites (two cholines per module) in contrast with the four sites found in C-LytA fragment structure.

(4) Based on the new complete structure of the Cpl-1 choline binding module, we have been able to reassign both the minimal structure of a choline binding site and the amino acids that form the characteristic repeats that built its primary structure. This new assignment is more coherent with the genetic evidence provided by compar-

ative evolutionary analyses than that previously suggested based on the interpretation of the incomplete C-LytA structure.

(5) The new Cpl-1 structure shows that the shape of the choline binding domain can be maintained in the absence of choline, in contrast with the previous assumption that it was completely dependent on choline presence.

(6) The binding of choline does not modify significantly the structure of the choline binding module, a result that is not in conflict with the previous observation that choline increases the stability of the enzyme and produces a dramatic change in CD spectrum. The interaction of choline with the aromatic rings rigidifies the protein and also explains that the CD changes are induced by modification of the spectra of aromatic amino acids and not by conformational alterations.

(7) The analysis of the 3D structure of Cpl-1 lysozyme together with the characterization of several mutants obtained by site-directed mutagenesis allowed us to assign, for the first time, the catalytic residues of a lysozyme from the G-25 family and postulate its most probable mechanism for cell wall hydrolysis. Besides, it would also account for the role played by cell wall acetylation in the lysozyme activity and settle the basis for constructing more active enzymes of the G-25 family with a broader antimicrobial spectrum that can be useful for pharmaceutical applications.

(8) The modular organization of Cpl-1 suggests that the choline recognition by choline binding module could help to orient the polysaccharide substrate in an optimal way within the catalytic cavity by selecting those glycosidic bonds that fulfill the geometrical constraints imposed by the linker and the intermodular contact surface.

Taking together these observations with the fact that the choline binding modules of the ChBPs not only contain a variable number of repeats but also can be positioned either in the C- or N-terminal regions of the protein, it can be envisioned that the peculiar left-handed superhelical structure that arises from the primary sequence of the module will be readapted to more complex structural patterns depending of the specific interactions with the corresponding functional module of each ChBP. The versatility of this adaptable structure would explain why it has been adopted by different functional modules of this protein family to fulfill specific roles in pneumococci since it provides many advantages for maintaining a close interaction with the cell envelope without perturbing the shape and function of the active module.

Experimental Procedures

Expression, Purification, and Activity of the Lysozymes

Cpl-1 lysozyme was purified from the cell extracts of *E. coli* DH1 (pCIP100) by affinity chromatography on DEAE-cellulose according to the method described elsewhere (Sánchez-Puelles et al., 1990). Assays for cell wall lysozyme activity were carried out according to standard conditions described elsewhere using as substrate pneumococcal cell walls that were radioactively labeled with [*methyl*-^{3}H] choline (Mosser and Tomasz, 1970). The homogeneity of the protein was confirmed by analytical ultracentrifugation and by SDS-polyacrylamide gel electrophoresis. Fractions containing Cpl-1 lysozyme were pooled together and extensively dialyzed against 20 mM Tris HCl buffer (pH 8.0). Then, the enzyme was concentrated at 277 K with a 10 kDa cutoff Amicon protein concentrator (YM-10) to approximately 13 mg/ml. The final protein concentration was determined spectrophotometrically, using a molar absorption coefficient of 117,350 M^{-1} cm^{-1} at 280 nm. Cpl-1 mutants (D92A, D182A, E94A, and E94Q) were constructed by PCR site-directed mutagenesis (Sambrook and Russell, 2001). According to this protocol, we used oligonucleotides complementary to the ends of the *cpl1* gene, primer cplext5′ (5′-GAAAAGAACATATGGTTAAAAAGAATGATTTATTTG-3′) and primer cplext3′ (5′-CGCGGATCCTTATGCTACGGTTATAAGCC-3′) for the 5′ end and 3′ end, respectively, and two complementary oligonucleotides for each mutation: primer cplD92A5′ (5′-TACCTTGTATTGGCCTACGAGGACGAC-3′) and primer cplD92A3′ (5′-GTCGTCCTCGTAGGCCAATACAAGGTA-3′) for D92A mutant; primer cplD182A5′ (5′-GTAACCCGTTTGCCAAGAATATTGTACTG-3′) and primer cplD182A3′ (5′-CAGTACAATATTCTTGGCAAACGGGTTAC-3′) for D182A mutant; primer cplE94A5′ (5′-GTATTGGACTACGCGGACGACCCAAGC-3′) and primer cplE94A3′ (5′-GCTTGGGTCGTCCGCGTAGTCCAATAC-3′) for E94A mutant; and primer cplE94Q5′ (5′-GTATTGGACTACCAGGACGACCCAAGC-3′) and primer cplE94Q3′ (5′-GCTTGGGTCGTCCTGGTAGTCCAATAC-3′) for E94Q mutant. After PCR amplification, the resulting *cpl1* mutant genes were cloned into plasmid pT7-7 and expressed, using 50 μM isopropyl-thio-β-D-galactopyranoside as inducer, in *E. coli* BL21 (DE3) strain. The Cpl-1 mutant enzymes were purified as described above. DNA sequences of the *cpl1* mutants were confirmed by the dideoxy chain-termination method with an automated Abi Prism 3700™ DNA sequencer (Applied Biosystems). All primers were synthesized on a Beckman model Oligo 1000M synthesizer.

Crystallization and Data Collection

The crystallization strategy followed to obtain suitable Cpl-1 crystals has been reported elsewhere (Monterroso et al., 2002). Briefly, crystals of Cpl-1 were grown using the hanging drop vapor diffusion technique and using protein solution containing 20 mM Tris HCl buffer (pH 8.0) and 8 mg/ml protein. Five microliters of reservoir solution (1.7 M sodium formate, 0.1 M sodium citrate buffer [pH 6.0]), 1 μl of 1.8 mM n-decyl-β-D-maltoside and 4 μl of protein solution were mixed and left to equilibrate at 20°C. Crystals appeared in 15–30 days and reached maximum dimensions of 0.5 × 0.3 × 0.3 mm^3. The crystals belong to the orthorhombic space group $C222_1$ with cell dimensions of $a = 77.949$, $b = 95.782$, $c = 129.282$ Å. A nonisomorphous Hg derivative was obtained by soaking the crystal in the mother liquor with 5 mM of phenyl mercury acetate for 24 hr. Native and SAD data sets were collected from frozen crystals at 100 K using a CCD detector on beamline BM14 at the ESRF (European Synchrotron Radiation Facility). Wavelengths for optimal data collection were determined using X-ray fluorescence at the absorption edge scan using a single crystal. Image data were processed and scaled using the programs MOSFLM (Leslie, 1987) and SCALA from the CCP4 package (CCP4, 1994). Crystals of Cpl-1:choline complex were obtained by soaking (15 min) the native crystals in the mother liquor with 10 mM choline chloride. This ligand concentration provides more than 95% saturation of choline binding sites in solution (data not shown).

Structure Determination and Refinement

The native crystal structure was determined by the SAD technique using a nonisomorphous Hg derivative. Initial Hg sites were observed in anomalous Patterson maps, and four sites were identified using SOLVE (Terwilliger and Berendzen, 1999). Initial phases were calculated at 2.8 Å resolution using the program SHARP (De la Fortelle and Bricogne, 1997) and two new Hg sites were added. Refined phases were solvent-flattened using RESOLVE (Terwilliger and Berendzen, 1999) on the basis of a solvent content of 50% (one molecule per asymmetric unit). Electron density maps clearly showed elements of secondary structure. Most of protein residues in the two modules were modeled manually at this stage by using the program O (Jones et al., 1991). However, the C-terminal tail, the linker, and some loops were not visible on this map. Molecular replacement procedure was carried out with the program AMoRe (Navaza, 1994), using native data up to 3.5 Å, and the previously constructed Cpl-1 modules used as independent search models. An unambiguous single solution for the rotation and translation func-

tions was obtained. Additional structure elements were identified and built manually (Jones et al., 1991). Further refinement including water molecules was performed using the program CNS (Brunger et al., 1998). The native structure was refined up to 2.1 Å with R of 20.7% and R_{free} of 25.9%. The resulting electron density was of great quality except for two side chains of the linker region. Crystal structure of the Cpl-1:choline complex was determined up to 2.4 Å resolution. No significant differences were observed between both native and complexed Cpl-1 structures except for a higher flexibility in the linker region of the Cpl-1:choline complex. Part of this region (residues 192–197) was modeled as Gly residues due to poor electron density definition. Structure determination parameters and refinement statistics for both structures are summarized in Table 1. Atomic coordinates and structure factors for Cpl-1 and Cpl-1:choline complex have been deposited in the PDB with accession codes 1h09 and 1oba, respectively.

Acknowledgments

We thank J. Sanz-Aparicio for helpful discussions and Roger A. Klein for kindly providing the atomic coordinates of the pentameric lipoteichoic acid. We also thank the staff of the BM14 beamline at ESRF (Grenoble) for support. This work was supported by grants BIO2000-1307 and BIO2002-02887 from Dirección General de Investigación and by grant of Contrato-Programa de Grupos Estratégicos (BMC2000-1002) de la Comunidad Autónoma de Madrid.

Received: May 5, 2003
Revised: June 18, 2003
Accepted: June 27, 2003
Published: September 30, 2003

References

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.

CCP4 (Collaborative Computational Project 4) (1994). The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.

Cleland, W.W., Frey, P.A., and Gerlt, J.A. (1998). The low barrier hydrogen bond in enzymatic catalysis. J. Biol. Chem. *273*, 25529–25532.

De la Fortelle, E., and Bricogne, G. (1997). Maximum-likelihood heavy atom parameter refinement in the MIR an MAD methods. Methods Enzymol. *276*, 472–494.

Din, N., Damude, H.G., Gilkes, N.R., Miller, R.C., Warren, R.A.G., and Kilburn, D.G. (1994). C_1-C_x revisited: intramolecular synergism in a cellulase. Proc. Natl. Acad. Sci. USA *87*, 11383–11387.

Dougherty, D.A., and Stauffer, D.A. (1990). Acetylcholine binding by a synthetic receptor: implications for biological recognition. Science *250*, 1558–1560.

Fernández-Tornero, C., López, R., García, E., Giménez-Gallego, G., and Romero, A. (2001). A novel solenoid fold in the cell wall anchoring domain of the pneumococcal virulence factor LytA. Nat. Struct. Biol. *8*, 1020–1024.

Fernández-Tornero, C., García, E., López, R., Giménez-Gallego, G., and Romero, A. (2002). Two new crystal forms of the choline-binding domain of the major pneumococcal autolysin: insights into the dynamics of the active homodimer. J. Mol. Biol. *321*, 163–173.

García, E., García, J.L., García, P., Arrarás, A., Sánchez-Puelles, J.M., and López, R. (1988). Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. Proc. Natl. Acad. Sci. USA *85*, 914–918.

García, J.L., Sánchez-Beato, A.R., Medrano, F.J., and López, R. (1998). Versatility of choline-binding domain. Microb. Drug Resist. *4*, 25–36.

Henrissat, B., and Bairoch, A. (1993). New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. *293*, 781–788.

Holm, L., and Sander, C. (1996). Mapping the protein universe. Science *273*, 595–603.

Jervis, E.J., Haynest, C.A., and Kilburn, D.G. (1997). Surface diffusion of cellulases and their isolated binding domains on cellulose. J. Biol. Chem. *272*, 24016–24023.

Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.

Klein, R.A., Hartmann, R., Egge, H., Behr, T., and Fischer, W. (1996). The aqueous solution structure of a lipoteichoic acid from *Streptococcus pneumoniae* strain R6 containing 2,4-diamino-2,4,6-trideoxy-galactose: evidence for conformational mobility of the galactopyranose ring. Carbohydr. Res. *281*, 79–98.

Kristinsson, K.G. (1997). Effect of antimicrobial use and other risk factors on antimicrobial resistance in pneumococci. Microb. Drug Resist. *3*, 117–123.

Leslie, A.G.W. (1987). Profile fitting. In Proceedings of the CCP4 Study Weekend, J.R. Machin and M.Z. Papiz, eds. (Warrington, UK: SERC Daresbury Laboratory), pp. 39–50.

Loeffler, J.M., Nelson, D., and Fischetti, V.A. (2001). Rapid killing of *Streptococcus pneumoniae* with a bacteriophage cell wall hydrolase. Science *294*, 2170–2172.

López, R., García, E., García, P., and García, J.L. (1997). The pneumococcal cell wall degrading enzymes: a modular design to create new lysins? Microb. Drug Resist. *3*, 199–211.

Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001). The kappa-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases. Structure *9*, 513–525.

Monterroso, B., Albert, A., Martínez-Ripoll, M., García, P., García, J.L., Menéndez, M., and Hermoso, J.A. (2002). Crystallization and preliminary X-ray diffraction studies of the complete modular endolysin from Cp-1, a phage infecting *Streptococcus pneumoniae*. Acta Crystallogr. D *58*, 1487–1489.

Monzingo, A.F., Marcotte, E.M., Hart, P.J., and Robertus, J.D. (1996). Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. Nat. Struct. Biol. *3*, 133–144.

Mosser, J.L., and Tomasz, A. (1970). Choline-containing teichoic acid as a structural component of pneumococcal cell wall and its role in sensitivity to lysis by an autolytic enzyme. J. Biol. Chem. *245*, 287–298.

Navaza, J. (1994). AMoRe: an automated package for molecular replacement. Acta Crystallogr. A *50*, 157–163.

Nelson, D., Loomis, L., and Fischetti, V.A. (2001). Prevention and elimination of upper respiratory colonization of mice by group A streptococci by using a bacteriophage lytic enzyme. Proc. Natl. Acad. Sci. USA *98*, 4107–4112.

Nicholls, A., and Honig, B.A. (1991). Rapid finite difference algorithm, utilizing succesive over-relaxation to solve the Poisson-Boltzmann equation. J. Comp. Chem. *12*, 435–445.

Pelton, S.I. (2000). Acute otitis media in the era of effective pneumococcal conjugate vaccine: will new pathogens emerge? Vaccine *19* (*Suppl. 1*), S96–S99.

Rau, A., Hogg, T., Marquardt, R., and Hilgenfeld, R. (2001). A new lysozyme fold. Crystal structure of the muramidase from *Streptomyces coelicolor* at 1.65 Å resolution. J. Biol. Chem. *276*, 31994–31999.

Sambrook, J., and Russell, D.W. (2001). Molecular Cloning: A Laboratory Manual, Third Edition (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press).

Sánchez-Puelles, J.M., Sanz, J.M., García, J.L., and García, E. (1990). Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. Gene *89*, 69–75.

Sanz, J.M., and García, J.L. (1990). Structural studies of the lysozyme coded by the pneumococcal phage Cp-1. Conformational changes induced by choline. Eur. J. Biochem. *178*, 409–416.

Sanz, J.M., García, P., and García, J.L. (1992a). Role of Asp-9 and Glu-36 in the active site of the pneumococcal CPL1 lysozyme: an evolutionary perspective of lysozyme mechanism. Biochemistry *31*, 8495–8499.

Sanz, J.M., Díaz, E., and García, J.L. (1992b). Studies on the structure and function of the N-terminal domain of the pneumococcal murein hydrolases. Mol. Microbiol. *6*, 921–931.

Sanz, J.M., García, J.L., Laynez, J., Usobiaga, P., and Menéndez, M. (1993). Thermal stability and cooperative domains of CPL1 lysozyme and its NH2- and COOH-terminal modules. Dependence on choline binding. J. Biol. Chem. *268*, 6125–6130.

Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. Acta Crystallogr. D *55*, 849–861.

Tettelin, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J., et al. (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science *293*, 498–506.

Tews, I., van Scheltinga, A.C.T., Perrakis, A., Wilson, K.S., and Dijkstra, B.W. (1997). Substrate-assisted catalysis unifies two families of chitinolytic enzymes. J. Am. Chem. Soc. *119*, 7954–7959.

Varea, J., Saiz, J.L., López-Zumel, C., Monterroso, B., Medrano, F.J., Arrondo, J.L.R., Iloro, I., Laynez, J., García, J.L., and Menéndez, M. (2000). Do sequence repeats play an equivalent role in the choline-binding module of pneumococcal LytA amidase? J. Biol. Chem. *275*, 26842–26855.

Vollmer, W., and Tomasz, A. (2000). The *pgdA* gene encodes for a peptidoglycan N-acetylglucosamine deacetylase in *Streptococcus pneumoniae*. J. Biol. Chem. *275*, 20496–20501.

Wang, Q., Graham, R.W., Trimbur, D., Warren, R.A.J., and Withers, S.G. (1994). Changing enzymatic reaction mechanisms by mutagenesis: conversion of a retaining glucosidase to an inverting enzyme. J. Am. Chem. Soc. *116*, 11594–11595.