Role of Novel Choline Binding Proteins in Virulence of *Streptococcus pneumoniae*

KHOOSHEH K. GOSINK, ELIZABETH R. MANN, CHRIS GUGLIELMO, ELAINE I. TUOMANEN,* AND H. ROBERT MASURE

Department of Infectious Diseases, St. Jude Children’s Research Hospital, Memphis, Tennessee 38105

Received 13 March 2000/Returned for modification 16 May 2000/Accepted 5 July 2000

The choline binding proteins (CBPs) are a family of surface proteins noncovalently bound to the phosphorylcholine moiety of the cell wall of *Streptococcus pneumoniae* by a conserved choline binding domain. Six new members of this family were identified, and these six plus two recently described cell wall hydrolases, LytB and LytC, were characterized for their roles in virulence. CBP-deficient mutants were constructed and tested for adherence to eukaryotic cells, colonization of the rat nasopharynx, and ability to cause sepsis. Five CBP mutants, CbpD, CbpE, CbpG, LytB, and LytC, showed significantly reduced colonization of the nasopharynx.

For CbpE and -G this was attributable to a decreased ability to adhere to human cells. CbpG, a putative serine protease, also played a role in sepsis, the first observation of a pneumococcal virulence determinant strongly operative both on the mucosal surface and in the bloodstream.

*Streptococcus pneumoniae* is currently the major invasive pathogen of children. Pneumococci attach to nasopharynx, lung, and vascular endothelial cells and invade causing pneumonia, bacteremia, and meningitis (23, 28). The pneumococcus has no fimbriae, like gram-negative organisms, and no fibrils, like other streptococci (28). The mechanism by which such a bald surface interacts with human cells is likely to be novel. The surface of *S. pneumoniae* is decorated with proteins that are covalently and noncovalently attached to the cell wall. These proteins fall into three classes. One well-characterized family of surface proteins, found in both streptococci and staphylococci, employs an LPXTGE motif that serves both as a cleavage site and an anchor for covalent attachment of a secreted protein to the cell wall (22). There are relatively few proteins with this motif in the pneumococcal genome. A second family consists of surface lipoproteins containing an LXXC motif in the N terminus that is cleaved and covalently attached to palmitic acid in the membrane (14, 17). Several members of this family of proteins have been implicated in pneumococcal pathogenesis (17, 18, 31). The most unique group of cell wall-associated proteins in pneumococci are the choline binding proteins (CBPs).

The pneumococcus contains phosphorylcholine on both the cell wall teichoic acid and the membrane-associated lipoteichoic acid (26). The presence of choline in the cell wall was thought to be unique to *S. pneumoniae*. However, recent data have indicated the presence of choline on the surface of a number of other respiratory tract pathogens, i.e., *S. oralis, S. mitis, S. constellatus, Clostridium* strain NI-4, *C. beijerinckii*, *Neisseria meningitidis, Pseudomonas aeruginosa*, and *Hemophilus influenzae* (9, 32, 33). In pneumococci, the CBPs bind to the phosphorylcholine of the cell wall noncovalently through a choline binding domain consisting of 2 to 10 repeats of a 20-amino-acid sequence (8, 12, 34). This choline binding motif, first described for pneumococci, has now been identified in other exported proteins, including toxins A and B of *C. difficile* (2, 6, 30), CspA of *C. beijerinckii* (21), glucan binding protein of *S. mutans*, and glycosyltransferases of both *S. mutans* and *S. downei* (1, 7, 13, 24, 29). It has been proposed that this domain forms a small ligand binding domain (34).

In pneumococci, six surface proteins that bind the phosphorylcholine moiety of the cell wall through their choline binding domain have been identified. The major autolysin of pneumococcus, LytA, was the first such protein characterized (9). LytA is required for daughter cell separation and pneumococcal lysis in stationary phase as well as in the presence of penicillin. It has been variably implicated to affect virulence by enabling release of the intracellular toxin pneumolysin (3, 5). Two other cell wall hydrolases, LytB and LytC, have recently been described, but their roles in virulence have not been assessed. LytB plays a role in pneumococcal daughter cell separation (10). LytC is reported to have lysozyme-like activity at 30°C (11). PepA was cloned recently and is thought to be involved in protein-protein and protein-lipid interactions (20). PepA is a 65-kDa protein that decreases complement deposition on the bacterial surface during sepsis (25, 27). Finally, CbpA (SpSA), the largest and most abundant of the CBPs, functions as a cell surface adhesin and plays a major role in colonization of the nasopharynx in the infant rat model (19). CbpA has also been shown to bind the secretory component of immunoglobulin A and the complement protein C3 (15; B. L. Smith, Q. Cheng, and M. K. Hostetter, Abstr. 98th Gen. Meet. Am. Soc. Microbiol., abstr. D122, 1998). CbpA, LytA, and PepA are subject to phase variation, CbpA and LytA being expressed strongly on mucosal surfaces whereas PepA is upregulated in the bloodstream. A surface-exposed virulence determinant operative in both sites would be a favored potential pneumococcal vaccine candidate.

Previous studies have indicated that many proteins can be eluted from the pneumococcal surface with soluble choline (4, 8, 19, 35). This is confirmed by a search of the pneumococcal genome with the sequence of the conserved choline binding domain. We sought to characterize the family of CBPs with respect to participation in colonization of the nasopharynx and in the pathogenesis of sepsis. In this report, we define a newly recognized role in colonization for two cell wall hydrolases and describe one new CBP, CbpG, active both on the mucosal surface and in the bloodstream.
TABLE 1. Primers used in PCR-based cloning into the E. coli expression system and in insertion duplication mutagenesis

<table>
<thead>
<tr>
<th>CBP gene</th>
<th>5' Primer</th>
<th>3' Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>hlyB</td>
<td>gacctggatccTATGATGTTACCTTGGCAAGAAAAACAG</td>
<td>acctgctgacATCTTTGCACACCATGTCTCATTTYG</td>
</tr>
<tr>
<td>hlyC</td>
<td>gacctggatccTGTGCCGAATAAAGCTTATGC</td>
<td>gactaaagtTATCCAAACGGCTGACATCCTGGC</td>
</tr>
<tr>
<td>cbpD</td>
<td>ggaatcttcATTAAATTTTTTGAATGCAAGAAG</td>
<td>tcccgagggagGTGAGACAAAGGCTTACA</td>
</tr>
<tr>
<td>cbpE</td>
<td>gacctggatccGAATGTTCCGCTCAAGAAGTGGATTG</td>
<td>gactaaggctTTCCTGCAGTGGAGCAAAGATTAC</td>
</tr>
<tr>
<td>cbpF</td>
<td>gacctggatccATTCCGATGATCATGGAAGATGG</td>
<td>tcacggatccTATACACCCTTCCGAGATTATT</td>
</tr>
<tr>
<td>cbpG</td>
<td>cgccgatccCGCCGCTGATGGGAAATTCCACTAC</td>
<td>tcccgagggagACATTAAATCCACTCA</td>
</tr>
<tr>
<td>cbpI</td>
<td>ctgactgatcGGGATTGGCAGCTTTAAAATC</td>
<td>caggtcagttGTTACCCATTCACCATTACC</td>
</tr>
<tr>
<td>cbpJ</td>
<td>tcgactggatccGGGATTGGCAGCTTTAAAATC</td>
<td>caggtcagttGTTACCCATTCACCATTACC</td>
</tr>
</tbody>
</table>

*Lowercase letters indicate added linkers used for cloning. cebp and cebpE primers were designed such that restriction sites were incorporated into the coding sequence by changing two of the bases (lowercase letters). Sequences used for insertion duplication mutagenesis are italicized.

**Nucleotide sequence accession numbers.** Accession numbers for the CBP sequences are AF278686 (CbpD), AF278687 (CbpE), AF278688 (CbpF/G), AF278689 (CbpI), and AF278690 (CbpJ).

**MATERIALS AND METHODS**

Bacterial strains and plasmids. *S. pneumoniae* Norway type 4 is a clinical isolate obtained from MedImmune Inc., Gaithersburg, Md. An uncapsulated derivative, 4R, was obtained from Rodger Novak, St. Jude Children’s Hospital. Strain R6 was obtained from the Rockefeller University collection. Cultures were grown without aeration at 37°C in 5% CO₂ in a defined semisynthetic medium (C+Y medium) (15a) or plated on tryptic soy agar supplemented with 3% (vol/vol) sheep blood. Pneumococci with integrated plasmids were grown in the presence of appropriate antibiotics (erythromycin [1 μg/ml] and/or chloramphenicol [5 μg/ml]).

Recombinant protein expression, purification, and antibody production. DNA techniques including PCR, plasmid isolation, chromosomal DNA purification, restriction endonuclease digestion, ligation, and transformation were done according to standard protocols (16, 18). For all CBPs, primers were designed to the N terminus and the C terminus of each CBP gene and used to amplify DNA fragments by PCR (Table 1). PCR reactions were performed according to the Qiagen Taq DNA polymerase protocol as follows: 94°C for 3 min, then 25 cycles at 94°C for 1 min, 52°C for 1 min, and 72°C for 1.5 min, and a final extension at 72°C for 10 min. The PCR products were digested with BamHI and Smal, inserted into pQE30 digested with BamHI and Smal, and transformed into the E. coli host strain M15[pREP4], (Qiagen). Clones were verified by nucleotide sequencing in St. Jude Children’s Hospital Biotechnology Center.

Recombinant proteins were purified according to the protocol provided by Qiagen. *Escherichia coli* strains with recombinant plasmids were grown to 100 ml of Luria broth with ampicillin (100 μg/ml) and kanamycin (25 μg/ml) to an optical density at 600 nm (OD₆₀₀) of 0.8 to 0.9 at 37°C with vigorous shaking. Cultures were induced with 1 mM isopropyl-β-D-thiogalactopyranoside for 3 h. Cells were harvested by centrifugation at 3,800 × g for 15 min. Pellets were resuspended in lysis buffer (6 μM guanidine, 0.1 M Na₂HPO₄, 0.01 M Tris, 10 mM imidazole [pH 8.0]) and lysed overnight at 4°C. The lysate was centrifuged at 10,000 × g for 30 min at room temperature; 1.5 ml of Ni-nitrilotriacetic acid resin was added to the supernatant and gently mixed by shaking for 1 h. The lysate-resin mixture was loaded onto a 1-ml column and washed with 10 to 15 column volumes of wash buffer (8 M urea, 0.1 M Na₂HPO₄, 0.01 M Tris, 20 mM imidazole [pH 8.0]). Recombinant protein was eluted four times with 0.5 ml of elution buffer (8 M urea, 0.1 M Na₂HPO₄, 0.01 M Tris, 500 mM imidazole [pH 8.0]). Protein was quantitated using the Bio-Rad Bradford protein assay, with bovine serum albumin as a standard. Purified recombinant proteins were run on sodium dodecyl sulfate (SDS) 10% to 12% polyacrylamide gels, and bands corresponding to the proteins were cut out and used for production of antisera. Antisera were generated in rabbits by Covance Inc. (Denver, Pa.).

Western blot analysis. Native CBPs were purified as previously described (19, 35). Pneumococcal cultures (100 to 400 ml) were grown to an OD of 0.4 to 0.6 and centrifuged at 3,000 × g for 10 min. Cells were washed once in phosphate-buffered saline (PBS), and CBPs were eluted by mixing the cells with 5 to 10 ml of PBS containing 2% choline and gently shaking at room temperature for 20 min. The eluate was dialyzed overnight against PBS at 4°C and concentrated on a Centriplus 10 concentrator (Amicon). The Bio-Rad protein assay was used to determine protein concentration. Ten microliters of eluate from 10³ cells was added to 2 μl of 5X loading dye; the sample was boiled at 100°C for 5 min and then loaded on a precast SDS–4 to 15% polyacrylamide gel (Bio-Rad). Following separation by gel electrophoresis, samples were transferred to Immobilon-P (Millipore Corp., Bedford, Mass.) and probed with the antisera raised against individual CBPs at a dilution of 1:5,000 to 1:10,000. Bands were visualized following incubation with peroxidase-conjugated goat anti-rabbit sera (diluted 1:8,000; Bio-Rad), using a chemiluminescence kit from Amersham.

**Insertion duplication mutagenesis.** CBP-deficient mutants were constructed by insertion duplication mutagenesis as previously described (18). PCR was used to amplify 200- to 457-bp fragments from the N-terminal domains of these genes (Table 1). EcoRI and BamH1 sites were introduced at the ends of the primers and used to clone the PCR fragments into the pIDC9 vector. For cebp, cbpE, and cebpD, fragments spanned amino acids 74 to 217, 7 to 114, and 262 to 379, respectively. For cbpB, cbpC, cbpI, and cebpG, the amplified DNA fragments corresponded to amino acid residues 7 to 89, 29 to 122, 20 to 128, 10 to 95, and 15 to 87, respectively. PCR fragments were digested with either EcoRI or EcoRI and BamH1, ligated into pIDC9 digested with either EcoRI or EcoRI and BamH1, and transformed into E. coli. Single transformants containing the insert were identified and plasmid DNA from these clones was transformed into pneumococcal strains serotype 4 and type 4R. Chromosomal integration of the vector at the right locus was verified by PCR, using primers homologous to plasmid sequences (M13 forward –21 and reverse primers) and to sequences upstream of the point of insertion of the plasmid.

**In vitro adhesion assays.** Adherence to Detroit nasopharyngeal cells was assayed as previously described (19). Pneumococci were grown to an OD₆₀₀ of 0.45, centrifuged, resuspended in 0.5 ml of carbonate buffer (0.05 M sodium carbonate, 0.1 M sodium chloride), and labeled with fluorescein isothiocyanate (1 mg/ml; Sigma) for 30 min at room temperature. Labeled bacteria were washed three times in PBS and dialyzed to 10⁷ CFU/ml in PBS. Cell monolayers established in Terasaki plates were incubated with fluorescein isothiocyanate-labeled bacteria (10⁷/g/ml) for 30 min at 37°C. Plates were washed four times with PBS and fixed with 2.5% glutaraldehyde. Then adherent bacteria were counted. Each strain was tested in six wells per experiment, and experiments were repeated four independent times.
proteins of size from 426 to 2,034 bp and were predicted to encode for except for. The genes were diversely located throughout the chromosome have a proline-rich linker region or a common CBP promoter. Additionally, six other contigs containing seven open reading frames with sequence homology to the choline binding domain 175, respectively. The degree of similarity between the choline repeats were located in the C-terminal domain of all CBPs proteins had between 2 and 10 choline binding repeats. These codons in all frames, suggesting that it did not encode a pro-

Animal models. Nasopharyngeal colonization of 1- to 5-day-old infant rats by the Norway T4 parental strain and CBP-deficient mutants was performed as follows. For each experiment, 8 to 10 rat pups were inoculated intranasally with 2.5 × 10^8 to 8 × 10^9 CFU of the CBP-deficient mutants or the isogenic parent in PBS. Colonization was assessed at 48 and 96 h postinoculation. The fluid from the nasal washes was diluted and plated, and colony counts were determined. For each mutant, the experiment was repeated at least three independent times. To assess virulence in a model of sepsis, 2- to 5-day-old infant rats were injected intraperitoneally with 2 × 10^8 to 4 × 10^9 CFU of the parental strain or a CBP-deficient mutant. Survival was assessed at 24 and 48 h postinjection.

RESULTS

Identification and cloning of CBP DNA fragments. A 180-amino-acid sequence corresponding to the C-terminal choline binding region of CbpA (amino acids 514 to 694 of S. pneumoniae type 4 strain) was used to search the partially completed genome of a Norway type 4 strain of S. pneumoniae, utilizing the National Center for Biotechnology Information BLAST search engine (www.ncbi.nlm.nih.gov). The search sequence contained all of the choline binding repeats of CbpA but did not include the proline-rich region or any N-terminal sequence. This search identified six discrete contigs containing previously identified cbpA, lytA, lytB, lytC, pspA, or pcpA. Additionally, six other contigs containing seven open reading frames with sequence homology to the choline binding domain of CbpA were identified (Table 2). The cbp genes ranged in size from 426 to 2,034 bp and were predicted to encode for proteins of ~20 to 80 kDa. The new CBPs did not appear to have a proline-rich linker region or a common CBP promoter. The genes were diversely located throughout the chromosome except for cbpF and cbpG. cbpG was located 13 bp upstream of cbpF on the same contig. cbpH had a large number of stop codons in all frames, suggesting that it did not encode a protein, and thus this locus was not studied further. The predicted proteins had between 2 and 10 choline binding repeats. These repeats were located in the C-terminal domain of all CBPs except for LytB and LytC (at amino acids 206 to 396 and 68 to 175, respectively). The degree of similarity between the choline binding domains of the CBPs and that of CbpA (the search sequence) ranged from 30 to 60%. The N-terminal domain of each protein was distinct, indicating that individual family members likely serve different functions. The diverse N-terminal functional domain of each of these predicted proteins was used individually to search the current protein databases. Significant homology was found only for CbpG, which showed ~56% similarity over 70% of the N-terminal sequence to a serine proteinase homolog of Enterococcus faecalis.

Expression of CBPs. A His tag expression system was used to overexpress and purify eight CBPs from E. coli: LytB, LytC, and the six new CBPs. CbpD (which proved lethal to E. coli in full length) and CbpG and were purified as N-terminal truncated forms missing the choline binding domains. Polyclonal antisera were raised against each recombinant protein. Although all of the CBPs harbor a choline binding domain, the antibodies did not cross-react among the CBPs, with one exception. Antibodies raised against recombinant CbpJ (rCbpJ) showed some cross-reactivity with rCbpE (Fig. 1), but this cross-reactivity was not observed with the native proteins. The

![FIG. 1. Recognition of recombinant and native CBPs by rCBP-specific antisera. The membrane was probed with antisera reactive with the indicated recombinant CBP: R, purified rCBP; C, CBP eluted from strain T4 by choline; K, CBP eluted from isogenic CBP-deficient strains. Sizes are indicated in kilodaltons.](https://journals.asm.org/journal/iai)
isolated choline binding domain was exceptionally poorly immunogenic, and its antisera failed to react with any native CBPs (data not shown), perhaps explaining the lack of cross-reactivity between CBP family members.

Antibodies raised against individual rCbpE, rCbpF, and rCbpJ and truncated rCbpD and rCbpG recognized native proteins in choline eluates of the wild-type Norway strain (Fig. 1). This indicated that these proteins were expressed and that they could be eluted from the bacterial surface with choline, consistent with membership in the CBP family. Conversely, antisera to a mixture of native CBPs (19) recognized rLytC, rCbpG, rCbpL, and rCbpJ (data not shown). However, native and recombinant proteins consistently migrated differently, suggesting possible structural or conformational differences. Antibodies against CbpE and CbpG recognized multiple bands which were specific, since mutants deficient in these CBPs lacked all bands. Antibodies against CbpD and CbpF recognized multiple bands, only one of which was specific, as evidenced by its absence in the isogenic knockout strain. No differences were found in the expression of any of these CBPs at 30 and 37°C or during competence (data not shown). Analysis of the expression of the CBPs over a growth cycle indicated that the expression of CbpE increased during logarithmic growth; the levels of other CBPs did not vary with growth phase (data not shown).

Analysis of CBP-defective mutants. To assess the in vivo roles of LytB, LytC, and the six new CBPs, mutants defective in expression of each CBP were constructed by insertion duplication mutagenesis. CBP-deficient mutants were constructed in both the Norway T4 strain and an isogenic, nonencapsulated derivative, T4R. All genes were distant from surrounding open reading frames, eliminating the possibility of polar effects of significance except for cbpF and cbpG. cbpG is located directly upstream of cbpF, and it is possible that mutations in cbpG are polar onto cbpF. However, since a mutant deficient in only CbpF had no observable phenotype, the phenotype observed for a CbpG-defective strain appears to be attributable to CbpG function alone. Furthermore, Western analysis of the CbpF- and CbpG-deficient mutants showed that the 25-kDa CBP was present but not detectable in the CbpG-defective strain (Fig. 2). It is possible that mutations in cbpG cause a loss of the open reading frame, leaving open the mechanism of secretion of this class of protein. All six new CBP family members reported here were surface expressed and eluted with choline, consistent with the behavior of known CBPs. Antibodies to recombinant CBPs recognized native CBPs eluted from pneumococci by choline. However, the native species uniformly migrated differently than the recombinant species, indicating possible differences in the folding or structure of the native versus recombinant proteins of this entire family.

Analysis of the CBP-deficient mutants indicated that a number of CBPs play a role in adhesion and colonization of the nasopharynx. Previous results have suggested that CbpA accounts for 40 to 50% of the adherence of wild-type bacteria to nasopharyngeal cells (19). A similarly strong phenotype was found for the putative serine protease, CbpG. This correlated with a strong defect in colonization of the nasopharynx in vivo. Mutants in CbpD and CbpE showed significant loss of nasopharyngeal colonization with only modestly decreased adherence in vitro at 30°C. Their primary function in colonization remains to be determined.

The role of cell wall hydrolases in virulence has long been sought because of the general belief that these suicidal enzymes must confer an in vivo advantage in order to be retained in the face of strong negative selection by penicillin. LytA is a well-characterized amidase triggered by treatment with penicillin. LytB has been designated a muramidase (10). LytC has been identified as the first streptococcal lysozyme (11). Deletion of any of the three hydrolases does not alter viability or transformability in vitro (10, 11, 26). Loss of function of LytA modestly affects the course of pneumonia but not sepsis (3, 5). This phenotype is thought to be due to the role of autolysis in the release of pneumolysin and neuraminidase, two well-described virulence determinants. The significant loss of ability to colonize the nasopharynx in both the LytB and LytC mutants with only modest changes in adherence in vitro raises the possibility that the two new hydrolases could also play this same role in toxin release. The suggestion that LytC functions optimally as a hydrolase at 30°C (11) is consistent with a potential role in the nasopharynx, where temperatures are cooler.

### Table 3. Effect of loss of CBP function on colonization of the infant rat nasopharynx

<table>
<thead>
<tr>
<th>CBP</th>
<th>% of WT CFU at 48 h (mean ± SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LytB</td>
<td>18 ± 8*</td>
</tr>
<tr>
<td>LytC</td>
<td>6 ± 5*</td>
</tr>
<tr>
<td>CbpD</td>
<td>43 ± 18*</td>
</tr>
<tr>
<td>CbpE</td>
<td>25 ± 12*</td>
</tr>
<tr>
<td>CbpF</td>
<td>67 ± 22*</td>
</tr>
<tr>
<td>CbpG</td>
<td>25 ± 12*</td>
</tr>
<tr>
<td>CbpJ</td>
<td>60 ± 30</td>
</tr>
<tr>
<td>6</td>
<td>68 ± 35</td>
</tr>
</tbody>
</table>

* 100% adherence = 243 ± 51 bacteria/10 μl of nasal wash. * P < 0.001 versus wild-type (WT) Norway strain.
PspA is the only previously known CBP with a dominant role in sepsis. The present analysis adds Cbp G to this group. CbpG appears to bear sequence similarity to a serine protease, the substrate of which may relate to its role in adherence to human cells. It is possible that CbpG modifies proteins on the surface of the pneumococcus, enabling them to bind to receptors on the eukaryotic cells. Conversely, CbpG may modify the eukaryotic cell surface, promoting ligand-receptor interactions.

In summary, CBPs appear to be a functionally significant group of pneumococcal surface proteins that are noncovalently and reversibly attached to the choline moiety of the cell wall through their C-terminal choline binding domain. These proteins have unique N-terminal domains that indicate varied functions for these proteins. Clearly a main function of the CBP family is in promoting colonization of the nasopharynx, as evidenced by the clear phenotype of CbpA, CbpD, CbpE, CbpG, LytB, and LytC mutants in the model of nasopharyngeal colonization. For CbpA and CbpG, the role in vivo is strongly implicated to be host cell recognition, and this may also apply to a much lesser extent for CbpB, CbpC, CbpD, and CbpE. Only PspA and CbpG are currently directly implicated in pneumococcal virulence in sepsis. The dual role of CbpG in sepsis is evidenced by the clear phenotype of CbpA, CbpD, CbpE, CbpG, LytB, and LytC mutants in the model of nasopharyngeal colonization.

### ACKNOWLEDGMENTS

This work was supported by NIH grants AI27913 and AI 39482, AI36445 and in part by Cancer Center Support CORE grant P30 CA21765 and the American Lebanese Syrian Associated Charities (ALSAC).

We are grateful to T. Wizemann, S. Johnson, and S. Koenig, MedImmune Inc., for helpful discussion. We thank Micha Ring and John Killmar for technical assistance.

### REFERENCES


ing unit in streptococcal glucosyltransferase glucan-binding regions and related sequences. J. Dental Res. 73:1133–1141.