JOURNAL OF BACTERIOLOGY, Feb. 1995, p. 1098–1103 Vol. 177, No. 4
0021-9193/95/$04.00+0

Tracking the Evolution of the Bacterial Choline-Binding Domain: Molecular Characterization of the *Clostridium acetobutylicum* NCIB 8052 *cspA* Gene

ANA R. SANCHEZ-BEATO, CONCEPCION RONDA, AND JOSE L. GARCIA*

Department of Molecular Microbiology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Madrid, Spain

Received 15 August 1994/Accepted 30 November 1994

The major secreted protein of *Clostridium acetobutylicum* NCIB 8052, a choline-containing strain, is CspA (clostridial secreted protein). It appears to be a 115,000-M_r glycoprotein that specifically recognizes the choline residues of the cell wall. Polyclonal antibodies raised against CspA detected the presence of the protein in the cell envelope and in the culture medium. The soluble CspA protein has been purified, and an oligonucleotide probe, prepared from the determined N-terminal sequence, has been used to clone the *cspA* gene which encodes a protein with 590 amino acids and an M_r of 63,740. According to the predicted amino acid sequence, CspA is synthesized with an N-terminal segment of 26 amino acids characteristic of prokaryotic signal peptides. Expression of the *cspA* gene in *Escherichia coli* led to the production of a major anti-CspA-labeled protein of 80,000 Da which was purified by affinity chromatography on DEAE-cellulose. A comparison of CspA with other proteins in the EMBL database revealed that the C-terminal half of CspA is homologous to the choline-binding domains of the major pneumococcal autolysin (LytA amidase), the pneumococcal antigen PspA, and other cell wall-lytic enzymes of pneumococcal phages. This region, which is constructed of four repeating motifs, also displays a high similarity with the glucan-binding domains of several streptococcal glycosyltransferases and the toxins of *Clostridium difficile*.

Phosphatidylcholine, a major membrane lipid in most eukaryotes, is found in only some prokaryotes (1). Interestingly, choline has also been described as a component of teichoic and lipoteichoic acids of a small number of bacteria (16). The functional role of choline in these bacteria has not been completely established. Most of what is currently known about choline has been collected from the study of *Streptococcus pneumoniae* and its bacteriophages. The choline-containing teichoic acids present in the cell walls of pneumococci participate in the regulation of the host autolytic enzymes and in the evolution of the pneumococcal phage lysins (19). A comparative study of the sequence of the major pneumococcal autolysin and those of several phage-encoded lysins has provided valuable information relative to their peculiar modular organization (12). These enzymes appear to have evolved by the fusion of two independent domains: the N-terminal domain, which contains the catalytic center and provides the enzymatic specificity, and the C-terminal one, which is responsible for the binding of the protein to specific components of the cell wall (27, 28). We have observed that the C-terminal domains of choline-dependent enzymes are homologous and contain six repeated motifs (12). The construction of different chimeric enzymes demonstrated that the acquisition of the choline-binding domain represents a noticeable evolutionary advantage for the enzymes that interact with a polymeric substrate, allowing them to achieve a higher catalytic efficiency (2, 4, 5). Hence, it has been suggested that the peculiar presence of choline in the pneumococcal cell wall can be explained by its role as an element of strong selective pressure for the choline-dependent murein hydrolases, preserving the existence of this characteristic choline-binding domain.

In a search for the evolutionary relationships among bacterial autolytic enzymes, we found that the supernatants obtained from cultures of *Clostridium acetobutylicum* CECT 806 (NCIB 8052) and the related strains *Clostridium saccharoperbutylacetonicum* NI-504 (ATCC 27022) and NI-4 had an enzyme capable of degrading pneumococcal cell walls containing choline, but not ethanolamine, in their teichoic acids (11). This activity was purified by affinity chromatography on choline-Sepharose or DEAE-cellulose columns (11). We also observed that the addition of 2% choline to the culture medium led to the formation of long chains of cells, as is already found with pneumococci (25). Podvin et al. (23) confirmed these results and demonstrated that choline was a component of the teichoic acids of *C. saccharoperbutylacetonicum* NI-4.

The aim of this work was to determine the evolutionary relationships between the choline-binding domains of the pneumococcal murein hydrolases and the major secreted protein of *C. acetobutylicum* NCIB 8052.

Characterization of the CspA protein. CspA is the most abundant protein in the culture medium of *C. acetobutylicum* NCIB 8052 and can be purified by affinity chromatography on DEAE-cellulose (11) (Fig. 1A). This preparation, which contained more than 95% CspA, showed a choline-dependent amidase activity on pneumococcal cell walls (11). Although we had postulated that it could be a cell wall-lytic enzyme, further chromatographic analyses revealed that the cell wall hydrolytic activity did not appear to be associated with the major component of this preparation. The CspA purified by electroelution (Bio-Rad electroeluter model 422) after sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (7, 18) was used to prepare a rabbit polyclonal antiserum (9). The polyclonal antibodies were not able to inhibit the cell wall-lytic activity on pneumococcal cell walls, suggesting that this protein

* Corresponding author. Mailing address: Department of Molecular Microbiology, Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Científicas, Velázquez 144, 28006 Madrid, Spain. Phone: 34-1-5611800. Fax: 34-1-5627518.

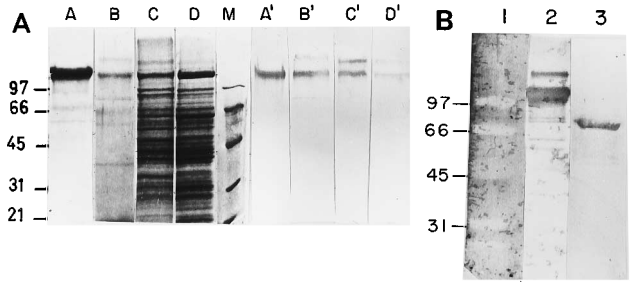

FIG. 1. Subcellular distribution and glycoprotein analysis of CspA. (A) Coomassie blue-stained profile (lanes A to D) and Western blot analysis (lanes A′ to D′) of the subcellular fractions. CspA was purified on DEAE-cellulose (lanes A and A′), culture medium (lanes B and B′), cellular envelope (lanes C and C′), and cytoplasmic fraction (lanes D and D′). The molecular masses of the protein markers (lane M) are expressed in kilodaltons. (B) SDS-PAGE of purified CspA. CspA was purified on DEAE-cellulose and stained with the DIG glycan detection kit. Lane 1, protein standards (negative control); lane 2, CspA; lane 3, transferrin (positive control). The numbers at the left are molecular masses in kilodaltons.

is probably not a precursor of the clostridial hydrolytic enzyme (data not shown). In addition, these antibodies were used for Western blot (immunoblot) analysis (24) of the CspA contents of different cell fractions of *C. acetobutylicum* NCIB 8052 cultured in TYA medium (11) (Fig. 1A). Although CspA was first detected in the culture medium, it is also an important component of the cell wall. In addition, this analysis revealed that the band with an M_r of 125,000 that was observed in Fig. 1A also reacts with these antibodies, suggesting that it might be a related protein.

To determine whether CspA could be glycosylated, the protein purified on DEAE-cellulose was stained with the DIG glycan detection kit (Boehringer, Mannheim, Germany) (Fig. 1B). CspA was positively stained, suggesting that it is a glycosylated protein. In addition, the carbohydrate content of the protein purified by electroelution after SDS-PAGE was determined by gas-liquid chromatography (13), which revealed the presence of 12% (wt/wt) rhamnose. However, the treatment of the protein with *N*-glycosidase or *O*-glycosidase (Boehringer) did not produce a reduction of the molecular weight (data not shown). Interestingly, Messner et al. (20) have described a new O-glycosidic linkage which was found in a surface layer protein of *Clostridium thermohydrosulfuricum* and which was not affected by treatment with *O*-glycosidase.

Cloning and sequencing of the *cspA* gene. The N-terminal amino acid analysis of the electroeluted CspA protein, the major band shown in Fig. 1A, rendered the sequence depicted in Fig. 2A. This sequence was used to synthesize two oligonucleotides, named O1 and O2 (Isogen, Amsterdam, The Netherlands) (Fig. 2A). Several genomic libraries of *C. acetobutylicum* NCIB 8052 were prepared in pUC18 (Pharmacia, Uppsala, Sweden) with chromosomal DNA (31) digested with different restriction enzymes and with *Escherichia coli* TG1 (Amersham, Little Chalfont, United Kingdom) as a host. These libraries were screened with the oligonucleotides O1 and O2 as DNA probes (26), and the recombinant plasmid pKIT10 (Fig. 2A) containing a 279-bp *Pst*I fragment was isolated and sequenced. This fragment encoded the N-terminal part of CspA. A new plasmid, named pKDI543 (Fig. 2C), containing three *Dra*I fragments was isolated from a *Dra*I library, with the *Pst*I fragment being used as a probe. Hybridization analyses (not shown) revealed that only the 2,052-bp *Dra*I fragment of pKDI543 encoded CspA, and it was subcloned into plasmid pKDE543 (Fig. 2A) and sequenced. Finally, a PCR strategy was used to clone a fragment containing the 5′ untranslated region of *cspA* and the initiation codon (Fig. 2B).

Nucleotide sequence of the *cspA* gene and analysis of the predicted amino acid sequence. The nucleotide sequence of *cspA* and the deduced amino acid sequence are shown in Fig. 3. The ATG codon is preceded by a potential ribosome-binding site (AGGAGG, positions −11 to −16) (3, 34). A sequence displaying similarity to that of the consensus *E. coli* promoter was observed 37 bp upstream of the start codon. The 5′ DNA region preceding the start codon is more AT rich than the *cspA* coding region (93 and 64%, respectively). A palindromic sequence able to form a hairpin loop with a calculated free energy of −17.9 kcal mol^{-1} was identified immediately downstream of the TAG stop codon. This structure is followed by a short T-rich sequence and likely functions as a transcription termination. The codon composition of the *cspA* gene showed a strong bias towards codons with predominantly A or U, as might be expected from the low G+C content of *C. acetobutylicum.*

Analysis of the deduced amino acid sequence revealed that the *cspA* gene consists of 1,774 bp and encodes a protein of 590 residues with a calculated M_r of 63,740 and an isoelectric point of 4.1. A comparison of the N terminus of the mature protein as determined by Edman degradation and the predicted N-terminal sequence deduced from the nucleotide sequence indicates that the protein is initially translated with a signal peptide of 26 amino acids. This signal peptide presents the typical characteristics of other bacterial signal peptides (29). Analysis of the peptide sequence as described by Kyte and Doolittle (17) indicated that, overall, the mature protein is hydrophilic. Eighteen percent of the protein is made up of acidic residues (Glu and Asp), and ten percent of it is made up of basic residues (Lys, Arg, and His). Only one cysteine was present in the deduced sequence. Four potential N-glycosylation sites (21) were found at positions 545, 558, 565, and 573 of the polypeptide chain (Fig. 3).

A comparison of the deduced amino acid sequence with those in the EMBL database revealed that the C-terminal part of the protein was very similar to those of the choline-binding domain of the pneumococcal major autolysin LytA, the cell wall-lytic enzymes of several pneumococcal phages (12, 19, 25), and the pneumococcal PspA antigen (33). However, the N-terminal part of the protein did not show a significant similarity with that of any protein in the database. A detailed analysis of the C-terminal region showed that it is composed of four repeated motifs plus a tail corresponding to an imperfect motif. These repeats present the consensus sequence (GWLKD NGSWYYLNANGAMAT) previously described for the choline-binding domain of pneumococcal cell wall hydrolases (12). Interestingly, two independent reports have shown that the C-terminal regions of toxins A and B from *Clostridium difficile*, a glucan-binding protein from *Streptococcus mutans*, and several glucosyltransferases from *Streptococcus downei* and *S. mutans* are similar to the C-terminal choline-binding domains of pneumococcal proteins (30, 32). All these proteins appear to be a family of ligand-binding proteins of modular design. In recent years, molecular techniques have identified repetitive amino acid sequences in the superfamily of the wall-associated proteins (8), particularly among gram-positive cocci (32). These repeated units form functional binding domains which appear to be involved in recognition processes.

The detection of CspA in the cell envelope as well as the existence of a signal peptide and a choline-binding domain suggests that CspA might be classified as a putative wall-associated protein. The presence of CspA in the culture medium

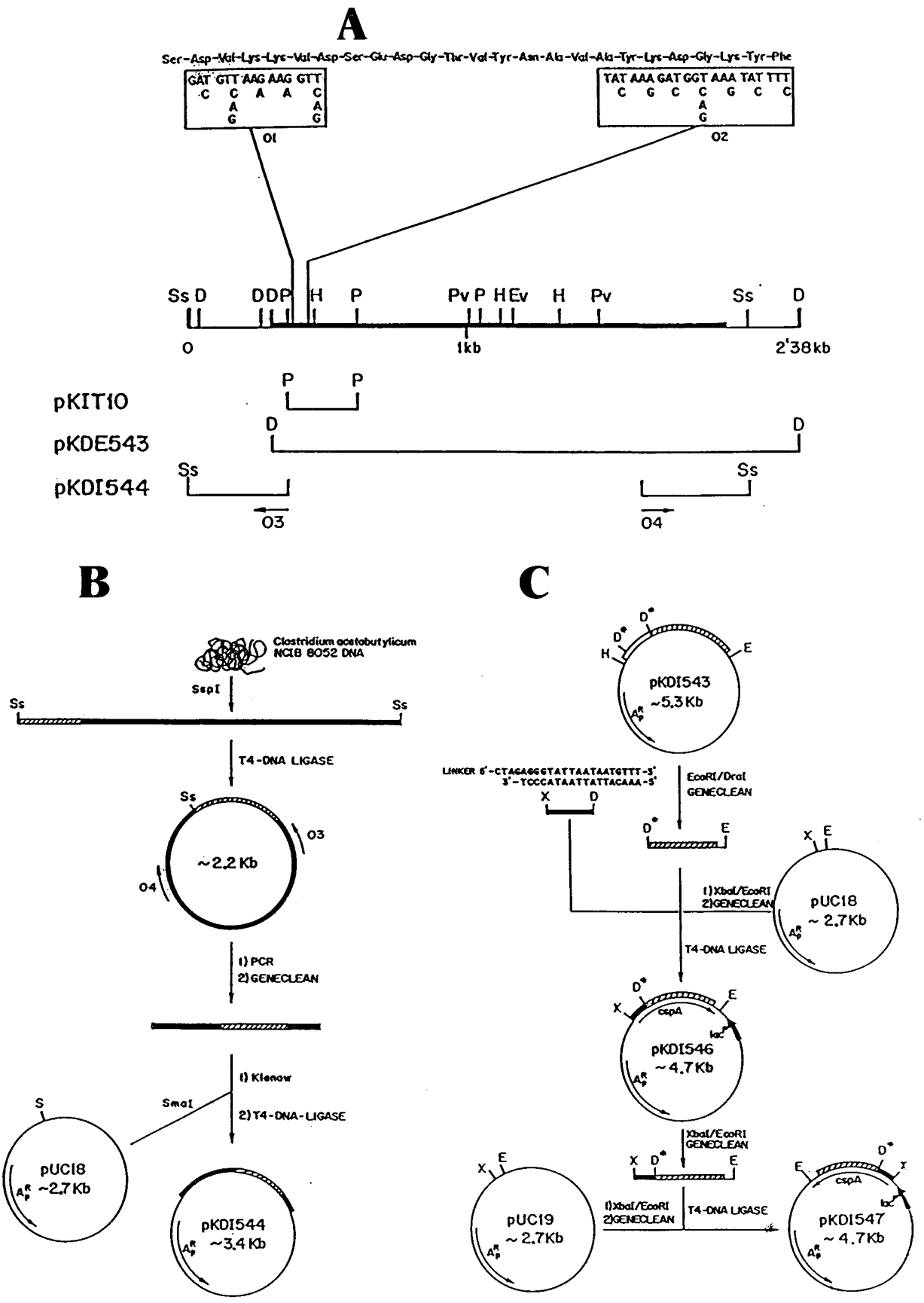
A
Ser-Asp-Val-Lys-Lys-Val-Asp-Ser-Glu-Asp-Gly-Thr-Val-Tyr-Asn-Ala-Val-Ala-Tyr-Lys-Asp-Gly-Lys-Tyr-Phe
GAT GTT AAG AAG GTT
C C A A C
A A
G G
TAT AAA GAT GGT AAA TAT TTT
C G C C G C C
A
G
O1
O2
Ss D DDP H P Pv P HEv H Pv Ss D
0 1kb 2'38kb
pKIT10
pKDE543
pKDI544
O3
O4
B
Clostridium acetobutylicum NCIB 8052 DNA
SspI
Ss
Ss
T4-DNA LIGASE
~2.2 Kb
1) PCR
2) GENECLEAN
1) Klenow
2) T4-DNA-LIGASE
SmaI
pUC18 ~2.7 Kb
pKDI544 ~3.4 Kb
C
pKDI543 ~5.3 Kb
LINKER 5'-CTAGAGGGTATTAATAATGTTT-3'
3'-TCCCATAATTATTACAAA-5'
EcoRI/DraI GENECLEAN
1) XbaI/EcoRI
2) GENECLEAN
pUC18 ~2.7Kb
T4-DNA LIGASE
cspA
pKDI546 ~4.7Kb
XbaI/EcoRI GENECLEAN
pUC19 ~2.7Kb
pKDI547 ~4.7Kb

FIG. 2. Cloning and reconstruction of *cspA*. (A) Restriction map of *cspA*. The N-terminal amino acid sequence of the protein is indicated in the three-letter code. The solid line denotes the structural sequence. The length of the fragment is indicated in kilobases. The inserts of pUC18-derivative plasmids pKIT10, pKDE543, and pKDI544 are indicated. (B) Illustration of the procedure used to clone *cspA* upstream and downstream sequences. (C) Reconstruction of *cspA*. The solid boxes represent the synthetic linker used to reconstruct the ribosome-binding site of the *E. coli lpp* gene. The hatched boxes represent the *cspA* structural gene. The open box represents a clostridial sequence. The heavy arrow indicates the *lac* promoter. Abbreviations (for all panels): Ap^R, ampicillin resistance; D, *Dra*I; E, *Eco*RI; Ev, *Eco*RV; H, *Hin*dIII; P, *Pst*I; Pv, *Pvu*II; S, *Sma*I; Ss, *Ssp*I; X, *Xba*I. Asterisks mean that plasmids contain additional *Dra*I sites. O3 (5′-ATAAGCAGGCACTAAAGC CATAA-3′) and O4 (5′-CTCAAACTTCAGCAGGATGGACTT-3′) indicate the oligonucleotides used for PCR cloning. Klenow represents the Klenow fragment of *E. coli* DNA polymerase I.

(Fig. 1) does not invalidate this hypothesis, since this property is shared by other proteins, such as WapA, the major wall-associated protein of *Bacillus subtilis* (8). This protein of unknown function is actively secreted, becoming wall bound as well as being released into the medium (8). Moreover, it has been recently demonstrated that the choline-binding domain of the pneumococcal surface protein PspA allows this protein to remain anchored to the bacterial cell surface, but some alterations in the composition of the culture medium can induce the release of PspA (33). We have also demonstrated that a serial deletion of the six motifs contained in the choline-binding domain of the pneumococcal LytA amidase causes a

```
1                                                AATATTAACTAAAAAAATATATTTATATATTTTTTAAAAAAAGTACCCAAAGTGAATTTTGACATTGTATTATTTAGTGATA

83   TTACAGTAAATACTAGGAAGAACAGTAAATTATGAAAGTAATAGCATTTTATGTACAAGCATTATGTAATGTGGTATCATGTTATAGTTAATGTAAATGTGTAATAAATTACATGTGATT
                                                          -35                -10                RBS
203  TATTTGATTACTAAGAAATTTATACAAAAATATACATTAGTAGTTATTATATAACTTTACATAGTTGTACAATAATGCTATAATTTATCTGAATTTTAAAATTTAGGAGGAAAAGAAATT

1'     M  F  K  R  A  N  K  I  T  S  L  L  V  A  A  A  S  V  M  A  L  V  P  A  Y  A  A  D  V  K  K  V  D  S  E  D  G  T  V  Y
323  ATGTTTAAAAGAGCAAACAAAATTACATCTTTATTAGTAGCTGCTGCTTCAGTTATGGCTTTAGTGCCTGCTTATGCTGCAGACGTAAAGAAAGTTGATTCAGAAGATGGTACTGTATAC

41'    N  A  V  A  Y  K  D  G  K  Y  F  V  D  G  E  I  N  D  D  E  E  A  Y  Y  V  A  D  G  K  F  N  K  L  E  D  V  D  S  G  D
443  AATGCAGTAGCATACAAAGATGGTAAATACTTTGTTGATGGAGAAATCAACGATGATGAAGAAGCTTACTATGTAGCTGACGGAAAATTCAACAAATTAGAAGATGTTGATTCAGGAGAT

81'    E  A  V  L  F  G  E  K  Y  L  D  V  S  D  G  D  Y  T  V  D  L  D  K  G  S  V  T  D  D  D  V  K  G  D  T  A  D  D  A  A
563  GAAGCAGTTCTATTTGGAGAAAAATACTTAGATGTATCAGACGGAGACTACACTGTTGATTTAGATAAAGGTAGTGTAACTGATGATGACGTTAAAGGTGATACAGCAGACGATGCTGCA

121'   A  A  L  R  K  K  I  K  D  D  T  D  D  R  Y  L  E  S  E  A  E  G  V  K  N  E  D  D  L  A  I  I  G  G  A  K  Y  D  K  P
683  GCAGCTTTAAGAAAGAAAATTAAAGATGATACAGATGATAGATATTTAGAAAGTGAAGCAGAAGGCGTTAAAAATGAAGATGATCTTGCTATAATTGGAGGAGCAAAGTATGACAAGCCT

161'   W  Y  A  T  T  Y  T  A  S  A  K  A  I  G  D  V  N  G  L  T  A  T  N  N  K  F  N  V  Y  T  D  T  N  G  A  Y  I  D  A  D
803  TGGTATGCTACAACATATACTGCTTCTGCAAAAGCTATAGGTGATGTTAATGGATTAACAGCAACAAATAATAAGTTCAATGTTTATACTGATACAAATGGTGCTTACATTGATGCAGAT

201'   Y  N  L  G  K  V  K  V  T  T  T  A  D  G  A  T  K  E  V  T  V  E  N  T  N  D  T  Y  D  A  A  G  S  S  D  S  G  K  D  K
923  TATAACTTAGGAAAAGTAAAAGTTACAACAACTGCTGACGGTGCTACAAAAGAAGTTACTGTAGAAAACACAAACGATACTTATGATGCAGCTGGTTCTAGCGATTCAGGAAAAGATAAA

241'   V  S  A  S  V  K  Q  T  A  V  L  T  Q  D  K  D  N  I  Y  R  L  V  E  V  T  V  K  T  G  H  G  V  I  T  E  I  N  G  V  K
1043 GTAAGTGCAAGTGTTAAACAAACTGCAGTTTTAACACAAGATAAGGATAATATCTACAGACTTGTAGAAGTAACTGTTAAAACAGGTCATGGTGTTATAACAGAAATTAATGGCGTAAAA

281'   I  A  D  I  G  A  N  T  V  F  G  G  T  N  T  E  V  T  F  P  A  I  Q  K  I  S  K  A  Q  A  S  D  D  V  D  G  A  K  Y  A
1163 ATAGCTGACATTGGTGCTAATACAGTATTTGGTGGTACAAACACTGAAGTAACATTCCCAGCTATCCAAAAGATATCAAAAGCACAAGCTTCTGATGATGTTGATGGAGCTAAATATGCT

321'   K  T  V  T  T  Y  A  L  S  D  D  S  G  N  K  L  D  E  Q  S  L  F  V  K  T  D  G  T  A  V  T  T  T  K  Y  T  V  V  N  G
1283 AAAACAGTAACTACTTATGCTTTATCTGATGATTCAGGAAATAAATTAGATGAACAAAGTTTATTTGTAAAAACTGATGGTACTGCTGTTACTACAACTAAGTATACTGTAGTCAATGGA

361'   K  L  I  A  Y  N  T  D  I  N  D  N  K  K  V  T  V  D  A  Y  T  L  K  T  K  G  G  Y  Y  Y  A  D  E  E  D  K  S  E  E  D
1403 AAGCTTATAGCTTACAATACTGATATTAATGACAATAAAAAAGTAACTGTTGATGCATATACATTAAAGACAAAAGGTGGATACTACTACGCTGATGAAGAAGATAAGAGCGAAGAAGAT

401'   C  E  V  S  A  Q  D  K  T  T  A  A  V  Q  T  D  V  D  G  N  L  W  R  L  D  G  G  Y  I  Y  K  F  D  N  T  D  D  W  D  K
1523 TGCGAAGTAAGTGCTCAAGATAAAACAACAGCAGCTGTTCAAACAGACGTTGATGGAAATCTTTGGAGATTAGATGGTGGATATATCTACAAATTCGATAACACTGATGATTGGGATAAA

441'   V  Y  K  V  D  G  S  F  D  E  L  S  V  Y  D  K  D  N  M  V  A  W  S  E  D  D  D  V  Y  S  L  I  G  G  K  S  D  D  N  K
1643 GTTTACAAAGTAGACGGATCTTTCGATGAATTATCAGTTTACGATAAAGACAACATGGTTGCTTGGAGCGAAGATGATGATGTATATTCATTAATCGGAGGAAAGAGCGATGATAATAAG
                                    |--> P1                                                    |--> P2
481'   G  D  D  Q  G  T  T  P  V  V  K  A  G  W  A  Q  T  S  A  G  W  T  Y  V  K  A  D  G  T  K  A  T  G  W  L  Q  D  G  G  A
1763 GGTGATGACCAAGGAACAACTCCTGTAGTTAAAGCTGGTTGGGCTCAAACTTCAGCAGGATGGACTTATGTTAAAGCTGATGGAACTAAAGCTACTGGATGGTTACAAGACGGTGGTGCT
                                    |--> P3                                                    |--> P4
521'   W  Y  Y  L  K  A  D  G  T  M  A  T  G  W  I  Q  D  G  A  T  W  Y  Y  L  N  G  S  G  A  M  Q  T  G  W  L  N  D  N  G  T
1883 TGGTACTACTTAAAAGCTGATGGTACAATGGCTACAGGTTGGATTCAAGATGGAGCAACTTGGTACTACTTAAACGGATCAGGTGCTATGCAAACTGGTTGGTTAAATGATAACGGAACT

561'   F  Y  Y  L  N  G  S  G  A  M  L  S  N  T  T  T  P  D  G  Y  Y  V  G  A  N  G  A  W  V  K  *
2003 TTCTACTACTTAAATGGATCAGGTGCTATGTTATCTAACACAACAACTCCTGATGGATACTATGTAGGAGCTAATGGAGCTTGGGTTAAATAGTTTTAATCTGAACTTATTAGATAAATA

2123 AAAAAGGACAAGGCTTAGCCTTGTTCTTTTTTTATCAATAATATTATTATATAATGAATTTACATTTGAAACTCTAATGGGCTAAGTGCTCAATTAAGATTAGAAAGTAGAAAATCAATT

2243 GTTATATATTAGATAGGAAAAGATTTTATGTATATGAATGCTTTGAAGAATATAAGAAAAAATAGAGTTGATCCTCCTATGTAGAGGTTTGATTATTTAATTTACTTATATTTAGTATAT

2363 AATAGTCTATATGAGTTTAAA
```

FIG. 3. Nucleotide sequence of *cspA*. Nucleotide sequencing was performed on both strands with the T7 DNA polymerase kit from Pharmacia, with universal and synthetic oligonucleotides (Isogen) being used as primers. The putative −35 and −10 promoter sequences, the ribosome-binding site, and the hairpin loop are underlined. The deduced amino acid sequence is shown in one-letter code above the nucleotide sequence. The numbering of nucleotides and amino acids is shown on the left margin. P1, P2, P3, and P4 indicate the sequence repeats.

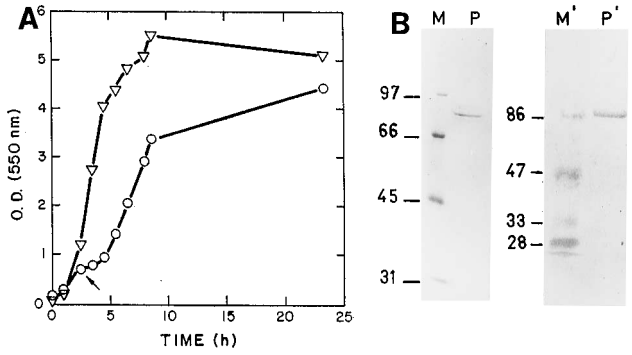

FIG. 4. Expression of *cspA* in *E. coli*. (A) The growth curves of *E. coli* TG1 (pUC19) (▽) and *E. coli* TG1(pKDI547) (○) are shown. The arrow indicates the time at which IPTG (1 mM) was added to the medium. (B) Coomassie blue-stained profile (left panel) and Western blot analysis (right panel) of CspA protein produced in *E. coli* TG1(pKDI547). CspA was purified on DEAE-cellulose (lanes P and P′). The molecular masses of the standard protein markers (lane M) and prestained protein markers (lane M′) are expressed in kilodaltons. O.D., optical density.

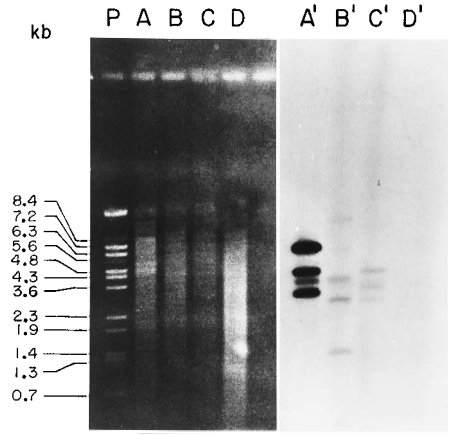

FIG. 5. Southern blot analysis of other clostridial strains. Autoradiographs of the *Hin*cII digestion of whole-cell DNA from different clostridial strains probed with the labeled *Pvu*II-*Ssp*I fragment of pKDE543 encoding the choline-binding domain of CspA are shown. Digestions stained with ethidium bromide are shown in the left panel of the figure, whereas the Southern blot analysis is shown in the right panel. The clostridial strains analyzed were NCIB 8052 (lanes A and A′), NI-504 (lanes B and B′), NI-4 (provided by H.-W. Ackermann, Université Laval, Quebec, Quebec, Canada) (lanes C and C′), and ATCC 824 (lanes D and D′). The molecular sizes of the markers (lane P) are indicated at the left in kilobases.

sequential reduction of its hydrolytic activity and its binding capacity (10). Whether the presence of only four motifs in the choline-binding domain could not be enough to keep CspA tightly bound to the cell envelope or whether this behavior forms part of a regulatory mechanism that responds to intracellular or extracellular physiological conditions is something that remains to be elucidated.

The glycosylated nature of CspA, the high content of acidic residues (18%) by comparison with the content of basic residues (10%) (this ratio appears to be responsible for the anomalous mobility observed by SDS-PAGE), the presence of an N-terminal signal sequence, the high molecular mass, and the very low content of cysteines (only one residue per molecule) are some of the characteristics that CspA shares with PS2, an S-layer protein that is one of the two major secreted proteins of *Corynebacterium glutamicum* (15, 22). The absence of similarity between the sequences of CspA and S-layer proteins is not unexpected since the S-layer proteins show little similarity among themselves (22). On the other hand, the glycosidic nature of CspA and the preponderance of rhamnose might also suggest a function in polysaccharide transport, as has been postulated for the platelet aggregation-associated glycoprotein from *Streptococcus sanguis* (6).

Expression of *cspA* in *E. coli*. The structural region of *cspA* was reconstructed according to the scheme shown in Fig. 2C. A synthetic linker was used to provide the ribosome-binding site of the *E. coli lpp* gene (14), the start codon, and the codon corresponding to Phe-2. Initial efforts to clone the complete *Ssp*I chromosomal fragment containing *cspA* or to reconstruct directly the gene into plasmid pUC19 or other expression vectors containing strong promoters, such as the *lpp* promoter of plasmid pINIIIA3 (14), were unsuccessful since all the plasmids were deleted (data not shown). Only when the *cspA* gene had previously been cloned in pUC18 in the opposite orientation to that of the *lac* promoter were we able to express the gene under the control of this promoter in pUC19. Nevertheless, Fig. 4A shows that the growth of *E. coli*(pKDI547) enters a lag phase after IPTG (isopropyl-β-D-thiogalactopyranoside) induction, suggesting that the expression of the gene strongly influences the metabolism of the host.

The production of CspA in *E. coli*(pKDI547) was detected by Western blot, and the protein was purified with DEAE-cellulose (Fig. 4B). Interestingly, CspA produced in *E. coli* had an M_r of 80,000, which is higher than the M_r deduced from the nucleotide sequence but lower than that observed for the protein produced by *C. acetobutylicum* (Fig. 1).

Analysis of other clostridial strains. To test for the presence of genes homologous to *cspA* in other clostridial strains, chromosomal DNA from different sources was extracted and subjected to Southern blotting and hybridization at low stringency, with the C-terminal encoding region of *cspA* being used as a probe. Figure 5 shows that along with strain NCIB 8052, the other choline-containing strains, NI-4 and NI-504, also contain DNA sequences that hybridize with *cspA*. However, no band was detected with strain ATCC 824, which does not contain choline in the cell wall. These results also suggest that the chromosome of strain NCIB 8052 contains other sequences encoding choline-binding proteins. Moreover, the choline-containing strains NI-504 and NI-4 showed a different pattern of hybridization bands. Although it has been proposed that strains NCIB 8052 and NI-4 are identical (31), our results clearly indicate that they are different. This discrepancy can probably be ascribed to the use of strains from different sources.

In summary, the high similarity found between the choline-binding domains of several pneumococcal proteins and the C-terminal region of the major secreted protein of *C. acetobutylicum* NCIB 8052 strongly supports the hypothesis that these domains have evolved from a common ancestral protein and reinforces the postulates of the modular theory of protein evolution. How these domains have been exchanged among these bacteria, which live in different ecological habitats, still remains an open question, but our findings strongly suggest that the trail of the choline-binding domain can be followed in other bacteria or phages. Our results reveal a new aspect of the role of choline in prokaryotic cells.

Nucleotide sequence accession number. The nucleotide sequence has been submitted to the GenBank and the EMBL Data Bank with the accession number Z37723.

We thank R. López, E. García, and P. García for helpful comments. We are grateful to B. Murdock for correcting the English version of

the manuscript. We also thank E. Cano and M. Carrasco for excellent technical assistance, E. Méndez for the N-terminal analyses, A. Prieto for the gas-liquid chromatography analyses, and A. Hurtado for the artwork.

This work was supported by a grant from the Programa Sectorial del Conocimiento (PB90-0069). A. R. Sánchez-Beato was supported by a fellowship from the Comunidad Autónoma de Madrid.

REFERENCES

1. **Arondel, V., C. Benning, and C. R. Somerville.** 1993. Isolation and functional expression in *Escherichia coli* of a gene encoding phosphatidylethanolamine methyltransferase (EC 2.1.1.17) from *Rhodobacter sphaeroides*. J. Biol. Chem. **268:**16002–16008.
2. **Croux, C., C. Ronda, R. López, and J. L. García.** 1993. Interchange of functional domains switches enzyme specificity: construction of a chimeric pneumococcal-clostridial cell wall lytic enzyme. Mol. Microbiol. **9:**1019–1025.
3. **Cummins, C. S., and J. L. Johnson.** 1971. Taxonomy of the clostridia: wall composition and DNA homologies in *Clostridium butyricum* and other acid-producing clostridia. J. Gen. Microbiol. **67:**33–46.
4. **Díaz, E., R. López, and J. L. García.** 1990. Chimeric phage-bacterial enzymes: a clue to the modular evolution of genes. Proc. Natl. Acad. Sci. USA **87:**8125–8129.
5. **Díaz, E., R. López, and J. L. García.** 1991. Chimeric pneumococcal cell wall lytic enzymes reveal important physiological and evolutionary traits. J. Biol. Chem. **266:**5464–5471.
6. **Erickson, P. R., and M. C. Herzberg.** 1993. Evidence for the covalent linkage of carbohydrate polymers to a glycoprotein from *Streptococcus sanguis*. J. Biol. Chem. **268:**23780–23783.
7. **Fernández-Patrón, C., L. Castellanos-Serra, and P. Rodríguez.** 1992. Reverse staining of sodium dodecyl sulfate polyacrylamide gels by imidazole-zinc salts: sensitive detection of unmodified proteins. BioTechniques **12:**564–573.
8. **Foster, S. J.** 1993. Molecular analysis of the three major wall-associated proteins of *Bacillus subtilis* 168: evidence for processing of the product of a gene encoding a 285 kDa precursor two-domain ligand-binding protein. Mol. Microbiol. **8:**299–310.
9. **García, E., J. M. Rojo, P. García, C. Ronda, R. López, and A. Tomasz.** 1982. Preparation of antiserum against pneumococcal autolysin. Inhibition of autolysin activity and some autolytic processes by the antibody. FEMS Microbiol. Lett. **14:**133–136.
10. **García, J. L., E. Díaz, A. Romero, and P. García.** 1994. Carboxy-terminal deletion analysis of the major pneumococcal autolysin. J. Bacteriol. **176:**4066–4072.
11. **García, J. L., E. García, J. M. Sánchez-Puelles, and R. López.** 1988. Identification of a lytic enzyme of *Clostridium acetobutylicum* that degrades choline-containing pneumococcal cell walls. FEMS Microbiol. Lett. **52:**133–138.
12. **García, P., J. L. García, E. García, J. M. Sánchez-Puelles, and R. López.** 1990. Modular organization of the lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. Gene **86:**81–88.
13. **Guarro, J., J. Cano, J. A. Leal, B. Gómez-Miranda, and M. Bernabé.** 1993. Composition of the cell wall polysaccharides in some geophilic dermatophytes. Mycopathologia **122:**69–77.
14. **Inouye, S., and M. Inouye.** 1985. Up-promoter mutations in the *lpp* gene of *Escherichia coli*. Nucleic Acids Res. **13:**3101–3110.
15. **Joliff, G., L. Mathieu, V. Hahn, G. Bayan, F. Duchiron, M. Renaud, E. Shechter, and G. Leblon.** 1992. Cloning and nucleotide sequence of the *csp1* gene encoding PS1, one of the two major secreted proteins of *Corynebacterium glutamicum*: the deduced N-terminal region of PS1 is similar to the *Mycobacterium* antigen 85 complex. Mol. Microbiol. **6:**2349–2362.
16. **Kilpper-Bälz, R., P. Wenzig, and K. H. Schleifer.** 1985. Molecular relationships and classification of some viridans streptococci as *Streptococcus oralis* and emended description of *Streptococcus oralis* (Bridge and Sneath 1982). Int. J. Syst. Bacteriol. **35:**482–488.
17. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157:**105–132.
18. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
19. **López, R., J. L. García, E. García, C. Ronda, and P. García.** 1992. Structural analysis and biological significance of the cell wall lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages. FEMS Microbiol. Lett. **100:**439–448.
20. **Messner, P., R. Christian, J. Kolbe, G. Schulz, and U. B. Sleytr.** 1992. Analysis of a novel linkage unit of O-linked carbohydrates from the crystalline surface layer glycoprotein of *Clostridium thermohydrosulfuricum* S102-70. J. Bacteriol. **174:**2236–2240.
21. **Paul, G., F. Loottspeich, and F. Wieland.** 1986. Asparaginyl *N*-acetylgalactosidase. J. Biol. Chem. **261:**1020–1024.
22. **Peyret, J. L., N. Bayan, G. Joliff, T. Gulik-Krzywicki, L. Mathieu, E. Shechter, and G. Leblon.** 1993. Characterization of the *cspB* gene encoding PS2, an ordered surface-layer protein in *Corynebacterium glutamicum*. Mol. Microbiol. **9:**97–109.
23. **Podvin, L., G. Reysset, J. Hubert, and M. Sebald.** 1988. Presence of choline in teichoic acid of *Clostridium acetobutylicum* NI-4 and choline inhibition of autolytic function. J. Gen. Microbiol. **134:**1603–1609.
24. **Romero, A., R. López, and P. García.** 1990. Characterization of the pneumococcal bacteriophage HB-3 amidase: cloning and expression in *Escherichia coli*. J. Virol. **64:**137–142.
25. **Ronda, C., J. L. García, E. García, J. M. Sánchez-Puelles, and R. López.** 1987. Biological role of the pneumococcal amidase. Cloning of the *lytA* gene in *Streptococcus pneumoniae*. Eur. J. Biochem. **164:**621–624.
26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Sánchez-Puelles, J. M., J. M. Sanz, J. L. García, and E. García.** 1990. Cloning and expression of gene fragments encoding the choline-binding domain of pneumococcal murein hydrolases. Gene **89:**69–75.
28. **Sanz, J. M., E. Díaz, and J. L. García.** 1992. Studies on the structure and function of the N-terminal domain of the pneumococcal murein hydrolase. Mol. Microbiol. **6:**921–931.
29. **Simonen, M., and I. Palva.** 1993. Protein secretion in *Bacillus* species. Microbiol. Rev. **57:**109–137.
30. **von Eichel-Streiber, C., M. Sauerborn, and H. K. Kuramitsu.** 1992. Evidence for a modular structure of the homologous repetitive C-terminal carbohydrate-binding sites of *Clostridium difficile* toxins and *Streptococcus mutans* glucosyltransferases. J. Bacteriol. **174:**6707–6710.
31. **Wilkinson, S. R., and M. Young.** 1993. Wide diversity of genome size among different strains of *Clostridium acetobutylicum*. J. Gen. Microbiol. **139:**1069–1076.
32. **Wren, B. W.** 1991. A family of clostridial and streptococcal ligand-binding proteins with conserved C-terminal repeat sequences. Mol. Microbiol. **5:**797–803.
33. **Yother, J., and J. M. White.** 1994. Novel surface attachment mechanism of the *Streptococcus pneumoniae* protein PspA. J. Bacteriol. **176:**2976–2985.
34. **Young, M., N. P. Minton, and W. L. Staudenbauer.** 1989. Recent advances in the genetic of clostridia. FEMS Microbiol. Rev. **63:**301–326.