Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 914–918, February 1988
Microbiology

Molecular evolution of lytic enzymes of *Streptococcus pneumoniae* and its bacteriophages

(autolytic enzymes/lysins/DNA sequencing/functional domains/host–parasite relationships)

ERNESTO GARCÍA, JOSÉ L. GARCÍA, PEDRO GARCÍA, ANA ARRARÁS, JOSÉ M. SÁNCHEZ-PUELLES, AND RUBENS LÓPEZ*

Centro de Investigaciones Biológicas, Consejo Superior de Investigaciones Cientfficas, Velázquez, 144, 28006 Madrid, Spain

Communicated by Lynn Margulis, September 21, 1987

ABSTRACT **A 2.9-kilobase *Acc* I fragment of the DNA of the pneumococcal bacteriophage Cp-1, containing the *cpl* gene, hybridizes with the *lytA* gene encoding the pneumococcal amidase. The nucleotide sequence of the *cpl* gene of Cp-1, encoding a muramidase (CPL), has been determined. The 3′ regions of the *cpl* and *lytA* coding sequences show considerable nucleotide sequence homology and the carboxyl-terminal domains of the deduced amino acid sequences of these lysins are quite similar: 73 of the carboxyl-terminal 142 amino acid residues are identical, and of the 69 substitutions, 55 are conservative. Comparisons between CPL, the pneumococcal amidase, and the muramidase of the fungus *Chalaropsis* sp. (an enzyme that also degrades the pneumococcal cell wall) strongly suggest that the carboxyl-terminal domains of CPL and of the amidase might be responsible for the specific recognition of choline-containing cell walls, as well as for the noncompetitive inhibition of the catalytic activity of these enzymes by the pneumococcal lipoteichoic acid or by high concentrations of choline. In addition, the active center of these enzymes should be located in their amino-terminal domains. Our results suggest an evolutionary relationship between phage and host lysins.**

Autolysins are enzymes that hydrolyze covalent bonds of the cell walls. Most bacterial species contain one or more autolysins of different specificity, suggesting that these enzymes fulfill essential physiological functions. Mutations in the structural genes of bacterial murein hydrolases have not been characterized except in the case of *Streptococcus pneumoniae* (1–3), making it difficult to draw general conclusions about the physiological role(s) of these ubiquitous enzymes. *S. pneumoniae* contains two autolysins, a powerful *N*-acetylmuramoyl-L-alanine amidase (mucopeptide amidohydrolase, EC 3.5.1.28) (4) and a glycosidase (5). The pneumococcal amidase is encoded by the *lytA* gene (6) and its nucleotide and deduced amino acid sequences have been determined (7). The amidase is the agent responsible for the separation of daughter cells at the end of cell division (3, 8). In addition, this enzyme participates in the liberation of progeny bacteriophage into the medium (9), although other enzymes encoded by the infecting bacteriophage are also involved in this process (10, 11). Two phage-coded lytic enzymes have been purified and biochemically characterized: (*i*) an amidase similar but not identical to the host one, from cultures infected with bacteriophage Dp-1 (10, 12), and (*ii*) a muramidase (CPL) encoded by the *cpl* gene of bacteriophage Cp-1 (11).

There is a characteristic of the pneumococcal cell wall that makes its peptidoglycan peculiar and unique. Pneumococcal teichoic acids contain choline phosphate as a structural component (13), the only case in nature in which choline residues are a component of a polysaccharide. Surprisingly, all the lytic enzymes found in *S. pneumoniae* and its bacteriophages share some important properties, especially an absolute requirement for choline-containing teichoic acid for activity. It has been pointed out that choline-containing teichoic acids act as adsorption ligands for the amidase; i.e., the enzyme needs to be attached to its insoluble substrate before it can degrade it (14). Furthermore, the catalytic activity of these enzymes is inhibited by high concentrations of choline (14, 15) or by pneumococcal lipoteichoic acid (16) (a lipopolysaccharide that also contains choline). On the other hand, some lytic enzymes isolated from other sources, the muramidases from *Chalaropsis* or from *Streptomyces globisporus*, can degrade pneumococcal cell walls even when choline has been replaced by the structural analogue ethanolamine at the position normally occupied by choline (17).

Knowledge of lysin functions is generally considered to be important for the evolutionary analysis of bacteriophages (18); the peculiar choline dependence of the pneumococcal lysins suggests that the unique presence of choline in the cell wall of *S. pneumoniae* serves as a strong selective pressure. It is conceivable that under these conditions, the divergence between bacterial and phage lysin genes would be considerably less than in other species. If this were the case, a comparative analysis of those genes would be the key for studying the evolutionary relationships between host and parasite genes. For investigating these possibilities, our experimental approach was to determine the nucleotide sequence of the muramidase encoded by Cp-1 DNA† and to compare it to the sequence of the host amidase already determined (7).

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Phages. The *Escherichia coli* strains used were HB101 (r^- m^- *recA13*) (19) and JM103 [*hsdR4* (F′ *traD36 proAB lacI*^{q} *lacZ*Δ*M15*)]. The latter was used as a host for phages M13mp10 and M13mp11 (20). The pneumococcal bacteriophages used, Cp-1, Cp-5, Cp-7, and Cp-9, have been described (21). Plasmid pGL80 (Ap^R *lytA*), encoding the pneumococcal amidase, has been described (7). Plasmid pCIP50 (Ap^R Tc^R *cpl*, where *cpl* is the structural gene of the muramidase encoded by bacteriophage Cp-1) was constructed by cloning the 2.9-kilobase (kb) *Acc* I fragment of Cp-1 DNA first into the replicative form of the M13mp11 DNA and then into the *Eco*RI–*Hin*dIII site of plasmid pBR325 (11).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

*To whom reprint requests should be addressed.

†This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03586).

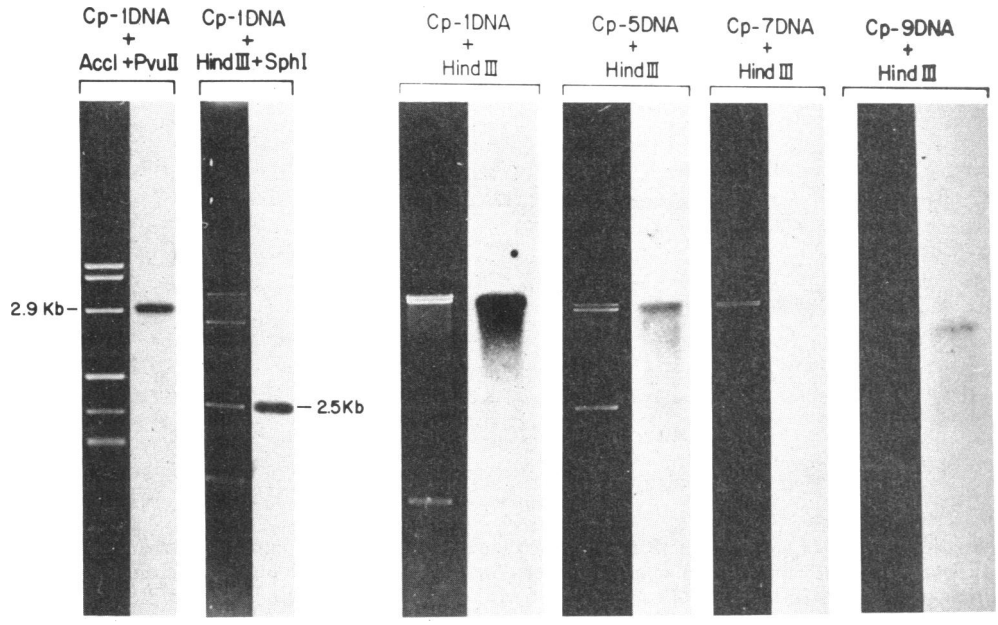

FIG. 1. Ethidium bromide-stained agarose gels and Southern blot analyses of restriction fragments obtained from Cp-1, Cp-5, Cp-7, and Cp-9 DNAs. Ethidium bromide-stained gels show the restriction patterns obtained from the digestion of the different DNAs used. The autoradiograph to the right of each gel shows the fragments blotted to Hybond membranes and hybridized at 50°C for 5 hr with ^{32}P-labeled pGL80. The sizes of two restriction fragments (kb) of Cp-1 DNA that hybridize with the probe are indicated.

Materials. Restriction endonucleases, phage T4 DNA ligase, DNA polymerase (Klenow fragment), BAL-31 nuclease, and proteinase K were purchased from commercial suppliers (New England BioLabs, Amersham International, and Merck). Radionuclides, Hybond membranes, and kits for DNA sequencing and for nick-translation were obtained from Amersham.

Preparation of DNA. Proteinase K-treated phage DNA was prepared as described (21). Plasmid DNA was prepared by the rapid alkaline method described by Birnboim and Doly (22).

DNA Sequence Analysis. DNA sequencing was performed by the dideoxy chain-termination method (23). Data analysis was simplified by computer programs run on a Zenith Z-100 personal computer. Protein similarity searches were conducted by using the Protein Sequence Database of the Protein Identification Resource‡ with the computer programs of Lipman and Pearson (24).

Southern Blot Hybridization. Hybridization (25) was performed according to Maniatis *et al.* (26). Radioactive probes were synthesized by nick-translation (26) in the presence of [α-^{32}P]dCTP. Radioactive bands were detected with Kodak X-Omat films and Dupont Cronex Lightning Plus intensifying screens at −70°C.

RESULTS

Sequence Homology Between Lysin-Coding Phage DNA and Amidase-Coding Host DNA. The biochemical similarities found between the host amidase and the CPL muramidase encoded by Cp-1 DNA prompted us to test whether the *lytA* and *cpl* genes also showed evidence of homology at the nucleotide level. Fig. 1 shows that pGL80 (*lytA*) hybridized intensely to the 2.9-kb *Acc* I fragment of Cp-1 DNA. This fragment has previously been cloned in *E. coli*, using pBR325 as a vector, and the recombinant plasmid (pCIP50) allowed the expression of the phage-coded muramidase (CPL) in the heterologous system (11). These results suggested that the *cpl* gene was responsible for the strong hybridization signal observed with the *lytA* gene. Additionally, Fig. 1 shows that the region of homology is located between the *Sph* I site and the right end of the 2.9-kb *Acc* I fragment (see also map in Fig. 2). Attempts to express the CPL by cloning either the left or the right *Acc* I–*Sph* I fragment in pBR325 were unsuccessful, suggesting that the *Sph* I site is located in an important region of the *cpl* gene (i.e., in the promoter or within the structural gene itself). Other phage genomes related to Cp-1 DNA also hybridized with pGL80. The intensity of the hybridization bands varied with the different phage DNAs, being least intense for Cp-7 DNA (Fig. 1). This result is in agreement with the average nucleotide homology between the four genomes calculated from restriction enzyme analyses, where Cp-7 DNA showed the highest divergence with respect to Cp-1, Cp-5, and Cp-9 DNAs (21). Restriction enzyme analysis of these phage DNAs (data not shown) indicated that the regions of homology are located, as in the case of Cp-1 DNA, near the right end of the genomes.

Nucleotide Sequence Analysis of the *cpl* Gene. The strategy for determining the sequence of the *cpl* gene is summarized in Fig. 2. The DNA sequence and the amino acid translation appear in Fig. 3. Only one open reading frame of significant length (1017 nucleotides) was found. This indicates that the phage-coded muramidase CPL would have a M_r of 39,146 (339 codons), which agrees with the value of 39,000 previously found (11). Ten bases upstream from the initiator ATG codon, a ribosome-binding sequence (AAGGAG) is found. However, as previously observed for the *lytA* gene (7), no canonical promoter-like sequences are found in the A+T-rich region preceding the *cpl* gene. A putative ρ-independent terminator (27) with a ΔG of −20.6 kcal (28) appears 11 bases downstream from the C-terminal codon of CPL. One of the most interesting features of CPL (and of the host amidase as discussed below) is the presence, in the carboxyl-terminal half of CPL (and also in the same region of the host amidase), of a set of six repeated sequences (P1–P6), each about 20 amino acid residues long (Fig. 4).

Comparison Between the *cpl* and *lytA* Genes. The nucleotide and amino acid sequences of the *cpl* gene were com-

‡Protein Identification Resource (1987) Protein Sequence Database (Natl. Biomed. Res. Found., Washington, DC), Release 13.

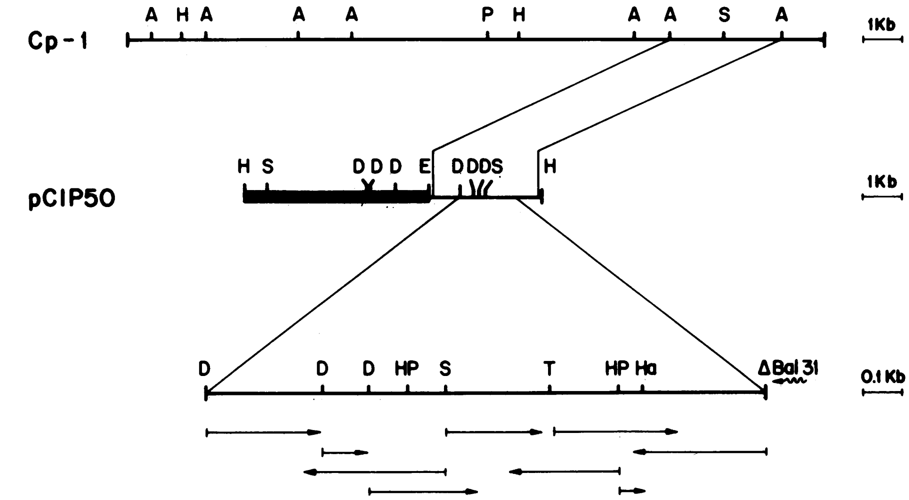

FIG. 2. Partial restriction maps of Cp-1 and pCIP50 DNAs and the strategy for determining the nucleotide sequence of the *cpl* gene. Locations of the restriction sites used for sequencing are indicated. Arrows below the map indicate the direction and extent of sequence determination. Restriction sites: A, *Acc* I; H, *Hind*III; P, *Pvu* II; S, *Sph* I; E, *Eco*RI; D, *Dra* I; HP, *Hin*PI; Ha, *Hae* II; T, *Taq* I. ΔBal31 indicates the direction of the deletion generated in pCIP50 by BAL-31 nuclease for purposes of sequencing.

pared to those of *lytA*. As shown in Fig. 3, no significant similarity can be found in the 5′ regions of the two genes (i.e., from positions 1–548 of *cpl* and from positions 1–468 of *lytA*). However, 43% of the 123 subsequent nucleotides of the two genes are identical, and this proportion increases to 54% identity from nucleotides 664 and 592 to the ends of the *cpl* and *lytA* genes, respectively. These results are in agreement with the hybridization analyses (Fig. 1), which re-

```
                                                 TTTAAAGGCTTGGAAGTGGAAAGTGACCGATAGCTGGACAGGGTTAAAAGGAGTTATCAAACATACCCTAACATT   -241'
TATTTTTTACTATTTTGTAGCGGTATTCTTAACATATATTCACGCTATGGCAGTCGGTCAGATTTTGCTTGTTATCATTAACTTGTACTATGCACTTTCAATTATGGAAAATCTTGCTGT   -121'
TATGGGTGTCTTTATCCCTAAATTTATGACGGCTAGAGTACAGGAAGAATTGCAGAAATACACCGCACAACTAGACGCAGGAAAAGACCTGCTAGAAGAATTTAAAGGAGAAAAGAAATA   -1'

M  V  K  K  N  D  L  F  V  D  V  S  S  H  N  G  Y  D  I  T  G  I  L  E  Q  M  G  T  T  N  T  I  I  K  I  S  E  S  T  T
ATGGTTAAAAAGAATGATTTATTTGTAGATGTTTCAAGTCACAACGGTTACGATATAACAGGTATCTTGGAGCAAATGGGAACAACTAACACCATCATTAAAATTTCTGAAAGTACGACC   120'
                                                                                    atggaaatt:atgtg:gt:aat:a:ga:c:ga::tgcctcaagtcgg:   48"
                                                                                     m  e  i  N  V  S  K  L  r  T  d  l  p  q  V  G

Y  L  N  P  C  L  S  A  Q  V  E  Q  S  N  P  I  G  F  Y  H  F  A  R  F  G  G  D  V  A  E  A  E  R  E  A  Q  F  F  L  D
TATTTAAACCGTTGCTTGTCTGCTCAAGTGGAGCAGTCAAACCCTATTGGCTTTTATCACTTCGCACGCTTTGGCGGAGACGTAGCAGAAGCCGAAAGAGAAGCGCAGTTTTTCCTTGAC   240'
gtgca:ccata:a:gcaagtacacgc:cactcaactggg::t::gca:tcaaccgta::gaat:a:gcgga:tatcactggcggaa:::cc:a:::tt::gtttttc:cgca:a:::tt   168"
v  q  p  y  r  q  v  h  A  h  S  t  G  :  :  h  S  t  v  q  n  E  a  d  y  h  w  r  k  D  P  :  l  G  f  f  s  h  I  v

N  V  P  M  Q  V  K  Y  L  V  L  D  Y  E  D  D  P  S  G  D  A  Q  A  N  T  N  A  C  L  R  F  M  Q  M  I  A  D  A  G  Y
AACGTGCCTATGCAAGTTAAATACCTTGTATTGGACTACGAGGACGACCCAAGCGGAGACGCACAAGCGAACACTAACGCATGCTTACGCTTTATGCAGATGATTGCTGACGCTGGATAT   360'
gggaacgg:tgcatca:gc:ggtaggacctg:t::ta:t:gt:c:tgggacgtt::g:g::gttggaatgctgag:c:tatgcagcggttgaac::attgaa:gcca:tcaa:caa:g:a   288"
G  n  g  c  i  M  Q  v  g  p  V  :  n  G  A  w  d  v  :  G  G  w  N  A  E  T  y  a  a  v  e  L  i  e  s  h  S  T  k  e

K  P  I  Y  Y  S  Y  K  P  F  T  H  D  N  V  D  Y  Q  Q  I  L  A  Q  F  P  N  S  L  W  I  A  G  Y  G  L  N  D  G  T  A
AAACCTATTTATTATAGTTATAAACCGTTTACACATGATAATGTGGACTATCAGCAAATCCTTGCACAGTTCCCTAATTCTCTATGGATTGCAGGCTATGGCTTAAACGATGGTACAGCT   480'
g:gttc::gacgg:ctaccgcctttata:cga::tcttacgcaatctagcag:tg::gcagg:ttg:c:aaaa:gct:ga:ac:g:::g:tt::ctggaattaa::cgc:c:agtattgc   408"
E  f  M  t  d  y  r  l  y  i  E  l  l  R  n  l  a  D  E  a  g  l  P  k  T  l  D  t  g  s  l  A  g  i  k  T  H  E  y  c

N  F  E  Y  F  P  S  M  D  G  I  R  W  W  Q  Y  S  S  N  P  F  D  K  N  I  V  L  L  D  D  E  E  D  D  K  P  K  T  A  G
AACTTTGAATACTTCCCAAGCATGGACGGGATAAGATGGTGGCAGTATTCTAGTAACCCGTTTGACAAGAATATTGTACTGTTAGACGATGAAGAAGACGACAAGCCAAAGACCGCTGGA   600'
:cgaa:a:cc:accaaacaa:cactca:accacgttgaccctt:tcca:a:ct:gctaaa:gg:g::tt:gccg::ag:a:::ta:gc::::tatt::ga::ggcttg:c::tt:aaac:   528"
T  n  N  q  p  n  N  h  S  D  h  V  d  p  y  p  y  l  A  k  W  G  i  S  r  e  q  F  K  H  D  i  E  N  g  l  T  I  E  T

┌P1                                                       ┐┌P2
T  W  K  Q  D  S  K  G  W  W  F  R  R  N  N  G  S  F  P  Y  N  K  W  E  K  I  G  G  V  W  Y  Y  F  D  S  K  G  Y  C  L
ACGTGGAAACAAGACAGCAAGGGGTGGTGGTTCAGACGAAACAATGGCAGTTTCCCTTATAATAAATGGGAAAAAATCGGTGGTGTGTGGTACTACTTCGATAGTAAAGGATATTGCTTA   720'
ggc:::c:ga:ga:tga::ct::c:ac::::a:gt::attcag:c:::tc::at::aa:ag:c::g:tt::g::::::aa:::cact:::::::::::t::c:::tc:::c:::atgc:t   648"
G  :  Q  K  N  D  T  :  Y  :  Y  v  H  S  D  :  :  Y  :  k  D  :  F  :  :  N  :  T  :  :  :  :  :  :  S  :  :  m  :

  ┐┌P3                                                 ┐┌P4
T  S  E  W  L  K  D  N  E  -  K  W  Y  Y  L  K  D  N  G  A  M  A  T  G  W  V  L  V  G  S  E  W  Y  Y  M  D  D  S  G  A
ACGAGCGAATGGCTCAAAGATAATGAA---AAATGGTACTACCTCAAGGACAACGGCGCAATGGCGACTGGTTGGGTGCTAGTCGGGTCAGAGTGGTATTATATGGACGATTCAGGCGCT   837'
g:aga:cgc:::agg::gc:c:ca::cggc::c:::::::ggt::g:ca::tca::::a::::::t::a::c:::aa:aa:a:::ctgata:::::::c:::t:ca::::aga:::t::c   768"
A  D  r  :  r  :  H  T  D  g  N  :  :  W  F  D  N  S  :  E  :  :  :  :  :  k  k  I  A  D  K  :  :  :  F  N  E  E  :  :

     ┐┌P5                                                    ┐┌P6
M  V  T  G  W  V  K  Y  K  N  N  W  Y  Y  M  T  N  E  R  G  N  M  V  S  N  E  F  I  K  S  -  -  G  K  G  W  Y  F  M  N
ATGGTTACTGGTTGGGTCAAGTATAAGAATAACTGGTACTATATGACAAATGAACGTGGTAACATGGTTTCTAATGAATTTATTAAGTCT------GGAAAAGGTTGGTATTTCATGAAC   951'
:::aag::a::c:::::::::::::c:::g:c:ct::::::::ct:agacgc:a::gaa::cgc::::::a::a::::cc::::::cc::::agcggac::::c:::c::::::c:a:c:c::a   888"
:  k  :  :  :  :  :  :  :  D  T  :  :  :  L  D  A  K  e  :  A  :  :  :  :  A  :  :  Q  :  a  d  :  T  :  :  :  Y  L  K

                   ┐
T  N  G  E  L  A  D  N  P  S  F  T  K  E  P  D  G  L  I  T  V  A end
ACAAACGGAGAGCTTGCAGACAATCCAAGTTTCACGAAAGAACCAGACGGGCTTATAACCGTAGCATAAAAAAGAAAAGCTAGTAGGATTTTCCTACTAGCTGTTTTTATAGTCTGCTAT 1071'
c::g:::::aca::g:::::::gg:::gaa:::::agt:::g::::::t::ct:g::t::a:::aa::::   957"
P  D  :  T  :  :  :  R  :  E  :  :  v  :  :  :  :  :  :  :  :  :  k end

AATTTTATAAGCATCTTCGTCTGGATTATCCAGAGCGATGGAACAGATTGCA 1123'
```

FIG. 3. Comparison of the nucleotide sequences and derived amino acid sequences of the bacteriophage muramidase CPL and the bacterial amidase. Only the DNA strand corresponding to the mRNA sequence is shown. Single primes indicate nucleotide sequence numbers of the *cpl* gene, and double primes indicate sequence numbers of the *lytA* gene (7). Identical nucleotides are indicated by colons in the sequence of the *lytA* gene. The top and bottom lines of the comparison are the deduced amino acid sequences (in one-letter code) for CPL and for pneumococcal amidase (7), respectively. When identical amino acids coincide, those corresponding to the amidase are indicated by colons. Conservative and nonconservative amino acid replacements are represented, in the amidase sequence, by uppercase and lowercase letters, respectively. P1–P6 are sets of repeated sequences. Boxes indicate presumptive −10 and −35 promoter sequences; the heavy underline represents the ribosome binding site; facing arrows show stem–loop structure; the light underline indicates the mRNA transcription stop sequence; dashes are gaps introduced to maximize homology.

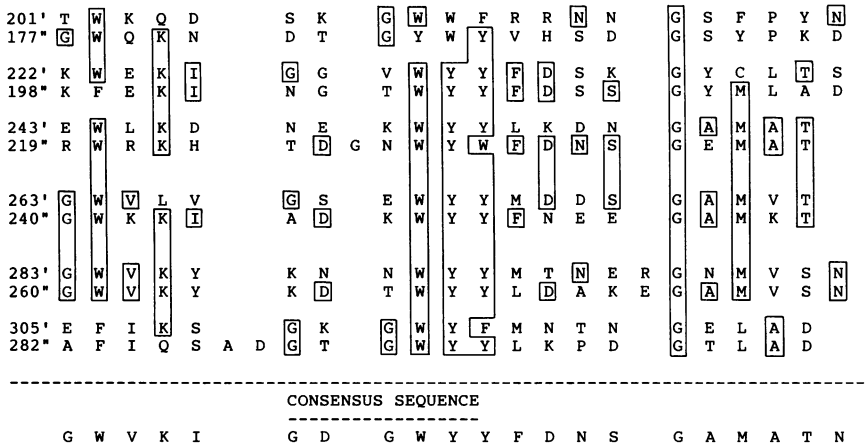

FIG. 4. Comparison of the repetitive modules (P1–P6) of CPL and the pneumococcal amidase. Standard one-letter symbols are used to represent amino acid residues. Numbers indicate the amino acid position in CPL (single prime) or the amidase (double prime). A putative consensus sequence is also shown.

vealed a high degree of DNA/DNA homology between the *lytA* gene and the region of Cp-1 DNA corresponding to the right of the *Sph* I restriction site (located between positions 319 and 324). As expected, similar results were found when the amino acid sequences of the two proteins were compared: 51% of the last 142 residues are identical, whereas in the rest of the molecules only about 3% are identical. When it is taken into account that 55 out of the 69 nonidentical amino acids in the homologous region are conservative substitutions, the carboxyl-terminal regions of CPL and the host amidase can be considered almost identical.

Comparison Between CPL and the Muramidase of *Chalaropsis*. When the amino acid sequence of the muramidase of *Chalaropsis*, an enzyme that degrades choline- or ethanolamine-containing pneumococcal cell walls, was compared to the amino acid sequence of CPL, it was observed that the amino termini of the two enzymes were unexpectedly similar. In fact, 18 of the first 70 residues (26%) are identical (Fig. 5). In addition, many other amino acids were conservative substitutions, which further increased the overall similarity. In order to analyze the significance of this, the amino-terminal ends of the two proteins were subjected to RDF analysis (24); the z value obtained, 4.8, suggests that the similarity may be significant. It has been reported that the catalytic activity of the muramidase of *Chalaropsis* is at least partially due to two amino acids, Asp-6 and Glu-33 (29). These two amino acids are located the same distance apart in the amino-terminal end of the CPL muramidase (Asp-10 and Glu-37). Although the pneumococcal amidase has Asp-11 and Glu-38, separated by 26 residues, no significant similarities were found when this protein was compared with the muramidase of *Chalaropsis*.

DISCUSSION

The idea of a modular organization of phage DNAs has been supported by the partial relationships between *E. coli* phages. Heteroduplexes constructed between the DNAs of phages like λ and φ80 showed that regions of near-perfect homology alternated with regions of total nonhomology. Botstein (30) extended this concept of modular evolution by comparing the genomes of λ and P22, two phages that have different hosts. Attempts to obtain viable hybrid phages showed that large segments of λ DNA could be exchanged for P22 DNA and vice versa (31). The genes specifying the lytic functions of these phages are completely nonhomologous, yet these genes are found at equivalent positions in the genetic maps of the two phages and are interchangeable in interbreeding phage populations, even though P22 degrades the wall of the host bacteria with lysozyme, whereas λ uses an endopeptidase. It has been pointed out that "modules" can be blocks of linked genes or individual genes, or even smaller modules that were shuffled during early phases of phage evolution. It has been suggested that the early modules were of exon size and that the proteins of contemporary phages were built up by rearranging such simple exons into new combinations (18). Our work appears to support this theory. The long regions (nucleotides 601–1332 in the *cpl* coding sequence and nucleotides 529–954 in the *lytA* gene) that are highly similar (Figs. 3 and 4) might represent a repetitive module corresponding to the protein domain that we believe to be involved in the specific recognition of the choline-containing cell wall teichoic acid, which has been suggested to be the allosteric ligand for these enzymes (15). Ethanolamine-containing teichoic acids are completely resistant to degradation by these lysins.

The conclusion that the carboxyl-terminal half of the molecule represents the substrate-recognition domain is also supported by the fact that the phosphatidylcholine-transfer protein from bovine liver, which acts as specific carrier of phosphatidylcholine, shows some amino acid similarity with the carboxyl-terminal region of both CPL and the host amidase, mainly with the group of amino acids that appears as a direct repetition in the lytic enzymes. The phosphatidylcholine-transfer protein has a recognition site for the phosphocholine head group (32), although this site has not been identified. It is interesting that other proteins related in some way to choline-containing compounds [e.g., the phospholipase A2 from *E. coli* (33)] also have some degree of similarity with the carboxyl-terminal domain of the lytic enzymes (data not shown). In the absence of additional data, the significance of these observations is still unclear, but it is tempting

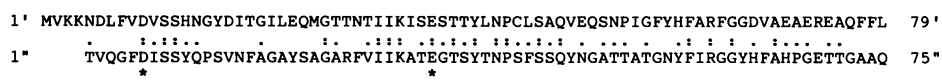

FIG. 5. Comparison of the amino termini of CPL (upper line) and the muramidase of *Chalaropsis* (lower line). Colons and periods indicate identical matches and conservative substitutions, respectively (24). Asterisks indicate the two active-site amino acids of the muramidase of *Chalaropsis* (29).

to assume that the choline-recognition sites of all these enzymes could have diverged from a common precursor.

The repetitive arrangement of the amino acids presumably involved in the recognition of choline suggests an intragenic duplication phenomenon. The functional role of these peptides in the pneumococcal lysins might be the simultaneous recognition of several molecules of choline, which could help to stabilize the enzyme, facilitating the degradation of the substrate by the active center of the lysin. Most probably, the active center would be located in the amino-terminal domain of the molecule. In fact, we have found a significant amino acid similarity between the amino ends of CPL and the muramidase of *Chalaropsis* (Fig. 5), one of the few heterologous enzymes that is capable of degrading pneumococcal cell walls containing either choline or ethanolamine. However, such homology does not appear to exist between this muramidase and the pneumococcal host amidase. Furthermore, the two amino acids of the muramidase of *Chalaropsis* that have been reported to be involved in the active center of the enzyme have also been found in the CPL molecule, where they are located at the same distance from one another. The activity of hen egg-white lysozyme is also due to two acidic residues (Glu-35 and Asp-52) (34), but no other amino acid similarities have been found between either CPL or *Chalaropsis* muramidase and other lysozymes. This lack of amino acid similarity has also been found when comparing T4, hen egg, and goose lysozymes but, strikingly, their three-dimensional structures are rather similar (35). Although no such structural studies have been carried out in the case of the CPL and *Chalaropsis* muramidases, the existence of significant amino acid similarities between their amino-terminal domains suggests that this similarity could be a result of either polyphyletic convergence or sequence divergence, notwithstanding that one is encoded by a phage and the other by a fungus. All these results suggest that the host amidase, as well as the phage muramidase genes, resulted from the fusion of two functionally different modules, one (the amino-terminal domain) conferring specificity of the enzymatic activity and the other being responsible for attachment to the choline residues present in the cell wall of *S. pneumoniae*.

Based on these results, our prediction is that the amino terminus of the amidase from another pneumococcal phage, Dp-1 (10, 12), should contain regions of homology with the amino-terminal domain of the host amidase. In fact, Southern blot hybridization experiments (data not shown) have indicated that the *lytA* gene is also homologous to a particular restriction fragment of Dp-1 DNA. In addition, the *cpl* gene hybridizes with the DNA of the pneumococcal strain M31 (Δ*lytA*). It is conceivable that the gene encoding the lytic glycosidase recently detected in this strain of *S. pneumoniae* (5) could be responsible for this hybridization. Whether or not this were the case, the system of lytic enzymes encoded by *S. pneumoniae* and its bacteriophages would provide an attractive model for the study of interesting and peculiar mechanisms of recognition and degradation of substrate by enzyme, as well as of the evolutionary relationships between host and parasite lysins.

We thank B. Dorritie for correcting the English version of the manuscript and M. Carrasco for his excellent technical assistance. The art work of A. Hurtado is gratefully acknowledged. This work was supported by grants from Comisión Asesora de Investigación Científica y Tecnológica (no. 144) and from the Spain–USA Joint Committee for Scientific and Technological Cooperation (CDB 840 20, 35). Fellowship support to J.M.S.-P. was provided by the Consejo Superior de Investigaciones Científicas.

1. García, J. L., Sánchez-Puelles, J. M., García, P., López, R., Ronda, C. & García, E. (1986) *Biochem. Biophys. Res. Commun.* **137**, 614–619.
2. López, R., Sánchez-Puelles, J. M., García, E., García, J. L., Ronda, C. & García, P. (1986) *Mol. Gen. Genet.* **204**, 237–242.
3. Sánchez-Puelles, J. M., Ronda, C., García, J. L., García, P., López, R. & García, E. (1986) *Eur. J. Biochem.* **158**, 289–293.
4. Howard, L. V. & Gooder, H. (1974) *J. Bacteriol.* **117**, 796–804.
5. Sánchez-Puelles, J. M., Ronda, C., García, E., Méndez, E., García, J. L. & López, R. (1986) *FEMS Microbiol. Lett.* **35**, 163–166.
6. García, E., García, J. L., Ronda, C., García, P. & López, R. (1985) *Mol. Gen. Genet.* **201**, 225–230.
7. García, P., García, J. L., García, E. & López, R. (1986) *Gene* **43**, 265–272.
8. Ronda, C., García, J. L., García, E., Sánchez-Puelles, J. M. & López, R. (1987) *Eur. J. Biochem.* **164**, 621–624.
9. Ronda, C., López, R., Tapia, A. & Tomasz, A. (1977) *J. Virol.* **21**, 366–374.
10. García, P., García, E., Ronda, C., López, R. & Tomasz, A. (1983) *J. Gen. Microbiol.* **129**, 489–497.
11. García, J. L., García, E., Arrarás, A., García, P., Ronda, C. & López, R. (1987) *J. Virol.* **61**, 2573–2580.
12. García, P., Méndez, E., García, E., Ronda, C. & López, R. (1984) *J. Bacteriol.* **159**, 793–796.
13. Tomasz, A. (1967) *Science* **157**, 694–697.
14. Giudicelli, S. & Tomasz, A. (1984) *J. Bacteriol.* **158**, 1188–1190.
15. Briese, T. & Hakenbeck, R. (1986) *Eur. J. Biochem.* **146**, 417–427.
16. Höltje, J. V. & Tomasz, A. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1690–1694.
17. Tomasz, A. (1968) *Proc. Natl. Acad. Sci. USA* **59**, 86–93.
18. Reanney, D. C. & Ackermann, H. W. (1982) in *Advances in Virus Research*, eds. Lauffer, M. A., Bang, F. B., Maramorosch, K. & Smith, K. M. (Academic, New York), Vol. 27, pp. 205–280.
19. Boyer, H. & Roulland-Dussoix, D. (1969) *J. Mol. Biol.* **41**, 459–474.
20. Messing, J. (1983) *Methods Enzymol.* **101**, 20–78.
21. López, R., Ronda, C., García, P., Escarmís, C. & García, E. (1984) *Mol. Gen. Genet.* **197**, 67–74.
22. Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
24. Lipman, D. J. & Pearson, W. R. (1985) *Science* **227**, 1435–1441.
25. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
26. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
27. Adhya, S. & Gottesman, M. (1978) *Annu. Rev. Biochem.* **47**, 967–996.
28. Tinoco, I., Jr., Borer, P. N., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London) New Biol.* **246**, 40–41.
29. Fouche, P. B. & Hash, J. H. (1978) *J. Biol. Chem.* **253**, 6787–6793.
30. Botstein, D. (1980) *Ann. N.Y. Acad. Sci.* **354**, 484–451.
31. Susking, M. M. & Botstein, D. (1978) *Microbiol. Rev.* **42**, 385–413.
32. Kamp, H. H., Wirtz, K. W. A., Baer, P. R., Slotboom, A. J., Rosenthal, A. F., Paltauf, F. & Van Deenen, L. L. M. (1977) *Biochemistry* **16**, 1310–1316.
33. de Geus, P., Verheij, H. M., Riegman, N. H., Hoekstra, W. P. M. & de Haas, G. H. (1984) *EMBO J.* **3**, 1799–1802.
34. Canfield, R. E. (1963) *J. Biol. Chem.* **238**, 2698–2707.
35. Grutter, M. G., Weaver, L. H., Gray, T. M. & Matthews, B. W. (1983) in *Bacteriophage T4*, eds. Mathews, C. K., Kutter, E. M., Mosig, C. & Berget, P. B. (Am. Soc. Microbiol., Washington, DC), pp. 356–360.