MM1, a Temperate Bacteriophage of the Type 23F Spanish/USA Multiresistant Epidemic Clone of *Streptococcus pneumoniae*: Structural Analysis of the Site-Specific Integration System

EMMANUEL GINDREAU,† RUBENS LÓPEZ,* AND PEDRO GARCÍA
Centro de Investigaciones Biológicas, CSIC, Velázquez 144, 28006 Madrid, Spain

Received 23 March 2000/Accepted 23 May 2000

We have characterized a temperate phage (MM1) from a clinical isolate of the multiply antibiotic-resistant Spanish/American 23F *Streptococcus pneumoniae* clone (Spain23F-1 strain). The 40-kb double-stranded genome of MM1 has been isolated as a DNA-protein complex. The use of MM1 DNA as a probe revealed that the phage genome is integrated in the host chromosome. The host and phage attachment sites, *attB* and *attP*, respectively, have been determined. Nucleotide sequencing of the attachment sites identified a 15-bp core site (5'-TTATA ATT CAT CGC-3') that has not been found in any bacterial genome described so far. Sequence information revealed the presence of an integrase gene (*int*), which represents the first identification of an integrase in the pneumococcal system. A 1.5-kb DNA fragment embracing *attP* and the *int* gene contained all of the genetic information needed for stable integration of a nonreplicative plasmid into the *attB* site of a pneumococcal strain. This vector will facilitate the introduction of foreign genes into the pneumococcal chromosome. Interestingly, DNAs highly similar to that of MM1 have been detected in several clinical pneumococcal isolates of different capsular types, suggesting a widespread distribution of these phages in relevant pathogenic strains.

*Streptococcus pneumoniae* is an important human pathogen and is presently the leading cause of pneumonia, meningitis, and bloodstream infections in the elderly and one of the main causes of middle ear infections in children. In addition, pneumococcal resistance to β-lactam antibiotics has been as high as 33.5% in the United States (54). Most of these resistances have been achieved as the result of interspecific gene transfers of DNA fragments between pneumococci and phylogenetically close species that colonize the same ecological niche (i.e., the nasopharynx), leading to acquisition of low-affinity penicillin-binding proteins (26). It has been reported that among the mechanisms of DNA transfer, lysogenic conversion by bacteriophages appears to be advantageous in several bacterial systems (36). The role of phages in the evolution and transfer of bacterial virulence determinants is a topic of increasing research (10, 58), and the potential use of bacteriophages for therapy and prophylaxis for antibiotic-resistant bacteria has been suggested (3, 37).

Pneumococcal phages have been a subject of continuous interest in our laboratory since the isolation of these phages was first reported (34, 55). The biological properties of several lytic and temperate phages infecting *S. pneumoniae* have been recently reviewed (24). The presence of temperate phages in fresh clinical isolates of *S. pneumoniae* was reported many years ago (4, 5). The outstanding similarity between the *lytA* gene, coding for the major lytic enzyme of pneumococcus, and the corresponding lytic genes coding for several pneumococcal phage amidases (19, 46) has led to the preparation of a probe, pCE3, based in the use of the 5'-end moiety of the *lytA* gene (16). This probe has been used to detect lysogenic strains of pneumococcus (A. Fenoll, personal communication). A recent survey carried out on clinical isolates of pneumococci by using the whole *lytA* as a probe confirmed and extended previous observations on the high incidence (about 75%) of prophage carriage among natural isolates (44). However, these two procedures have severe limitations, since strains containing remnants of the *lytA* gene in the genome might provide erroneous data on the real presence of pneumococcal phages in clinical samples (45). Nevertheless, the lysis of fresh isolates after treatment with mitomycin C and the observation of phage-like particles in the crude supernatants of these lysates also suggested the presence of a high proportion of temperate phages in clinical strains of pneumococcus. Due to the well-documented difficulties in isolation and purification of pneumococcal phages (24), none of these interesting observations provides an easy way to carry out detailed molecular characterization of some of the temperate phages in order to develop reliable studies that might document the real value of these phages as vehicles of virulence genes. Furthermore, the abundant presence of temperate phages in pneumococcus might influence genetic variation in natural populations of *S. pneumoniae*. That is, the bacterium-phage coevolution might result in several attributes in pathogenic microorganisms. For example, it has been suggested that phage infection may be a requirement in the pathogenesis of Shiga-like toxin-producing *Eschericha coli*-associated diseases (58). Currently, microbial pathogens such as *S. pneumoniae* are developing great new strategies to guarantee their own survival and expansion. As already documented for many other bacteria, phages might be important vehicles to introduce new factors that microbes can eventually use to cause infection and disease (38).

The multiresistant 23F Spanish clone (Spain23F-1) is the best example to illustrate the rapid spread of drug resistance, in this case originally detected in Spain and then rapidly disseminated to other parts of the world (40). A recent study conducted in 38 states of the United States revealed that of 328 isolates highly resistant to penicillin (≥2.0 μg/ml), about 40% belonged to the Spain23F-1 clone (35). In this study, we have purified and characterized a temperate phage, named MM1, isolated from the
23F strain 949. Moreover, we have also determined the attP and attB attachment sites, as well as the phage integrase gene required for site-specific recombination. A nonreplicative vector based on phage integration elements has been constructed and shown to be able to integrate in a specific attB site in the S. pneumoniae chromosome. To our knowledge, a detailed analysis of the phage integration system in pneumococcus had not been previously documented.

MATERIALS AND METHODS

Bacterial strains, bacteriophages, plasmids, and growth conditions. The bacterial strains, bacteriophages, and plasmids used in this study are listed in Table 1. S. pneumoniae was grown in C medium (29) supplemented with yeast extract (0.8 mg/ml) (Difco Laboratories) at 37°C without shaking, and the growth was monitored with a Karch 2100N nephelometer. E. coli was grown in Luria-Bertani medium at 37°C with shaking. Phage MM1 was induced from the lysogenic strain S. pneumoniae.

<table>
<thead>
<tr>
<th>Material Description</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Bacterial strains</strong></td>
<td></td>
</tr>
<tr>
<td>S. pneumoniae</td>
<td></td>
</tr>
<tr>
<td>949 Lysogenic for phage MM1; serotype 23F</td>
<td>40</td>
</tr>
<tr>
<td>496 Serotype 23F</td>
<td>40</td>
</tr>
<tr>
<td>499 Serotype 23F</td>
<td>40</td>
</tr>
<tr>
<td>622 Serotype 23F</td>
<td>40</td>
</tr>
<tr>
<td>8249 Serotype 19A</td>
<td>32</td>
</tr>
<tr>
<td>CSUB 3409 Serotype 9V</td>
<td>J. Linhares</td>
</tr>
<tr>
<td>CSUB 4086 Serotype 14</td>
<td>J. Linhares</td>
</tr>
<tr>
<td>SSISP33C/1 Serotype 33C</td>
<td>Statens Seruminstitut</td>
</tr>
<tr>
<td>SSISP33F/1 Serotype 33F</td>
<td>Statens Seruminstitut</td>
</tr>
<tr>
<td>746 Lysogenic for phage HB-746; serotype 8</td>
<td>47</td>
</tr>
<tr>
<td>708 Hex</td>
<td>53</td>
</tr>
<tr>
<td>M24 S′ hytA Hex∥</td>
<td>20</td>
</tr>
<tr>
<td>M222 Hex</td>
<td>17</td>
</tr>
<tr>
<td>PM11 708 pAPU1 integrant strain</td>
<td>This study</td>
</tr>
<tr>
<td>E. coli DH5α superE44 lacU169 (λ 80 lacZDM15) hsdR17 recA1 endA1 gyrA96 thi-1 relA1</td>
<td>50</td>
</tr>
<tr>
<td><strong>Plasmids</strong></td>
<td></td>
</tr>
<tr>
<td>pUC191 pUC derivative; Ln′; 4.1 kb</td>
<td>1</td>
</tr>
<tr>
<td>pBR325 derivative; Cm′; harbors the 5′ moiety of hytA</td>
<td>16</td>
</tr>
<tr>
<td>pAPU1 pUC191::EcoRI-PstI M1M int-attP; 5.6 kb</td>
<td>This study</td>
</tr>
<tr>
<td><strong>Phages</strong></td>
<td></td>
</tr>
<tr>
<td>MM1 Temperate phage from strain 949</td>
<td>This study</td>
</tr>
<tr>
<td>HB-746 Temperate phage from strain 746</td>
<td>47</td>
</tr>
<tr>
<td>Dp-1 Virulent phage</td>
<td>34</td>
</tr>
<tr>
<td>Cp-1 Virulent phage</td>
<td>48</td>
</tr>
<tr>
<td>EJ-1 Temperate phage from strain 101/87</td>
<td>13</td>
</tr>
<tr>
<td>L2 Virulent phage</td>
<td>55</td>
</tr>
<tr>
<td><strong>Primers</strong></td>
<td></td>
</tr>
<tr>
<td>EGP2 5′-GCAATTATATCATTATTCTCTCC-3′</td>
<td>This study</td>
</tr>
<tr>
<td>EGP4 5′-GAAGATAGGAGGATAAACTGG-3′</td>
<td>This study</td>
</tr>
<tr>
<td>EGP8 5′-GGAATTCCCCACACTCAAAATTTGCGC-3′</td>
<td>This study</td>
</tr>
<tr>
<td>EGP9 5′-AACTGCAGAAATTGTTCTTTCACCGCAGG-3′</td>
<td>This study</td>
</tr>
<tr>
<td>EGP14 5′-CCATACAGACACATCTGCC-3′</td>
<td>This study</td>
</tr>
<tr>
<td>EGP15 5′-CATATTGTAGACCATGAGGC-3′</td>
<td>This study</td>
</tr>
</tbody>
</table>

The bacterial strains, bacteriophages, plasmids, and growth conditions. The bacterial strains, bacteriophages, and plasmids used in this study are listed in Table 1. S. pneumoniae was grown in C medium (29) supplemented with yeast extract (0.8 mg/ml) (Difco Laboratories) at 37°C without shaking and the growth was monitored with a Karch 2100N nephelometer. E. coli was grown in Luria-Bertani medium at 37°C with shaking. Phage MM1 was induced from the lysogenic strain S. pneumoniae.

**Preparation of MM1 antiserum.** Purified MM1 phage particles were mixed with an equal volume of Freund's complete adjuvant. The preparation was injected subcutaneously, and the rabbit was reinoculated four more times, at 15-day intervals, with Freund's incomplete adjuvant. Each inoculation was done

**Southern hybridization.** Restricted DNA fragments were separated on a 0.7% (wt/vol) agarose gel and transferred to Hybond N⁺ membranes (Amersham Pharmacia Biotech) by vacuum blotting with blotter model 785 (Bio-Rad Laboratories) as described by the supplier. For determination of the attB chromosomal location, DNA from the S. pneumoniae strain M24 was restricted with Apal, SacII, or Smal enzyme and run in a 1% agarose gel by the pulsed-field gel electrophoresis (PFGE) technique as previously described (2), using a contour-clamped homogenous electric field DRII apparatus (Bio-Rad). DNA fragments were then blotted as described by Southern (52). DNA probes were digoxigenin labeled with a DNA labeling and detection kit (Roche, Mannheim, Germany). Hybridizations were carried out at 65°C, and detections were performed as recommended by the supplier.

**Electron microscopy.** Phage particles purified as described above were dia- lyzed against 0.1 M ammonium acetate (pH 7.0) and negatively stained with 1% sodium phosphotungstate. Samples were examined at 80 kV in a Philips EM 300 electron microscope.
using 15 μg of phage proteins. The serum was collected 15 days after the last inoculation.

**DNA sequencing.** DNA sequencing was carried out by using an ABI Prism 377 DNA Sequencer (Applied Biosystems, Inc.). DNA and protein sequences were analyzed with the PC/GENE software package version 6.85 (Intelligenetics, Mountain View, Calif.) or using the programs present in the Deambulum (http://www.infobiogen.fr) and National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov) sites. Sequence similarity searches were performed using the EMBL/GenBank, SWISS-PROT, and PIR databases.

**Nucleotide sequence accession numbers.** The nucleotide sequence data for the attP, attR, attB, and attL-containing fragments have been deposited in GenBank under accession numbers AJ400629, AJ400630, AJ400631, and AJ400632, respectively.

**RESULTS**

**Isolation and characterization of phage MM1.** In a survey to look for the presence of temperate phages in freshly clinical isolates of pneumococcus, using plasmid pCE3 as a probe, we paid special attention to the pneumococcal strain 949, which belongs to the Spain23F-1 clone. The presence of two hybridization bands was the first hint suggesting that this strain contained a temperate phage. The lysis of the pneumococcal strain 949 when the culture was treated with mitomycin C gave additional support to the hypothesis of the presence of a temperate phage. Moreover, purification of the lysed culture revealed a bluish band after two CsCl gradients. Electron microscopy of the purified particles showed that this phage, named MM1, belongs to the Siphoviridae family, with an icosahedral head (60 nm in diameter), and a long tail (160 nm in length) (Fig. 1A). SDS-PAGE of MM1 and five different pneumococcal phages showed that MM1 virions contained two main bands of 36 and 22 kDa (Fig. 1C). A Western blot analysis using a polyclonal antiserum raised against the MM1 virion revealed common bands between phages MM1 and HB-746, and a faint band with a protein of phage Cp-1, whereas no signal was detected with any other phages in our collection (Fig. 1D). It should be mentioned that HB-746 is a phage that was originally isolated from a type 8 strain (5). From this preliminary characterization, it was concluded that phage MM1 might share several traits with HB-746, a temperate phage that has the peculiarity of having a protein covalently bound to the 5’ ends of its DNA (47). To test whether phage MM1 also has this characteristic, we prepared DNA from purified virions of MM1 that had been treated or not with proteinase K before phenol extraction. As clearly illustrated in Fig. 1B, the DNA remained at the top of the gel in the sample that was not treated with proteinase K, a peculiarity attributed to the presence of a DNA-protein complex (47), whereas the proteinase K-treated DNA migrated normally into the gel. The stability of the DNA complex was also tested by treatment with different chaotrophic agents and conditions that affect ionic and hydrophobic associations (21); e.g., MM1 DNA treated with 2% SDS and 2% mercaptoethanol at 65°C for 10 min or with 6 M urea at 37°C for 30 min did not penetrate the agarose gel (data not shown). In spite of these common traits between the MM1 and HB-746 phages, restriction enzyme digestions with HindIII, PstI, and PvuII revealed that they are different phages. A molecular size of about 40 kb for MM1 DNA was determined from the sum of the sizes of DNA fragments obtained with several restriction enzymes. Furthermore, PFGE analysis of the entire DNA also indicated that the molecular size was ca. 40 kb (see Fig. 8A).

**Identification of the attachment sites.** To locate the phage attachment site of MM1, attP, DNAs from the lysogenic strain 949 and the MM1 phage were digested with different restriction enzymes, the resulting fragments were separated by electrophoresis and blotted, and the membrane was hybridized with MM1 DNA as a probe. When we compared the DraI digestions, all of the hybridization bands of MM1 were present in the lysogenic strain DNA, except for a 1.7-kb fragment and a new 4-kb band found in the prophage pattern (Fig. 2A). This finding was consistent with a recombination event occurring between the attP site, located in the 1.7-kb fragment, and the bacterial attachment site, attB, resulting in the splitting of attP and the formation of a new 4-kb junction fragment. The second junction fragment is probably overlapped by one of the other restriction fragments. The phage 1.7-kb DraI fragment was isolated, cloned into Smal-digested pUC19, and used as a probe in the same membrane. Following this procedure, a second junction fragment (1.3-kb band) was detected (Fig. 2B). The faintness of the hybridization signal was due to the small portion of phage DNA present in this junction site (see below).

We used an inverse PCR strategy to clone the junction fragments. attR was amplified as follows. Fifteen micrograms of DraI-digested S. pneumoniae 949 DNA was self-ligated in a 400-μl reaction volume, precipitated, and used for PCR amplification with oligonucleotides EG2P and EG4P. The resulting 2.5-kb fragment was purified from an agarose gel and sequenced. Despite several attempts, attL could not be amplified by this method. We then took advantage of the partial S. pneumoniae genome sequence (http://www.tigr.org) and our sequence of attP and attR to look for a locus of identity, probably located in the attB zone. This region was found in the contig sp_100, and an oligonucleotide, EGP14, deduced from the bacterial sequence was designed. The attL sequence was amplified, using oligonucleotides EGP9 and EGP14, as a 0.6-kb fragment that was purified and sequenced. Finally, taking into account that most lysogenic bacteria suffer spontaneous phage excision from the bacterial chromosome resulting in attB site restoration, we could amplify attB with oligonucleotides EGP14 and EGP15, deduced from the bacterial attL and attR, respectively, using DNA from strain 949 as the template. The resulting 1,852-bp amplicon was sequenced.

Alignment of the sequence obtained from the att sites revealed a 15-bp core site (5'TTATAATTCATCCGC-3') where the site-specific recombination process presumably takes place (Fig. 3 and 4). Searches in the databases revealed a single site in the S. pneumoniae genome but not in the other bacterial genomes already sequenced. Furthermore, comparison of the bacterial sequences from the attB, attL, and attR sites showed that strain 949 displays 99.6% identity at the nucleotide level with the corresponding region of the S. pneumoniae type 4 genome. The observed point mutations do not affect the genomic arrangement of the attB-containing region.

Analysis of the 1.7-kb DraI fragment, which contains the attP site, showed that the core was present in this fragment but that the entire attP region probably was not, since a DraI site lies 4 bp downstream of the 3’ end of the core. This observation explains the low intensity of the hybridization signal obtained with the 1.3-kb junction site (Fig. 2B), as this fragment overlaps with the probe only in the core region. The complete sequence of the attP region was obtained from a cloned 3.2-kb HaeIII fragment that contains the lytic enzyme gene of phage MM1, mml, and overlaps with the 1.7-kb DraI fragment (unpublished results). Sequence analysis of the attP-containing region showed the presence of several direct repeats, including the 8-bp-long sequence 5’-TGCCCCCTT-3’, which is repeated four times in the core surrounding region. Furthermore, four putative integration host factor binding sites, very similar to the consensus sequence deduced from attP sites of lambdoid phages 5’-(CTT/AANNNTGTG(A/T)-3’) (30), were also found. Two hairpin structures (L1 and L2) with free energies of −19.4 and −7.8 kcal/mol, respectively, are present in this
region (Fig. 4). They could behave as rho-independent terminators for the two open reading frames (ORFs) that flank the core region.

Analysis of the attB region showed that the core overlaps the 3’ end of an ORF (O06975) coding for a 303-amino-acid-long protein which has 42% identity and 78% similarity with a protein of unknown function of *Bacillus subtilis* (accession no. O06975). Remarkably, integration of the phage DNA in the bacterial chromosome does not disrupt the sequence of this ORF, as its stop codon is present in the core. Sequence analysis of this region also showed the presence of the 3’ end of an ORF (O05268) lying 34 bp to the left of the attB core (Fig. 3) and oriented in the opposite direction compared to O06975.

The amino acid sequence of O05268 was deduced from the contig sp_100 and showed that O05268 could code for a 321-amino-acid-long protein displaying 54% identity and 80% similarity with the thioredoxin reductase from *B. subtilis* (accession no. O05268). To locate the attB site on the pneumococcal chromosomal map (25), we used an attB-specific probe amplified with oligonucleotides EGP14 and EGP15, deduced from the bacterial regions of attL and attR, respectively. Restriction fragments from the pneumococcal DNA isolated from strain M24 that were Southern blotted, subjected to PFGE, and hybridized with this probe revealed a single band in each case, corresponding to *Apa*I fragment 5, *Sac*II fragment 8, and *Sma*I number 4 (Fig. 5A). We also show in Fig. 5B a fragment of the
contig sp._29. We have observed that contigs sp._100 and sp._29 contained partial regions of O85254, and the gene lytC, recently demonstrated to code for the first identified pneumococcal lysozyme (23), is located in the 3’ end of contig sp._29.

Identification of the integrase. The complete sequencing of the 1.7-kb DraI fragment reported above allowed the identification of a first ORF coding for a 116-amino-acid protein and a second ORF, named int, which coded for a 375-amino-acid protein. It is preceded by a Shine-Dalgarno sequence (5’-GA GGT-3’) located 8 bp upstream of the start codon. The stop codon of the int gene is located 87 bp upstream of the attP core. BLAST searches performed with the Int sequence identified similar site-specific recombinases belonging to the λ integrase family. The best score, 33% identity and 70% similarity, was obtained with the Staphylococcus aureus phage φPVL integrase. An alignment performed with the MM1, φPVL, and λ phage integrases is shown in Fig. 6, where only the regions corresponding to the conserved boxes and patches defined by Esposito and Scocco (15) and Nunes-Dúby et al. (41) are shown. Box I contains the conserved Arg residue, and the triad His-Arg-Tyr is present in box II. These four amino acids, which are involved in the recombinase activity and considered a hallmark of the λ Int family recombinases, are present at the expected positions in the MM1 Int. In addition, the three patches containing charged amino acids and highly conserved, precisely spaced, hydrophobic residues could be found in the MM1 Int sequence. The presence of the previously identified boxes and patches, the global sequence similarity with other λ integrases, and the location of int close to the attP site strongly suggested that this ORF encodes the MM1 integrase. A sche-
matic representation of the site-specific integration of MM1 DNA into the pneumococcal chromosome is depicted in Fig. 3.

**Construction of an integrative vector.** To demonstrate that attP and the int gene were actually sufficient to mediate site-specific integration into the pneumococcal chromosome, a 1,527-bp DNA fragment, embracing the int-attP cassette, was PCR amplified using oligonucleotides EGP8 and EGP9. These oligonucleotides are located 112 bp upstream from the int start codon and 186 bp downstream from the 3' end of the attP core, respectively. This fragment contains the putative promoter, the structural gene of the integrase, and 186 bp downstream of the core of the attP site. The 1,527-bp amplified product was restricted with the enzymes EcoRI and PstI and ligated into the EcoRI- and PstI-digested pUCE191 plasmid (Fig. 7A). The 5,599-bp recombinant plasmid (pIAPU1) was first introduced into *E. coli DH5α* and then transferred into *S. pneumoniae* 708 by transformation and selected for lincomycin resistance. Since pIAPU1 is a nonreplicative plasmid in *S. pneumoniae*, lincomycin resistance is expressed upon chromosomal integration. Two clones, named PM11 and PM12, were chosen, although the amplification analysis revealed the same pattern for both transformants. Site-specific integration of a single copy of pIAPU1 in the attB site should lead to the detection of two fragments of 5,600 and 3,150 bp when chromosomal DNA is cut with NcoI. In fact, this was the case when the DNA prepared from PM11 was restricted with this enzyme, electrophoresed, and blotted and the membrane was probed with the 1,527-bp int-attP-containing fragment (Fig. 7B).

To confirm the site specificity and to evaluate the stability of the integration event of PM11, we also performed PCR analysis. attL and attR sites could be amplified using genomic DNAs from the lysogenic parental strain 949 and from the transformant PM11 strain as templates (Fig. 7C, lanes 8, 9, 11, and 12), demonstrating the specificity of the integration of the plasmid in the bacterial chromosome. attB-containing amplicons could be detected using DNAs from 949 and 708 (Fig. 7C, lanes 4, and 6), and attP was amplified when using DNA from 949 (Fig. 7C, lane 3). The fact that attB and attP were amplified...
when using DNA from 949 as a template clearly reflects a spontaneous induction event leading to the release of phage progeny in the lysogenic culture and then restoring intact attB and attP sites. These two sites could not be amplified when using DNA from PM11, showing the stability of the integrated copy of the integrative plasmid pLAPU1 in the bacterial chromosome.

**Lysogeny in different pneumococcal strains.** To test the incidence of this particular type of phage among different clinical isolates, we analyzed three other strains belonging to the Spain23F-1 clone, as well as other isolates from serotypes 19A (strain 8249, an important multiresistant strain, originally isolated in South Africa), 14 (CSUB 3409), 9V (CSUB 4086), 33C (SSISP33C/1), and 33F (SSISP33F/1). A first experimental approach was based on the comparison of the growth curves of several strains treated with mitomycin C or not. Later, total DNAs prepared from these pneumococcal cultures were subjected to PFGE and Southern blotting using MM1 DNA as a probe. As shown in Fig. 8A, three isolates belonging to the 23F serotype, as well as strains 8249 and SSISP33C/1, showed an extra chromosomal hybridization band of about 40 kb, whereas in the case of the 496, CSUB 3409, CSUB 4086, and SSISP33F/1 strains there were no visible bands. These results might suggest that some of the phages released after induction had a close relationship. Nevertheless, since we had reported the high similarity between the lytic genes from pneumococcal phages and their host (19), the common hybridization bands could be simply attributed to the presence of htaA-like genes in phages that otherwise could be very different when more accurate analysis are performed. Hence, we digested the bacterial DNAs of the strains mentioned above, along with 746 and M222 as controls, with HindIII, and the fragments generated by these digestions were probed with MM1 DNA. The result showed three different patterns (Fig. 8B): (i) strains 949, 499, 622, 746, and 8249 had very similar, but not identical, hybridization profiles (lanes 1 to 5); (ii) strain SSISP33C/1 showed one strong and several weak hybridization bands (lane 9); (iii) strains CSUB 3409 and CSUB 4086 only had two weak bands at different positions from the other ones (lanes 6 and 7). Strains 496 and SSISP33F/1 gave no bands, except the common 1.2-kb fragment containing the host lytic gene, htaA (22).

The simplest explanation for these findings could be that, in the first case, the temperate phages are very similar, despite the distinct strains from which they were isolated and different geographic origins of the corresponding strains. Besides, the phage of strain SSISP33C/1 is more distant from these phages in terms of genome similarity, and strains CSUB 3409 and CSUB 4086 harbor the most different phages of the strains tested here, which most likely are defective phages because there were no bands visible in the PFGE-Southern blot analysis (Fig. 8A). Another possibility was that strains CSUB 3409 and CSUB 4086 contained some phage remnants in their chromosomes, as has already demonstrated for other organisms (33), including pneumococcus (45).

**DISCUSSION**

Double-stranded-DNA-containing bacteriophages infect a large diversity of bacterial hosts and probably are the most abundant group of similar organisms in the biosphere (27). In *S. pneumoniae* only a few examples of well-characterized phages have been described, but they have revealed a striking morphological and physiological variety (24). In two recent reports, it has been claimed that 76% of the pneumococcal clinical isolates carried a prophage in the chromosome (44, 51). This was based on the presence of two or more chromosomal *Sma*I fragments that hybridized with a htaA probe in lysogenic strains, with one of these bands corresponding to the gene coding for the major host autolysin and the other corresponding to the gene present in the temperate phage coding for the phage lysis (16). Nevertheless, in these reports there was no indication of phage purification, which is a limiting factor to establish any relationship among these temperate phages. To investigate a precise biological role that provides clues about the ubiquitous presence of phages in clinical isolates of virulent pneumococci, we decided to investigate several clones of the multiresistant 23F serotype and tried to purify one phage (MM1) from the Spain23F-1 clone. This clone was selected because strains resistant to penicillin, tetracycline, and chloramphenicol and variably resistant to erythromycin have been spread worldwide (35). The genome of phage MM1 appears to be quite similar, but not identical, to that of phage HB-746 isolated from strain 746, a type 8 pneumoccus. In fact, from preliminary sequence data of the HB-746 genome obtained from PhageTech, Inc. (Montreal, Canada), we know that the DNA fragment encompassing the integrase gene of MM1 does not display significant similarity to the HB-746 genome, although at the protein level, the corresponding ORF of HB-746 has 41% identity with the integrase of MM1. These two phages also share the peculiar characteristic of having DNA-protein complexes that are capable of becoming integrated into the host chromosome. The determination of the precise biological role of the proteins covalently linked to the DNAs of these pneumococcal temperate phages still remains a mystery. We have previously suggested a protective role for incoming DNA after the phage has injected the DNA into the...
host bacteria or a function during integration, assuming that the bound protein has retained an enzymatic activity (47). Similar mechanisms have been postulated to explain the integrative process of the T-DNA molecules generated in the *Agrobacterium tumefaciens* system for transferring the DNA into the genome of the host plant (28). In the case of these pneumococcal phages there must be an additional mechanism allowing the phage genome to regain the terminal protein when these peculiar temperate phages enter the lytic cycle. Furthermore, the study of the genome of MM1 might provide a reasonable way to investigate genes involved in the mechanism leading to the programmed loss and recovery of the terminal protein when shifting from the lytic to the lysogenic cycle and vice versa.

In work on the molecular characterization of MM1, we have analyzed the genetic determinants required for phage DNA integration. Temperate bacteriophages integrate their DNAs into the host chromosome by a site-specific recombination process following the Campbell model (9). Two specific attachment sites, one on the bacterial chromosome (*attB*) and the other on the phage genome (*attP*), are recombined by the activity of a phage-encoded integrase. Although the integration system of phages of gram-positive bacteria is less well documented, data are available for several phages of *S. aureus* (11, 31, 59), for bacteriophage T12 of *Streptococcus pyogenes* (36), for the actinophage RP3 (18), and for several lactic acid bacterial phages (8, 14, 42, 57). We have now identified the *attP*-containing phage DNA, the bacterial attachment site *attB*, and the host-phage junctions *attL* and *attR* of the MM1 prophage. These sequences share a 15-bp identity region, a typical size for the cores of other phages that integrate through a site-specific recombination mechanism. The *attP* re-
region, 313 bp long, is located between two ORFs, coding for the integrase and the lytic enzyme, that are convergently transcribed. The attP site has several traits in common with other attP sites, such as a high percent A+T (67%) (although not significantly higher than that of the host DNA) (49) and a complex array of direct and inverted repeats. These repeat sequences have been postulated to be the binding or recognition sites for phage-encoded proteins such as integrase or excisionase or for host factors analogous to the E. coli integration host factor (12), although precise assignment of these sites in phage MM1 must await the purification of these proteins. From the nucleotide sequence of the MM1 integration region we deduced an ORF encoding a polypeptide of 375 amino acids located adjacent to attP and transcribed towards it. Evidence suggesting that this ORF encoded the MM1 integrase came from the location, size, and similarity to site-specific recombinases of the λ integrase family. This was confirmed by the construction of a functional vector promoting site-specific integration. This report represents the first demonstration of such a mechanism carried out for a pneumococcal phage, since with the other two temperate phages previously studied, we did not succeed in accurately sequencing the attachment sites (13, 45).

The elucidation of the determinants required for the integration of MM1 has allowed the construction of a site-specific integration vector for S. pneumoniae. Although this gram-positive bacterium possesses a remarkable and well-characterized mechanism for incorporating foreign DNA, because of its natural competence, a site-specific vector like pIAPU1 described here presents some new advantages over the alternative methods developed to take up genetic material. As an example, this vector will help in the specific insertion of any heterologous gene into the attB site in a single copy, which could eventually be useful for gene expression studies with this important human pathogen.

We do not know whether orf116, which precedes the integrase gene, could function as an excisionase, since the typical traits of this kind of protein have not been identified in the orf116 product. On the other hand, 208 bp downstream of the attP core region (Fig. 4) we could locate the 3’ end of the lytic gene, which, together with the holin gene, forms part of the lytic cassette of this phage (unpublished observations). Concerning attB, it is interesting that MM1 integrates into a gene of unknown function with the peculiarity that the stop codon of this gene is included in the core region of attP, which implies that lysogenization of the host does not inactivate this gene.

The unexpected similarities found between phage MM1 and the HB family of phages prompted us to undertake a broader examination of the presence of this phage among different multiresistant isolates belonging to other capsular serotypes. From the examples presented here we can conclude that an MM1-like phage appears to be spread among several of the most abundant pneumococcal strains studied. Moreover, the presence in strains SSISP33C/1, CSUB 3409, and CSUB 4086 of another phage(s) very different from MM1 can be expected (Fig. 8). On the other hand, we do not favor the idea, suggested by Bernheimer (4, 5), that lysogeny is associated with only certain pneumococcal capsular types, since in a more general context, we have found that the presence of fully functional defective or remnant prophages in the chromosome is indeed a general trait among pneumococcal isolates, including some atypical pneumococcci that are deoxycholate insensitive using a classical taxonomic test (13). Botstein has proposed that evolution of lambdoid phages happens by exchange of genes organized in functional modules (7). The biological and functional similarities between phage-encoded enzymes and the host pneumococcal amidase have been well documented (46). One of the highest levels of identity between bacterial and phage genes (87.1%) was observed when comparing the host lytA and lbbH from the HB-3 temperate phage (46). This nucleotide sequence similarity allows recombination between both genomes that permits restructuring and evolutionary adaptation in both organisms. In addition, it has been established that lactococcal phages are able to acquire pieces of the host chromosome (39).

The high incidence of lysogeny among clinical strains has raised the possibility that part of the exchange of genetic information found in pneumococcus in vivo is carried out through transduction or is facilitated by phage functions (44). Actually, it is well documented that some temperate phages bear virulence-related genes in many bacterial systems (reference 27 and references therein), and a process similar to transduction but requiring competence development was described previously for pneumococci (43). The interchange of capsular polysaccharides has been revealed to be a quite common process in nature as a mechanism to improve serotype replacement, which can provide to pathogenic species like pneumococcus an excellent and profitable way to escape from a vaccine prepared against a limited number of capsular types, like the newly developed heptavalent vaccine. The possibility of in vivo
DNA exchange by a transfection-like mechanism is attractive, since the higher efficiency of pseudotransduction of large fragments of DNA compared to transformation could give this mechanism an advantage over transformation for the observed in vitro capsular switch events between the cassette-like organization of the genes coding for capsules (44). Phages capable of lysing unencapsulated (nonlysogenic) indicator strains have been readily isolated from carriers or patients (34, 48, 55), and the high incidence of temperate phage carriage in S. pneumoniae could strongly influence the structure of natural populations of pneumococci in their ecological niche. Incidentally, it has been suggested that defense against phage infection may be another selective pressure that facilitates the structure and maintenance of capsular polysaccharide in this species (44). In addition to all of these relevant biological roles of pneumococcal phages, the observation that a high proportion of isolates with clinical relevance carry phages (e.g., Spain234-1) also invites speculation that these phages might contribute to transfer of antibiotic resistance markers between strains through a generalized transduction-like mechanism. We now have available the possibility of sequencing the complete genomes of temperate phages isolated and purified from very relevant pneumococcal strains. This approach together with the development of an integrative phage vector will be an important tool to facilitate the study of potential virulence genes.

ACKNOWLEDGMENTS

We acknowledge E. García for encouragement and many fruitful discussions and J. L. García for critical reading of the manuscript. We thank M. Carrasco and E. Cano for technical assistance. V. Muñoz, M. Fontenla, and A. Hurtado are also acknowledged for the artwork. This work was supported by grants from Dirección General de Investigación Científica y Técnica (PB96-8089) and the Comunidad Autónoma de Madrid (08.2/0014.2/98). E.G. was the recipient of a Marie Curie Research Training Grant.

REFERENCES


