Zinc metalloproteinase genes in clinical isolates of *Streptococcus pneumoniae*: association of the full array with a clonal cluster comprising serotypes 8 and 11A

Romina Camilli,¹ Elena Pettini,² Maria Del Grosso,¹ Gianni Pozzi,^{2,3}
Annalisa Pantosti¹ and Marco R. Oggioni^{2,3}

Correspondence
Marco R. Oggioni
oggioni@unisi.it

¹Dipartimento di Malattie Infettive, Parassitarie e Immunomediate, Istituto Superiore di Sanità, Rome, Italy

^{2,3}Laboratorio di Microbiologia Molecolare e Biotecnologia², Dipartimento di Biologia Molecolare, Università di Siena, and UOC Batteriologia³, Azienda Ospedaliera Universitaria Senese, Siena, Italy

Pneumococci display large zinc metalloproteinases on the surface, including the IgA protease, which cleaves human IgA1 in the hinge region, the ZmpC proteinase, which cleaves human matrix metalloproteinase 9 (MMP-9), and two other proteinases, ZmpB and ZmpD, whose substrates have not yet been identified. Surface metalloproteinases are antigenic and have been linked to virulence. The genes encoding these proteinases reside in three distinct loci: two loci specific for *zmpB* and *zmpC*, and a third, the *iga* locus, containing *iga* and *zmpD*. Data obtained by this and other groups have shown that pneumococcal metalloproteinase genes are transcribed and yield mature and enzymatically active proteins. Since the presence of the four proteinase genes is variable in the pneumococcal strains whose genomes have been sequenced, the presence of these genes in a collection of 218 pneumococcal isolates, mostly from invasive disease, was investigated. The data showed that *zmpB* and *iga* were present in all the isolates examined, while *zmpC* and *zmpD* were present in a variable proportion of the isolates (in 18 and 49 %, respectively). Interestingly, isolates carrying both *zmpC* and *zmpD* were found to belong mainly to two serotypes (sts), 8 and 11A. By molecular typing, st 8 and st 11A isolates appeared to belong to the same clonal cluster. The presence of these two additional metalloproteinases could contribute to the fitness of particular pneumococcal clones.

Received 9 August 2005
Revised 23 September 2005
Accepted 7 October 2005

INTRODUCTION

Streptococcus pneumoniae (pneumococcus) is a common and important human pathogen. Pneumococci are the chief cause of community-acquired pneumonia and bacterial meningitis, the second-most-prevalent cause of sepsis, and one of the main causes of otitis media (Gillespie, 1989). The main virulence factors of *S. pneumoniae* are the capsule, the pneumolysin that is a cholesterol-dependent cytolysin, and some pneumococcal surface proteins and enzymes, including PspA, PspC, neuraminidase and hyaluronidase (Paton *et al.*, 1993; Mitchell, 2000; Chapuy-Regaud *et al.*, 2003; Iannelli *et al.*, 2004). A characteristic feature of *S. pneumoniae* is the presence of large zinc metalloproteinases on the cell surface. These enzymes have a common structure (Pfam Accession no. PF02395) and are present not only in *S. pneumoniae* but also in related oral streptococci, such as

Streptococcus sanguis, *Streptococcus gordonii*, *Streptococcus mitis* and *Streptococcus oralis* (Poulsen *et al.*, 1996).

From the published genome sequences, it can be inferred that *S. pneumoniae* possesses two (strain R6, rough derivative of serotype 2; Hoskins *et al.*, 2001), three (strain TIGR4, serotype 4; Tettelin *et al.*, 2001) or even four (strain G54, serotype 19F; Dopazo *et al.*, 2001) zinc metalloproteinase genes, depending on the strain. All the above-mentioned strains possess *iga* and *zmpB*, encoding the IgA protease and ZmpB, respectively. In addition, strains TIGR4 and G54 possess the metalloproteinase gene *zmpC*. In G54 only, a duplicate gene is present in the *iga* locus that is predicted to code for another zinc metalloproteinase, and that we suggest should be designated ZmpD. The most studied enzyme is the IgA protease, which cleaves human IgA1 in the hinge region (Kilian *et al.*, 1996; Poulsen *et al.*, 1996; Wani *et al.*, 1996). The importance of the IgA protease in pneumococcal virulence is evidenced by assays of bacterial adherence to respiratory epithelial cells in culture (Weiser *et al.*, 2003).

Abbreviations: MLST, multilocus sequence typing; MMP-9, matrix metalloproteinase 9; st, serotype; ST, sequence type.

Recently, the target of pneumococcal zinc metalloproteinase ZmpC has been identified as human matrix metalloproteinase 9 (MMP-9), a gelatinase which participates in the protease cascade responsible for remodelling the extracellular matrix (Oggioni *et al.*, 2003). Pathogenic processes in which MMP-9 has been implicated include the opening of the blood–brain barrier during inflammation, tissue destruction in periodontal disease and tissue invasion by tumour cells (McCawley & Matrisian, 2000). The pneumococcal enzyme ZmpC could play a role in favouring *S. pneumoniae* invasion and the crossing of tissue barriers through the cleavage and activation of human MMP-9 (Oggioni *et al.*, 2003). Knock-out TIGR4 mutants that have been deleted for *iga*, *zmpB* or *zmpC* have been shown to be attenuated in murine models of infection (Chiavolini *et al.*, 2003; Blue *et al.*, 2003).

The great majority of pneumococcal genes share from 97 to 99 % nucleotide identity among isolates (Oggioni & Pozzi, 2001; Tettelin & Hollingshead, 2004). However, the pneumococcal zinc metalloproteinases fall within a restricted group of hypervariable surface proteins that includes other important antigens, such as PspA and PspC. The hypervariability of these antigens has been recognized to be due to frequent horizontal gene transfer in these regions, enabling antigenic escape (Iannelli *et al.*, 2002). The purpose of this work was to investigate the presence of the zinc metalloproteinase genes *iga*, *zmpB*, *zmpC* and *zmpD* in a collection of *S. pneumoniae* isolates belonging to a variety of different serotypes.

METHODS

Pneumococcal strains. A total of 218 *S. pneumoniae* isolates, representing 35 different serotypes, were examined in this study. One hundred and fifty-three invasive isolates were chosen from a collection of 322 invasive isolates obtained during the years 1999–2000 as part of a nationwide surveillance of antimicrobial resistance (Dicuonzo *et al.*, 2002; Pantosti *et al.*, 2003). The strains were selected to be representative of the different serotypes circulating in Italy during those years. Out of the 153 invasive isolates, 116 (75·8 %) were from blood and 32 (20·9 %) from cerebrospinal fluid (CSF); 12·4 % of the isolates were from children aged 0–5 years and 33·9 % from elderly people (≥65 years). Forty-three strains, isolated from the nasopharynx of healthy children in different day-care centres

in Rome in 1999 (Petrosillo *et al.*, 2002), were included in the study. Only one isolate per serotype per day-care centre was examined. Additional strains belonging to serotype (st) 8 (10 isolates) and st 11A (13 isolates), obtained during surveillance conducted in Italy in the years 1997–2003, were included in the molecular typing study. The strains TIGR4 (st 4; Tettelin *et al.*, 2001), R6 (rough derivative of st 2 D39; Hoskins *et al.*, 2001) and G54 (st 19F; Dopazo *et al.*, 2001), whose complete or partial genome sequences are available, were used as reference strains. The incomplete pneumococcal genome sequence data used for BLASTN screens were produced by the Sequencing Group at the Sanger Institute and can be obtained from <ftp://ftp.sanger.ac.uk/pub/pathogens/spn/>. All pneumococcal isolates were kept as frozen stock cultures in Microbank vials (Biolife Italiana), containing a glycerol broth, at –80 °C.

PCR assays for metalloprotease genes (locus-specific PCR).

Detection of the locus for each of the four zinc metalloproteinase genes (*iga*, *zmpB*, *zmpC* and *zmpD*) was accomplished by PCR, using the Roche Expand Long Template PCR system, following the recommended procedure. Frozen pneumococcal stock culture (2 µl) was used as template. The primer pairs were designed for conserved sequences of each locus, on the basis of the published pneumococcal genomes (Table 1). The primers used to amplify *zmpB* and *zmpC* annealed upstream and downstream of the respective ORFs, thus amplifying the whole genes, if present (Fig. 1), and yielding amplicons of approximately 5800 and 8000 bp, respectively. If gene *zmpC* was missing, the predicted amplicon of the *zmpC* locus was of approximately 600 bp. Two different PCR reactions were designed to detect the presence of *iga* and *zmpD*, both of which are contained in the *iga* locus. In both assays, the forward primer (GM192) annealed to the conserved sequence upstream of *iga*; the reverse primer annealed to the sequence immediately downstream of *iga* (GM9) or inside *zmpD* (GM191), for the two reactions, respectively (Fig. 1). The presence of *iga* alone was detected by an amplification product of approximately 6500 bp in the first PCR assay and by no amplification in the second assay. If *iga* was associated with *zmpD*, the first PCR assay would give no amplification, while the second assay would amplify a segment of approximately 7000 bp spanning *iga* and a portion of *zmpD*. The PCR products were examined in UV light after electrophoresis in a 0·8 % agarose gel.

ZmpC activity assay. The activity of ZmpC was indirectly evidenced by the detection of enzymically active cleavage products of its substrate, human MMP-9, as previously described (Oggioni *et al.*, 2003). MMP-9 and enzymically active forms of MMP-9 were detected by gelatin zymography, since this human protease has gelatinase activity. In brief, pneumococcal supernatants were incubated with 2 ng recombinant human proenzyme MMP-9 (Oncogene Research Products). After 1 h incubation at 37 °C, samples were stored at –20 °C until use. Gelatin zymogram gels were prepared as

Table 1. Oligonucleotide primers used in the study

Gene detected	Primer	Oligonucleotide sequence (5'–3')	Nucleotides	GenBank accession no.
<i>zmpB</i>	GM10	TTGAGAATGAATAAAGGAGAT	634754–634774	NC_003028*
	GM11	ATCAATCACTGTTTTCTATGC	640583–640562	NC_003028
<i>zmpC</i>	GM12	AACCTTCACCGTAAGACACGT	71937–71958	NC_003028
	GM13	CTTATCTCGATAGATTCAATGCC	80064–80041	NC_003028
<i>iga</i> + <i>zmpD</i>	GM192	TTGAAATCGCACCAGAACAAGTC	1083530–1083552	NC_003028
<i>iga</i>	GM9	AAGAACTAACTACCAACAGCTA	1090078–1090056	NC_003028
<i>iga</i> + <i>zmpD</i>	GM191	GCTAGGATCAGATTCTTCTATAGTGG	119772–119797	AL449925†

*Genome sequence of *S. pneumoniae* TIGR4; †unfinished genome sequence of *S. pneumoniae* G54 (NCBI htgs database).

Fig. 1. Schematic drawing of the zinc metalloproteinase genes in the three strains for which the genome has been published (TIGR4, R6 and G54). The codes above the genes identify the ORFs in the respective genomes. The positions of the amplicons are shown below the respective genes. Primers are indicated by arrowheads whose shading matches that of the sequences from which they have been designed.

previously described (Oggioni *et al.*, 2003). After electrophoresis, gels were renatured, washed and incubated overnight to allow gelatinase activity to develop. Finally, gels were stained with Coomassie blue.

Quantitative real-time RT-PCR. Quantitative RT-PCR was performed essentially as described previously (Oggioni *et al.*, 2004). In brief, RNA extraction was performed using the SV Total RNA isolation system (Promega) according to the manufacturer's instructions. Retrotranscription was performed with random octamer primers ($10 \text{ pmol } \mu\text{l}^{-1}$), using RNAsin ($40 \text{ U } \mu\text{l}^{-1}$ RNAsin Plus RNase inhibitor, Promega) and ImProm II reverse transcriptase (Promega). Retrotranscribed samples were stored frozen. Quantitative real-time PCR was performed essentially as previously described (Oggioni *et al.*, 2004), in a LightCycler (Roche) using LightCycler DNA-Master SYBR Green I for detection. The relative gene expression was analysed using the $2^{-\Delta\Delta\text{CT}}$ method (Livak & Schmittgen, 2001). The reference gene was the housekeeping gene *gyrB* (Oggioni *et al.*, 2004). The efficiency of all primer pairs was controlled by doing PCRs on serially diluted samples of cDNA. Only primer sets with a comparable efficiency were used for quantification. The primer pairs used for gene expression analysis were designed from the sequences of strain TIGR4 and were for *iga* (SP1154) TGCGAGATTCAGAGCTATCATC and GCCTCTCATTCTTGCTTCC (159 bp amplicon), for *zmpB* (SP0664) AATGAGCGCTAGAAATGTTGT and ATTAATAATGGATGTCCAAT (124 bp amplicon), and for *zmpC* (SP0071) GATGAGTCAAGGGATTCAATCG and AAGGGCCTCTACCAAGCAAG (118 bp amplicon). All data were obtained from quadruplicate biological replicates.

Western blot. The production of zinc metalloproteinases was analysed by Western blot using standard conditions. The antiserum for the IgA protease was kindly provided by Mogens Kilian (Poulsen *et al.*, 1996), and the antiserum for ZmpB of D39 by Tim Mitchell and Gavin Paterson (Blue *et al.*, 2003). The serum for ZmpC was obtained from mice challenged with a DNA vaccine vector expressing the *zmpC* gene of TIGR4 lacking the sequence encoding the hydrophobic N terminus (results not shown).

PFGE and multilocus sequence typing (MLST). Molecular typing was performed by PFGE, following a method previously described (Pantosti *et al.*, 2000). Isolates differing by one to six bands were considered genetically related and were assigned to the same

PFGE type, but to different subtypes (Tenover *et al.*, 1995). A computer-assisted dendrogram of fragment patterns was constructed using Diversity Database Fingerprinting software, version 2 (Bio-Rad Laboratories). Clustering was obtained by the unweighted pair group method with arithmetic averages (UPGAMA) with the Dice similarity coefficient. MLST (Enright & Spratt, 1998) was performed on representative isolates as recommended at the MLST web site (<http://spneumoniae.mlst.net/>).

RESULTS

Nomenclature

The three zinc metalloproteinase genes of the TIGR4 genome deposited as SP0664, SP0071 and SP1154 (Tettelin *et al.*, 2001) have been named, respectively, *zmpB* (zinc metalloproteinase B; Bergé *et al.*, 2001), *zmpC* (zinc metalloproteinase C; Oggioni *et al.*, 2003) and *iga* (IgA protease; Poulsen *et al.*, 1996; Kilian *et al.*, 1980) (Fig. 1). In addition to these three genes, a fourth zinc metalloproteinase gene (SPN03142) is carried by strain G54 in the *iga* locus (Dopazo *et al.*, 2001) (Fig. 1). When constructing a tree using the amino acid sequences of the protease domains of the zinc metalloproteinases (Pfam accession no. PF02395), this fourth protein falls within a distinct cluster with respect to the three metalloproteinase groups above (Chiavolini *et al.*, 2003). We propose to name this fourth zinc metalloproteinase, already identified in a genomic screen for virulence factors (Polissi *et al.*, 1998), ZmpD (zinc metalloproteinase D).

Expression of *zmp* genes and characterization of gene products

In order to characterize the *zmp* genes and their gene products, we monitored the gene expression, protein production and enzymic activity of the proteases. The detection of

gene expression requires the use of probes or primers internal to the genes of interest. Especially in the case of genes under high selective pressure and which show extensive allelic variation, this restricts the analysis to a single strain. We designed real-time RT-PCR primers for the three *zmp* genes of the TIGR4 strain. The expression of the *zmp* genes was assayed by real-time RT-PCR in liquid culture, after induction of competence and during growth on agar plates. In liquid exponentially growing culture, all three *zmp* genes were expressed at a level similar to that of the control gene *gyrB*. We have reported in previous work the effect of competence-stimulating peptide (CSP)1 and CSP2 on the induction of gene expression of some selected competence genes in TIGR4 (Oggioni *et al.*, 2004). By reusing the same frozen cDNA samples collected at 5 and 10 min after the addition of CSP to pneumococcal cells, no change in the expression of any *zmp* gene was detected in response to competence induction. When assaying *zmp* expression in non-confluent colonies on agar plates after 12 h incubation, a threefold relative decrease of expression (*iga*, 3.6 ± 0.3 ; *zmpB*, 3.5 ± 0.2 ; and *zmpC*, 3.0 ± 0.6 -fold decrease) was observed.

Protein production was evaluated by Western blotting of supernatants of late-exponential-phase broth cultures of both D39 and TIGR4 (Fig. 2). IgA protease was detected in the supernatant of both strains; this was expected, since the gene is present in both strains, and the identity at the protein level is 87 %. A band reactive to ZmpB was identified only in D39. When overstaining a comparable blot, ZmpB-specific

reactivity could also be seen in supernatants of TIGR4, revealing a single protein band of similar size to that of D39 (data not shown). Since the anti-ZmpB serum was generated against a protein fragment derived from D39, the absence of reactivity in TIGR4 is consistent with the low identity of the two allelic forms of ZmpB (below 50 %). ZmpC was detected only in TIGR4, as expected, since D39 is one of the strains which does not contain the *zmpC* gene.

Since IgA protease activity is known to be a trait common to all pneumococci, and since no substrate is known for ZmpB and ZmpD, we analysed the activity of the ZmpC proteinase. Zymography revealing cleavage of human MMP-9 by ZmpC was performed using the supernatants of strains identified by our locus-specific PCR as being either *zmpC* positive or negative. As shown in Fig. 3, PCR results for *zmpC* are in complete agreement with the presence or absence of enzymic activity.

Detection of the zinc metalloproteinase loci

As our PCR approach for *zmp* detection is based on conserved nucleotide sequences flanking the *zmp* genes (locus-specific PCR), we had to exclude that the genes were carried in alternative positions on the genome. To confirm the absence of *zmp* homologues in alternative positions, we used BLAST analysis of eight complete and nearly complete genome sequences, as an 'up-to-date' alternative to Southern blots. The BLASTN query with the first 120 nucleotides of *zmpD* yielded positive results in strains G54 (type 19F; htgs NCBI BLAST database) and Spanish 23F-1 (type 23F; Sanger Institute), and no hit in strains TIGR4, R6, 670 (serotype 6B; TIGR), INV104B (type 1; Sanger Institute), INV200 (type 14; Sanger Institute) and OXC141 (type 3; Sanger Institute). In both strains positive for *zmpD*, the gene was located downstream from *iga*, excluding any alternative location for *zmpD*, at least in this set of strains. Also, for *iga*, *zmpB* and *zmpC*, genome screens confirmed a single possible location for each of the genes.

All pneumococcal strains yielded amplification products for the three loci examined, confirming that high sequence conservation exists between the pneumococcal genomes. No large variation in the apparent size was observed between the amplicons obtained for the isolates under investigation and the amplicons obtained with the control strains. The few variations consisted of an increase in the apparent size of *zmpB* amplicons: 14 isolates yielded a *zmpB* amplicon of approximately 6500–7500 bp, instead of the 5800 bp of the control. In six isolates, an insertion sequence, IS1167, located downstream from the coding sequence of ZmpB, was found to account for the larger amplicons. No further investigation was carried out of the other isolates.

The analysis of the three metalloproteinase loci in 218 pneumococcal isolates yielded four different combinations, as described in Table 2. The zinc metalloproteinase gene *zmpB* and *iga* were present in all the strains analysed, while *zmpC* was found in 40 isolates only (18.4 %) and *zmpD*

Fig. 2. Western blot for detection of pneumococcal zinc metalloproteinases. The equivalent of 100 µl of culture supernatant at late-exponential phase was loaded to each well (6% SDS-PAGE). Supernatants were from the type 2 strain D39 (lanes 1, 3 and 5) and from the type 4 strain TIGR4 (lanes 2, 4 and 6). Lanes 1 and 2 were developed with anti-IgA-protease serum, lanes 3 and 4 were developed with anti-ZmpB serum, and lanes 5 and 6 with anti-ZmpC serum. When overstaining the blot, ZmpB-specific reactivity could also be seen in lane 4 (not shown).

Fig. 3. Detection of ZmpC activity in pneumococcal culture supernatants. Indirect detection of ZmpC activity was performed by gelatin zymogram, which revealed the gelatinase activity of the ZmpC substrate MMP-9. Samples run on the SDS-PAGE contained purified human MMP-9 and pneumococcal supernatants. The single major gelatin cleavage band determined by intact MMP-9 (96 kDa) is shown in control samples in lane 7 of (A) and lane 9 of (B). The positive-control supernatant of the TIGR4 strain is shown in lane 6 of (A) and lane 8 of (B). In this case, ZmpC cleavage of MMP-9 is evidenced by a second band of gelatin cleavage of slightly lower molecular mass. Pneumococcal supernatants shown are: (A) lane 1, strain SP131 type 19F *zmpC*⁺; lane 2, SP285 type 19F *zmpC*⁻; lane 3, SP277 type 7F *zmpC*⁻; lane 4, SP068 type 4 *zmpC*⁻; (B) lane 1, SP135 type 11A *zmpC*⁺; lane 2, SP082 type 4 *zmpC*⁺; lane 3, SP225 type 8 *zmpC*⁺; lane 4, SP165 type 14 *zmpC*⁻; lane 5, SP239 type 8 *zmpC*⁺; lane 6, SP028 type 3 *zmpC*⁻. Molecular mass markers are shown in lanes (A) 5 and (B) 7.

in approximately half of the isolates examined (48.6%). Interestingly, *zmpC* was more frequently found in isolates carrying *zmpD* also, and therefore positive for all four metalloproteinase genes (Table 2).

Association of the zinc metalloproteinase genes with the source of the isolates and with serotypes

As *zmpB* and *iga* were present in all the isolates examined, subsequent comparisons were carried out taking into account the presence of *zmpC* and *zmpD* only. No association was observed between the presence of *zmpC* or *zmpD* and the type of infection (whether the isolates were from meningitis, bacteraemic pneumonia or sepsis cases) or the age of the patients. Similarly, no difference was found

between isolates from cases of disease and those from healthy carriers. However, a relationship was observed between the presence of *zmpC* and/or *zmpD* and the st of the isolates. *zmpC* was found in all isolates belonging to st 8 and st 11A, and only in rare isolates belonging to other sts. *zmpD* was present in all strains belonging to sts 7F, 8, 9N, 11A and 12F, in the vast majority of st 14, st 19A and st 9V strains, and sporadically in other sts (Table 3).

To confirm the association of *zmpC* and *zmpD* and certain sts, additional isolates belonging to st 8 and st 11A from various sources were examined. Overall, 13 out of 16 st 8 isolates and 14 out of 16 st 11A isolates were positive for both *zmpC* and *zmpD*. Only one st 11A isolate carried *zmpC* alone. The remaining isolates were negative both for *zmpC* and *zmpD*.

Association of the metalloprotease genes with particular clones

In order to verify the nature of the association of *zmpC* and *zmpD* with certain sts, all isolates belonging to st 8 and st 11A were examined by PFGE to determine their genetic relatedness. Isolates with > 80 % genetic relatedness on the PFGE-based dendrogram (Fig. 4) were considered to belong to the same clone. Twenty-eight out of 32 isolates belonging to st 8 and st 11A showed PFGE type 1, which included nine different subtypes (1.1–1.9). All these isolates, with the exception of a single strain (AP200), were both *zmpC* and

Table 2. Different combinations of metalloproteinase genes at the three loci in 218 *S. pneumoniae* isolates

Number of strains (percentage of total)	<i>zmpB</i>	<i>zmpC</i>	<i>iga</i> locus	
			<i>iga</i>	<i>zmpD</i>
106 (48.6 %)	+	–	+	–
72 (33.0 %)	+	–	+	+
34 (15.6 %)	+	+	+	+
6 (2.8 %)	+	+	+	–

Table 3. Correlations between serotypes and the presence of *zmpC* and *zmpD* genes

Serotype	Number of strains	Number of <i>zmpC</i> -positive strains	Number of <i>zmpD</i> -positive strains
1	7	0	0
3	12	0	0
4	10	2	0
6A	13	0	0
6B	14	1	1
7F	12	0	12
8	16	13	13
9N	8	0	8
9V	6	0	4
11A	16	15	14
12F	6	1	6
14	27	1	25
18C	5	2	2
19A	9	1	8
19F	12	2	4
23F	22	0	5
Others	23	2	4
Total	218	40	106

zmpD positive. Divergent PFGE profiles were shown by three st 8 isolates (PFGE type 2) and one st 11A isolate

(PFGE type 3) that were negative for both *zmpC* and *zmpD* (Fig. 4).

Based on PFGE results, five st 8 isolates and five st 11A isolates were selected for MLST analysis. The results are summarized in Table 4. Two sequence types (STs) were found in PFGE type 1 isolates positive for both *zmpC* and *zmpD*: ST53 was found in three st 8 isolates and ST62 in three st 11A isolates. In the MLST database, ST53 contains only st 8 strains (originating from several different countries) and ST62 only st 11A isolates. In addition, ST53 and ST62 differ by two alleles only, at the *gki* and *recP* loci. The single st 11A isolate carrying *zmpC* but not *zmpD* showed a novel combination of alleles which was a single-locus variant of ST62. The *zmpC*- and *zmpD*-negative isolates belonged to unrelated STs: two st 8 isolates belonged, respectively, to ST404 and to the single-locus variant ST1480, both already found in st 8 isolates according to the MLST database; one st 11A strain, PFGE type 3, belonged to ST383, which is unrelated to all the other STs obtained, and has previously been found only in an st 19A isolate.

In order to investigate the possibility that *zmpC*- and *zmpD*-positive strains belonging to sts other than st 8 and st 11A (Table 3) had undergone capsule switch and were derivatives of st 8 or st 11A strains, PFGE analysis was performed on six *zmpC*- and *zmpD*-positive isolates belonging to sts 4, 12F, 14, 18C, 19A and 19F, and on one st 6B isolate positive

Fig. 4. Genetic relatedness of the 16 st 8 and 16 st 11A isolates examined in this study. Comparison of the PFGE banding patterns was performed by the unweighted pair group method with arithmetic averages (UPGAMA) by using the Dice similarity coefficient. The strain codes (from left to right) for representative isolates, PFGE subtypes, numbers of isolates with identical PFGE subtype (in parentheses), st and ST for selected isolates are shown to the right of the figure. The presence of *zmpC* (■) and *zmpD* (●) is indicated for positive strains.

Table 4. Correlation between presence of *zmpC* and *zmpD*, PFGE type and allelic profile (MLST) for selected pneumococcal isolates belonging to serotypes 8 and 11A

In each group, variant alleles are in bold type.

Strain	Serotype	<i>zmpC</i>	<i>zmpD</i>	PFGE-type subtype	ST	Allele no.						
						<i>aroE</i>	<i>gdh</i>	<i>gki</i>	<i>recP</i>	<i>spi</i>	<i>xpt</i>	<i>ddl</i>
SP225	8	+	+	1·1	53	2	5	1	11	16	3	14
SP320	8	+	+	1·2	53	2	5	1	11	16	3	14
PT176	8	+	+	1·2	53	2	5	1	11	16	3	14
PN175	11A	+	+	1·3	62	2	5	29	12	16	3	14
AP156	11A	+	+	1·4	62	2	5	29	12	16	3	14
AP/PT136	11A	+	+	1·4	62	2	5	29	12	16	3	14
AP200	11A	+	—	1·5	ST _{NEW}	2	5	29	10	16	3	14
AP/PT085	8	—	—	2·1	1480	7	9	15	11	93	1	70
AP118	8	—	—	2·3	404	7	9	15	11	42	1	70
PT013	11A	—	—	3	383	1	10	62	10	15	1	6

for *zmpC* only. Each isolate exhibited a unique PFGE pattern, different from those shown by st 8 or st 11A isolates (data not shown), suggesting that capsular switch did not occur among these isolates.

DISCUSSION

The availability of complete and partially sequenced genomes of *S. pneumoniae* has allowed the identification of several new putative virulence factors, including a family of large surface-located zinc metalloproteinases. This family comprises the well-known streptococcal IgA protease (Kilian *et al.*, 1980) and three other metalloproteinases, named ZmpB, ZmpC and (newly named) ZmpD (Bergé *et al.*, 2001; Oggioni *et al.*, 2003). The genes for the IgA protease and ZmpD are adjacent in the same *iga* locus in the G54 genome, and also in the partially sequenced type 23F genome (http://www.sanger.ac.uk/Projects/S_pneumoniae/). On the basis of a variety of studies that demonstrate the role of these proteins in the pathogenicity of *S. pneumoniae* (Blue *et al.*, 2003; Chiavolini *et al.*, 2003; Hava & Camilli, 2002; Oggioni *et al.*, 2003; Polissi *et al.*, 1998; Weiser *et al.*, 2003), we examined the presence of the zinc metalloproteinase genes in a large collection of *S. pneumoniae* isolates both from invasive infections and from healthy carriers.

The genes for the zinc metalloproteinases in *S. pneumoniae* have been recognized to fall within a streptococcal gene family with a high allelic variability (Chiavolini *et al.*, 2003; Oggioni *et al.*, 2003; Poulsen *et al.*, 1998). The extremely high nucleotide variability, resulting in a mosaic structure of the *zmp* genes, hinders epidemiological approaches using nucleotide probes or primers directed to the coding sequences (Hakenbeck *et al.*, 2001). We therefore decided to trace the *zmp* genes by employing primers for conserved sequences outside the genes. In order to link this locus-specific PCR approach to the proteinase genes, we performed BLAST genome screens, gene expression analysis, the determination of protein production, and enzymic activity

assays. *In silico* genome screens confirmed that the four *zmp* genes are always associated to the respective chromosomal loci. Gene expression analysis showed that the three *zmp* genes of strain TIGR4 are expressed in comparable amounts during growth. These *in vitro* results are in accordance with data from whole-genome screens for virulence determinants *in vivo* that have identified *zmp* genes by signature-tagged mutagenesis (Hava & Camilli, 2002; Polissi *et al.*, 1998). In order to control for whether *zmp* gene transcription yields mature proteins, we performed Western blots for IgA1 protease, ZmpB and ZmpC, and activity assays for ZmpC. Both types of assays confirmed that mature and active proteases are produced, and that the phenotypes correlate with the data obtained by our locus-PCR approach, although the variability of Zmp proteins and the specificity of antisera limit the possibility of a large screening by Western blotting. Our data add to a series of reports on the pneumococcal IgA protease which show that all pneumococci produce IgA1 protease and have IgA1 protease activity (Poulsen *et al.*, 1996), and that patients' sera recognize IgA1 protease (Romanello *et al.*, 2006). To summarize, data obtained by others and by us indicate that the pneumococcal *zmp* genes are (1) always linked to a certain chromosomal locus and are (2) transcribed *in vitro* and *in vivo*, and that (3) they yield mature and active proteins, which are (4) antigenic in humans and (5) linked to virulence. These observations prompted us to use our locus-specific PCR approach to evaluate the molecular epidemiology of these genes in a large number of pneumococcal isolates.

As a consequence of primer design based on conserved sequences outside the *zmp* genes, we were able to amplify the three metalloproteinase loci in all the isolates examined by using primers designed towards conserved sequences from the published pneumococcal genomes. With few exceptions (involving *zmpB*), we found that the size of the metalloproteinase genes did not appear to differ from that of the control strains. In some isolates, in which the *zmpB* amplicon was larger than expected, an insertion sequence (IS),

IS1167, was found inserted immediately downstream from the *zmpB* coding sequence. This IS element is commonly found in multiple copies in the pneumococcal genome, and has also been found to cluster, in some isolates, downstream from the *pspC* locus (Iannelli *et al.*, 2002).

Our results confirmed that, as in the strains for which the genome has been published, the genes for two of the metalloproteinases, IgA protease and ZmpB, were consistently present in all the strains examined. Both IgA protease and ZmpB have been demonstrated to play a role in virulence. The IgA1 protease is a well-known virulence factor that cleaves human IgA1 in the hinge region (Kilian *et al.*, 1996; Poulsen *et al.*, 1996; Wani *et al.*, 1996). Previous studies have shown that all pneumococcal isolates exhibit this activity (Kilian *et al.*, 1996; Oggioni *et al.*, 2003). The substrate of ZmpB has not been identified; however, the protein has been demonstrated to be a mediator of inflammation in the lung (Bergé *et al.*, 2001; Blue *et al.*, 2003).

The other two zinc metalloproteinase genes, *zmpC* and *zmpD*, were found in some of the strains only: 18 and 49% of the isolates, respectively. *zmpC* was more frequently present in isolates that also harboured *zmpD*, and were therefore endowed with the complete array of the four metalloproteinase genes. ZmpC has been demonstrated to cleave MMP9 (Oggioni *et al.*, 2003). Cleavage of MMP9 by isolates carrying *zmpC* has been shown previously, and has also been confirmed in 10 randomly selected *zmpC*-positive isolates from this study. To date, no target has been found for ZmpD. Although a previous study indicated an association between the presence of *zmpC* and strains isolated from cases of pneumonia (Oggioni *et al.*, 2003), in this study, no link was observed between the presence of *zmpC* or *zmpD* and the site of pneumococcal infection. In addition, the occurrence of *zmpC* or *zmpD* did not differ in isolates from healthy carriers, confirming that carriage and invasive isolates do not have peculiar characteristics, since invasive isolates originate from carriage ones (Bogaert *et al.* 2004). However, the presence of *zmpC* and *zmpD* did not appear to be random, but was associated with certain serotypes. Previously, among a small number of clinical isolates, *zmpC* has been found associated with serogroup 9 strains (Oggioni *et al.*, 2003). In this study, *zmpC* and *zmpD* were found mainly associated with isolates belonging to st 8 and st 11A. Molecular typing by PFGE revealed that isolates within st 8 and st 11A that carried the whole array of the four metalloproteinase genes were clonally related. MLST analysis confirmed this finding. The *zmpC/zmpD*-positive st 8 and st 11A isolates belonged to STs (ST53 and ST62) that were specific for the respective sts, according to the MLST database and to recently published reports from Scotland (Jefferies *et al.*, 2004) and the USA (Gertz *et al.*, 2003). Interestingly, ST53 and ST62 differ for two alleles out of seven (<http://spneumoniae.mlst.net/>) and therefore they can be considered to enclose clonally related isolates. On the basis of genotyping analysis, the presence of *zmpC/zmpD* in isolates

of sts other than st 8 or st 11A cannot be explained by the occurrence of a capsular switch. In fact, the presence of *zmpC/zmpD* in unrelated isolates belonging to different sts could be due to the horizontal transfer of metalloproteinase genes.

Although only pneumococci isolated in Italy were examined in this study, the ST data for st 8 and st 11A strains from other countries suggest that the clonal cluster 8/11A is not a peculiarity of our geographical area. Sts 8 and 11A are not the most prominent sts causing invasive disease; however, they are fairly common and are incorporated in the 23-valent pneumococcal vaccine. These sts are more commonly found in adults than in young children (Hausdorff *et al.*, 2000). In an Italian study of invasive disease, st 8 and st 11A ranked, respectively, twelfth and seventeenth in the general population and sixth and fourteenth in elderly subjects (Pantosti *et al.*, 2003). It could be speculated that the presence of the whole array of the four metalloproteinases, which are putative virulence factors, has contributed to the stability and success of the clonal cluster 8/11A among clinical isolates.

In conclusion, our study shows the molecular epidemiology of the pneumococcal zinc metalloproteinases, an important family of surface-exposed antigenic virulence determinants. These data are of prime importance when studying the virulence of pneumococcal isolates or screening genomes for novel vaccine candidates, as only extensive molecular epidemiology of variable genes, which are the part of the genome under selective pressure, permits their products to be linked to specific disease situations and their usefulness as vaccine antigens to be predicted.

ACKNOWLEDGEMENTS

The work was supported by the Commission of the European Union (contract LSHM-CT-2005-512099) and FIRB (RBAU01X9TB) (to M.R.O.), by the Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR, Italian Ministry of Research) (Cofinanziamento 2004) (to G.P.), and by funds from the Italian Ministero della Salute (Progetti Finalizzati 2003) to A.P. We would like to thank Mogens Kilian, Gavin Paterson and Tim Mitchell for kindly providing antisera. We would also like to thank Guido Memmi for primer design, Rita Privitera for the anti-ZmpC serum, Claudia Trappetti for data on *zmp* expression and Fabio D'Ambrosio for the PFGE dendrogram.

REFERENCES

- Bergé, M., Garcia, P., Iannelli, F., Prère, M. F., Granadel, C., Polissi, A. & Claverys, J. P. (2001). The puzzle of *zmpB* and extensive chain formation, autolysis defect and non-translocation. *Mol Microbiol* **39**, 1651–1660.
- Blue, C. E., Paterson, G. K., Kerr, A., Bergé, M., Claverys, J. P. & Mitchell, T. J. (2003). ZmpB, a novel virulence factor of *Streptococcus pneumoniae* that induces tumor necrosis factor alpha production in the respiratory tract. *Infect Immun* **71**, 4925–4935.
- Bogaert, D., de Groot, R. & Hermans, P. W. M. (2004). *Streptococcus pneumoniae* colonization: the key to pneumococcal disease. *Lancet Infect Dis* **4**, 144–154.

- Chapuy-Regaud, S., Ogunniyi, A. D., Diallo, N., Huet, Y., Desnottes, J. F., Paton, J. C., Escaich, S. & Trombe, M. C. (2003). RegR, a global LacI/GalR family regulator, modulates virulence and competence in *Streptococcus pneumoniae*. *Infect Immun* **71**, 2615–2625.
- Chiavolini, D., Memmi, G., Maggi, T., Iannelli, F., Pozzi, G. & Oggioni, M. R. (2003). The three extra-cellular zinc metalloproteinases of *Streptococcus pneumoniae* have a different impact on virulence in mice. *BMC Microbiol* **3**, 14.
- Dicuonzo, G., Gherardi, G., Gertz, R. E., D'Ambrosio, F., Goglio, A., Lorino, G., Recchia, S., Pantosti, A. & Beall, B. (2002). Genotypes of invasive pneumococcal isolates recently recovered from Italian patients. *J Clin Microbiol* **40**, 3660–3665.
- Dopazo, J., Mendoza, A., Herrero, J. & 13 other authors (2001). Annotated draft genomic sequence from *Streptococcus pneumoniae* type 19F clinical isolate. *Microb Drug Resist* **7**, 99–125.
- Enright, M. & Spratt, B. (1998). A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* **144**, 3049–3060.
- Gertz, R. E., McEllistrem, C., Boxrud, D. J. & 8 other authors (2003). Clonal distribution of invasive pneumococcal isolates from children and selected adults in the United States prior to 7-valent conjugate vaccine introduction. *J Clin Microbiol* **41**, 4194–4216.
- Gillespie, S. H. (1989). Aspects of pneumococcal infection including bacterial virulence, host response and vaccination. *J Med Microbiol* **28**, 237–248.
- Hakenbeck, R., Balmelle, N., Weber, B., Gardes, C., Keck, W. & de Saizieu, A. (2001). Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. *Infect Immun* **69**, 2477–2486.
- Hausdorff, W. P., Bryant, J., Paradiso, P. R. & Siber, G. R. (2000). Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* **30**, 100–121.
- Hava, D. L. & Camilli, A. (2002). Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. *Mol Microbiol* **45**, 1389–1406.
- Hoskins, J., Alborn, W. E., Arnold, J. & 36 other authors (2001). Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* **183**, 5709–5717.
- Iannelli, F., Oggioni, M. R. & Pozzi, G. (2002). Allelic variation in the highly polymorphic locus *pspC* of *Streptococcus pneumoniae*. *Gene* **284**, 63–71.
- Iannelli, F., Chiavolini, D., Ricci, S., Oggioni, M. R. & Pozzi, G. (2004). Pneumococcal surface protein C (PspC) contributes to sepsis caused by *Streptococcus pneumoniae*. *Infect Immun* **72**, 3077–3080.
- Jefferies, J. M. C., Smith, A., Clarke, S. C., Dowson, C. & Mitchell, T. J. (2004). Genetic analysis of diverse disease-causing pneumococci indicates high level of diversity within serotypes and capsule switching. *J Clin Microbiol* **42**, 5681–5688.
- Kilian, M., Mestecky, J., Kulhavy, R., Tomana, M. & Butler, W. T. (1980). IgA1 proteases from *Haemophilus influenzae*, *Streptococcus pneumoniae*, *Neisseria meningitidis* and *Streptococcus sanguis*: comparative immunochemical studies. *J Immunol* **124**, 2596–2600.
- Kilian, M., Reinholdt, J., Lomholt, H., Poulsen, K. & Frandsen, E. V. G. (1996). Biological significance of IgA1 proteases in bacterial colonization and pathogenesis: critical evaluation of experimental evidence. *APMIS* **104**, 321–338.
- Livak, K. J. & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods* **25**, 402–408.
- McCawley, L. J. & Matrisian, L. M. (2000). Matrix metalloproteinases: multifunctional contributors to tumor progression. *Mol Med Today* **6**, 149–156.
- Mitchell, T. J. (2000). Virulence factors and the pathogenesis of disease caused by *Streptococcus pneumoniae*. *Res Microbiol* **151**, 413–419.
- Oggioni, M. R. & Pozzi, G. (2001). Comparative genomics for identification of clone-specific sequence blocks in *Streptococcus pneumoniae*. *FEMS Microbiol Rev* **200**, 137–143.
- Oggioni, M. R., Memmi, G., Maggi, T., Chiavolini, D., Iannelli, F. & Pozzi, G. (2003). Pneumococcal zinc metalloproteinase ZmpC cleaves human matrix metalloproteinase 9 and is a virulence factor in experimental pneumonia. *Mol Microbiol* **49**, 795–805.
- Oggioni, M. R., Iannelli, F., Ricci, S., Chiavolini, D., Parigi, R., Trappetti, C., Claverys, J.-P. & Pozzi, G. (2004). Antibacterial activity of a competence stimulating peptide in experimental sepsis caused by *Streptococcus pneumoniae*. *Antimicrob Agents Chemother* **48**, 4725–4732.
- Pantosti, A., D'Ambrosio, F., Tarasi, A., Recchia, S., Orefici, G. & Mastrantonio, P. (2000). Antibiotic susceptibility and serotype distribution of *Streptococcus pneumoniae* causing meningitis in Italy, 1997–1999. *Clin Infect Dis* **31**, 1373–1379.
- Pantosti, A., Boccia, D., D'Ambrosio, F., Recchia, S., Orefici, G. & Moro, M. L. (2003). Inferring the potential success of pneumococcal vaccination in Italy: serotypes and antibiotic resistance of *Streptococcus pneumoniae* isolates from invasive diseases. *Microb Drug Resist* **9**, 61–68.
- Paton, J. C., Andrew, P. W., Boulnois, G. J. & Mitchell, T. J. (1993). Molecular analysis of the pathogenicity of *Streptococcus pneumoniae*: the role of pneumococcal proteins. *Annu Rev Microbiol* **47**, 89–115.
- Petrosillo, N., Pantosti, A., Bordin, E., Spanò, A., Del Grosso, M., Tallarida, B. & Ippolito, G. (2002). Prevalence, determinants, and molecular epidemiology of *Streptococcus pneumoniae* isolates colonizing the nasopharynx of healthy children in Rome. *Eur J Clin Microbiol Infect Dis* **21**, 181–188.
- Polissi, A., Pontiggia, A., Feger, G., Altieri, M., Mottl, H., Ferrari, L. & Simon, D. (1998). Large-scale identification of virulence genes from *Streptococcus pneumoniae*. *Infect Immun* **66**, 5620–5629.
- Poulsen, K., Reinholdt, J. & Kilian, M. (1996). Characterization of the *Streptococcus pneumoniae* immunoglobulin A1 protease gene (*iga*) and its translation product. *Infect Immun* **64**, 3957–3966.
- Poulsen, K., Reinholdt, J., Jespersgaard, C., Boye, K., Brown, T. A., Hauge, M. & Kilian, M. (1998). A comprehensive genetic study of streptococcal immunoglobulin A1 proteases: evidence for recombination within and between species. *Infect Immun* **66**, 181–190.
- Romanello, V., Marcacci, M., Molin, F. D., Maschioni, M., Censini, S., Covacci, A., Baritussio, A. G., Montecucco, C. & Tonello, F. (2006). Cloning, expression, purification, and characterization of *Streptococcus pneumoniae* IgA1 protease. *Protein Expr Purif* **45**, 142–149.
- Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. & Swaminathan, B. (1995). Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33**, 2233–2239.
- Tettelin, H. & Hollingshead, S. K. (2004). Comparative genomics of *S. pneumoniae*: intra-strain diversity and genome plasticity. In *The Pneumococcus*, pp 15–29. Edited by E. I. Tuomanen. Washington, DC: American Society for Microbiology.
- Tettelin, H., Nelson, K. E., Paulsen, I. T. & 36 other authors (2001). Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* **293**, 498–506.
- Wani, J. H., Gilbert, J. V., Plaut, A. G. & Weiser, J. N. (1996). Identification, cloning, and sequencing of the immunoglobulin A1 protease gene of *Streptococcus pneumoniae*. *Infect Immun* **64**, 3967–3974.
- Weiser, J. N., Bae, D., Fasching, C., Scamurra, R. W., Ratner, A. J. & Janoff, E. N. (2003). Antibody-enhanced pneumococcal adherence requires IgA1 protease. *Proc Natl Acad Sci U S A* **100**, 4215–4220.