Variation in a surface-exposed region of the Mycoplasma pneumoniae P40 protein as a consequence of homologous DNA recombination between RepMP5 elements

Emiel B. M. Spuesens, Nick van de Kreeke, Silvia Estevão, Theo Hoogenboezem, Marcel Sluijter, Nico G. Hartwig, Annemarie M. C. van Rossum and Cornelis Vink

Erasmus MC-Sophia Children’s Hospital, Laboratory of Pediatrics, Pediatric Infectious Diseases and Immunity, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

Mycoplasma pneumoniae is a human pathogen that causes a range of respiratory tract infections. The first step in infection is adherence of the bacteria to the respiratory epithelium. This step is mediated by a specialized organelle, which contains several proteins (cytadhesins) that have an important function in adherence. Two of these cytadhesins, P40 and P90, represent the proteolytic products from a single 130 kDa protein precursor, which is encoded by the MPN142 gene. Interestingly, MPN142 contains a repetitive DNA element, termed RepMP5, of which homologues are found at seven other loci within the M. pneumoniae genome. It has been hypothesized that these RepMP5 elements, which are similar but not identical in sequence, recombine with their counterpart within MPN142 and thereby provide a source of sequence variation for this gene. As this variation may give rise to amino acid changes within P40 and P90, the recombination between RepMP5 elements may constitute the basis of antigenic variation and, possibly, immune evasion by M. pneumoniae. To investigate the sequence variation of MPN142 in relation to inter-RepMP5 recombination, we determined the sequences of all RepMP5 elements in a collection of 25 strains. The results indicate that: (i) inter-RepMP5 recombination events have occurred in seven of the strains, and (ii) putative RepMP5 recombination events involving MPN142 have induced amino acid changes in a surface-exposed part of the P40 protein in two of the strains. We conclude that recombination between RepMP5 elements is a common phenomenon that may lead to sequence variation of MPN142-encoded proteins.

INTRODUCTION

Mycoplasma pneumoniae is a common pathogen of the human respiratory tract that can cause both upper and lower respiratory tract infections, in particular in childhood and adolescence (Waites & Talkington, 2004). The first and crucial step in colonization and infection with M. pneumoniae is attachment of the bacterium to the epithelium of the respiratory tract. An essential role in this process is played by a specialized attachment organelle at the tip of M. pneumoniae cells. This attachment organelle (also referred to as the ‘terminal tip’) is composed of a complex of adherence and accessory proteins. The loss of function of one of the proteins from this complex almost invariably results in a non-functional attachment organelle and, hence, a non-virulent phenotype of the bacteria (Krause et al., 1982). One of the most prominent adherence proteins in the attachment organelle is the ~170 kDa P1 protein. This protein forms a complex with two accessory proteins, called P40 and P90 (Layh-Schmitt & Herrmann, 1994), which function at the distal tip of the attachment organelle and are required for the correct localization of P1 (Layh-Schmitt & Herrmann, 1992; Layh-Schmitt & Harkenthal, 1999). In line with this close functional relationship, it has been shown that the genes encoding the P1, P40 and P90 proteins are linked at both the transcriptional and translational level and are located in a single operon (Waldo & Krause, 2006). In this operon, the MPN141 gene encodes the P1 protein, whereas the MPN142 gene encodes a 130 kDa precursor protein which is cleaved into an N-terminal protein of ~40 kDa (P40).
Apart from the functional linkage of MPN141 and MPN142, these genes also share an important structural feature. Both genes contain stretches of DNA of which homologous sequences exist at other loci within the *M. pneumoniae* genome. These stretches of DNA and their homologous counterparts are known as RepMP elements (or repetitive DNA elements). The MPN141 gene contains a RepMP4 element and a RepMP2/3 element, of which, respectively, seven and nine versions are present at other loci in the genome. MPN142 contains a RepMP5 element of which seven other versions are present throughout the genome of *M. pneumoniae* strain M129 (Dandekar et al., 2000; Himmelreich et al., 1996) (Fig. 1). These elements range in size from 1825 to 3361 bp (Table 1), and share 48–91% identity among each other (Supplementary Table S1). It has been hypothesized that these elements could rearrange with their homologous counterparts by means of homologous DNA recombination. Obviously, such rearrangements could generate sequence variation within the MPN141 and MPN142 genes, which in turn could give rise to amino acid changes in the proteins they encode. Although no rearrangements have been reported (Rocha & Blanchard, 2002; Ruland et al., 1990), it has been hypothesized that these elements could rearrange with their homologous counterparts by means of homologous DNA recombination. Obviously, such rearrangements could generate sequence variation within the MPN141 and MPN142 genes, which in turn could give rise to amino acid changes in the proteins they encode.

![Diagram of RepMP elements](image)

### Table 1. Primer sequences and genomic positions of the RepMP5 elements within strains M129 and FH

<table>
<thead>
<tr>
<th>Element</th>
<th>Genome position*</th>
<th>Primer 1 (5'-3')</th>
<th>Primer 2 (5'-3')</th>
<th>Amplicon length (bp)</th>
<th>Element length (bp)†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Primer 1</td>
<td>Primer 2</td>
<td>M129</td>
<td>FH</td>
</tr>
<tr>
<td>RepMP5-a</td>
<td>113 676–116 092</td>
<td>tga tgg cat taa tta cta ttg c</td>
<td>cta cta ata gta aag tgg agt</td>
<td>2417</td>
<td>2415</td>
</tr>
<tr>
<td>RepMP5-b</td>
<td>131 974–133 863</td>
<td>gat tcc aat acc acc act gat</td>
<td>ttc aaa agt gaa aac cct act</td>
<td>1891</td>
<td>1891</td>
</tr>
<tr>
<td>RepMP5-c</td>
<td>185 648–188 602</td>
<td>cag tgc ggc taa acc agg a</td>
<td>tgg gag gtc gtc ctg act g</td>
<td>2955</td>
<td>2754</td>
</tr>
<tr>
<td>RepMP5-d</td>
<td>196 236–199 306</td>
<td>aca cat cag gtt act cat cag</td>
<td>gaa cac aat tgg aac tag ctg</td>
<td>3071</td>
<td>3071</td>
</tr>
<tr>
<td>RepMP5-e</td>
<td>435 836–439 626</td>
<td>tga gag gtt taa tgc ctc tca</td>
<td>tgg cct tgg ctt tga aag gt</td>
<td>3791</td>
<td>3792</td>
</tr>
<tr>
<td>RepMP5-f</td>
<td>497 466–499 820</td>
<td>ggt gtt gtc gat tgg ttt gtc</td>
<td>ttc ttt tta aca cta cta ggc aa</td>
<td>2373</td>
<td>2373</td>
</tr>
<tr>
<td>RepMP5-g</td>
<td>566 089–568 240</td>
<td>gtt tat gga tta aat acc tca c</td>
<td>ctc tac atc aaa cta acc gct</td>
<td>2168</td>
<td>2168</td>
</tr>
<tr>
<td>RepMP5-h</td>
<td>609 094–612 464</td>
<td>ccc act aaa att gag ttc</td>
<td>att tag tgt gag cag aca g</td>
<td>3371</td>
<td>3371</td>
</tr>
</tbody>
</table>

*Genomic positions are derived from GenBank accession number U00089.2.
†Except for RepMP5-c, the lengths of all RepMP5 elements were redefined with respect to the originally published sequences (in parentheses) from *M. pneumoniae* M129 U00089.2.
1983; Layh-Schmitt & Herrmann, 1992; Leith et al., 1983; Seto et al., 2005; Sperker et al., 1991), it is tempting to speculate that amino acid changes within these proteins could play a role in the immune-evasive strategies of *M. pneumoniae*.

Currently, there is strong evidence that recombination among RepMP2/3 elements and RepMP4 elements indeed occurs in *M. pneumoniae* strains (Kenri et al., 1999; Spuesens et al., 2009), and that these recombination events are characterized by a unidirectional DNA sequence transfer mechanism reminiscent of gene conversion (Dumke et al., 2006; Kenri et al., 1999; Pereyre et al., 2007; Spuesens et al., 2009). Moreover, in several clinical strains, such recombination events have likely led to sequence changes within the MPN141 gene, resulting in amino acid sequence alterations within the P1 protein (Dorigo-Zetsma et al., 2001; Dumke et al., 2006; Kenri et al., 1999; Pereyre et al., 2007; Spuesens et al., 2009). While there is considerable knowledge of the mechanism underlying the sequence variation of the MPN141 gene, not much is known about the dynamics of variation of the MPN142 gene. The potential role of the MPN142-localized RepMP5 element in this process is also still unclear. This is remarkable, because this element, which has a length of ~2.7 kb, spans a significant part of the 5’ end of the MPN142 ORF, and includes the sequence that encodes the proposed cleavage site separating the P40 and P90 proteins (Fig. 3) (Catrein et al., 2005; Layh-Schmitt & Herrmann, 1992; Ruland et al., 1994; Sperker et al., 1991). Nevertheless, homologous recombination between RepMP5 elements has been hypothesized to occur on the basis of two findings: (i) that the sequence of the MPN142-borne RepMP5 element from strain FH differs significantly from that of strain M129; and (ii) that the variant sequence of this element from strain FH is also found at a site distant from MPN142 within the bacterial genome, in both strain FH and strain M129 (Ruland et al., 1994). In this regard, it is important to note that strain FH is a so-called subtype 2 strain, whereas strain M129 is a subtype 1 strain. Subtype 1 and subtype 2 strains represent distinct evolutionary lineages of *M. pneumoniae* that can be distinguished on the basis of DNA sequence differences [mostly single-nucleotide polymorphisms (SNPs), small deletions and insertions] throughout the bacterial genome; these differences between subtype 1 and 2 strains have also been found in all variants of RepMP2/3 and RepMP4 (Spuesens et al., 2009).

### Table 2. Names, genotypes and P1 subtypes of *M. pneumoniae* strains used in this study

<table>
<thead>
<tr>
<th>No.</th>
<th>Strain</th>
<th>Genotype or subtype*</th>
<th>P1 operon subtype†</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>M129 (ATCC 29342)</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>2</td>
<td>Mac (ATCC 15492)</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>3</td>
<td>PI1428 (ATCC 29085)</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>4</td>
<td>Mp5</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>5</td>
<td>Mp22</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c; a/f/c)</td>
</tr>
<tr>
<td>6</td>
<td>Mp72</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c; a/f/c)</td>
</tr>
<tr>
<td>7</td>
<td>Mp1042</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>8</td>
<td>Mp1116</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>9</td>
<td>Mp1286</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>10</td>
<td>Mp1397</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>11</td>
<td>Mp1842</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>12</td>
<td>Mp2004</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>13</td>
<td>Mp2018</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>14</td>
<td>Mp2157</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>15</td>
<td>Mp4599</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>16</td>
<td>Mp4817</td>
<td>1</td>
<td>1-(4-c; 2/3-d; d; 5-c)</td>
</tr>
<tr>
<td>17</td>
<td>Mp5181</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>18</td>
<td>Mp5191</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>19</td>
<td>Mp5192</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>20</td>
<td>Mp5194</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>21</td>
<td>Mp5196</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>22</td>
<td>Mp5245</td>
<td>1</td>
<td>1-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>23</td>
<td>Ofo</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>24</td>
<td>FH (ATCC 15531)</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
<tr>
<td>25</td>
<td>R003</td>
<td>2</td>
<td>2-(4-c; 2/3-d; 5-c)</td>
</tr>
</tbody>
</table>

*The genotype or subtype of the strains was previously determined by Dorigo-Zetsma et al. (2001), Dumke et al. (2003) and Spuesens et al. (2009).
†The P1 operon subtype was defined according to the classification scheme of Spuesens et al. (2009).
To investigate the sequence variation among RepMP5 elements in *M. pneumoniae*, as well as the putative role of these elements in antigenic variation of the surface-exposed P40 and P90 proteins, we have determined the sequences of all eight RepMP5 elements in a collection of 25 *M. pneumoniae* strains. Our data reveal that RepMP5 rearrangements are common among strains and can lead to amino acid changes in MPN142-encoded polypeptides.

**METHODS**

**Strains.** A total of 25 different *M. pneumoniae* strains, including 15 subtype 1 and 10 subtype 2 strains, were used in this study. The names and characteristics of these strains are listed in Table 2. All strains, except strain R003, have been described by Dorigo-Zetsma et al. (2001), and were isolated in Denmark and The Netherlands between 1962 and 1995. Strain R003 was isolated in Germany in 1991 and has been described previously by Dumke et al. (2003) and Maquelin et al. (2009). The complete genome sequence of *M. pneumoniae* strain M129 (ATCC 29342; a subtype 1 strain) was used as a reference (Dandekar et al., 2000; Himmelreich et al., 1996). In addition, we used strain PI 1428 (ATCC 29085) as a representative of subtype 1, and strains Mac (ATCC 15492) and FH (ATCC 15531) as representatives of subtype 2 strains.

**M. pneumoniae cultures and genomic DNA purification.** Strains were grown in 3 ml Mycoplasma medium containing 1.4% Difco PPLO broth (Becton Dickinson), 0.15% Difco TC Yeastolate, UF (Becton Dickinson), 1.4% glucose, 20% horse serum, 1000 U penicillin G ml⁻¹, 500 U polymyxin B ml⁻¹, and 0.02 mg phenol red ml⁻¹, at 37°C/5% CO₂ in 25 cm² tissue culture flasks (Greiner). Cells were harvested upon colour change of the medium (from red/orange to yellow). Genomic DNA was isolated from *M. pneumoniae* cells as described by Spuesens et al. (2009).

**Amplification and cloning of RepMP5 elements.** Eight specific PCR primer sets were designed on the basis of the genome sequence of *M. pneumoniae* strain M129 (GenBank accession no. U00089.2). The primers, which were synthesized and purchased from Eurogentec, are listed in Table 1. For all strains, purified chromosomal DNA was used to amplify each RepMP5 element by high-fidelity PCR. For most elements, the PCR mixture (25 μl) contained 0.4 μM primer 1, 0.4 μM primer 2, 0.2 mM of each dNTP (Fermentas), 0.02 U μl⁻¹ *Pfu* DNA polymerase (Fermentas), 1 μM forward primer (Eurogentec), and 1 μl of chromosomal DNA. The conditions for the long template protocol were as follows: 5 min at 94°C, followed by 25 cycles of 10 s at 94°C, 30 s at 55°C, and 4 min at 68°C, with an increment of 10 s per cycle for the last 15 cycles. All resulting PCR fragments, of which the lengths are shown in Table 1, were purified by ethanol precipitation and cloned into a *HincII*-digested pBluescript SK⁺ vector (Stratagene). In some cases, purified PCR products were used directly in DNA sequencing reactions (described below).

**Sequencing of RepMP5 elements.** Automated DNA sequencing was performed using ABI PRISM BigDye Terminator (BDT) Cycle Sequencing Ready Reaction kits and a 3130xl Genetic Analyzer (Applied Biosystems). Sequencing reactions (10 μl) contained 1 μl BDT Ready Reaction premix (Applied Biosystems), 5 μl sequencing buffer (Applied Biosystems), 3.2 pmol of either the *M13* reverse or forward primer (Eurogentec), and 1 μl plasmid DNA or PCR product (150–300 ng). Since most of the RepMP5 elements are larger than 2 kb, new primers were designed in order to close all sequences double-stranded using a ‘primer-walking’ strategy. Raw DNA sequence data were processed and assembled using the application SeqMan II (DNASTAR). Alignments were made with the online multiple sequence alignment program CLUSTAL W, available at http://www.ebi.ac.uk/Tools/msa/clustalw2/.

**RESULTS**

**Amplification and sequencing of RepMP5 elements**

To determine the nucleotide sequences of the RepMP5 elements within the genomes of the 25 *M. pneumoniae* strains, we designed eight element-specific PCR primer sets (Table 1). The sequences of the primers were derived from the complete genome sequence of strain M129 (Himmelreich et al., 1996), and were selected so as to allow specific amplification of all known RepMP5 elements. Since no systematic nomenclature for the RepMP5 elements was available, the elements were provided with a suffix, i.e. -a to -h, respectively, in a manner similar to that described previously for the RepMP2/3 and RepMP4 elements (Spuesens et al., 2009) (Table 1). This nomenclature is based on the consecutive appearance of each repetitive element within the genome of *M. pneumoniae*.

**Fig. 2.** Sequence variation among the eight RepMP5 elements in 25 *M. pneumoniae* strains. Nucleotide positions are derived from GenBank accession number U00089.2. The sequences of RepMP5-a to -h are schematically represented in (a) to (h), respectively. Only subtype-specific polymorphisms, tandem repeats and putative inter-RepMP5 recombination events are shown. Subtype-independent sequence polymorphisms are listed in Supplementary Table S2. The arrowed boxes indicate the length (to scale) and orientation of each element. Elements RepMP5-b, -f, -g and -h are directly flanked by a RepMP4 element, whereas RepMP5-e is flanked by a RepMP1 element; these flanking elements are indicated in grey. A detailed description of RepMP5-c is shown in Fig. 3. ORFs that were identified within the elements are shown as grey bars (including their ORF numbers). Subtype-specific polymorphisms are indicated below the arrowed boxes in two rows. The top row represents the subtype 1 signature sequence (1) and the bottom row the subtype 2 sequence (2). A ‘Δ’ indicates a deletion. Putative recombination events were found in four elements (RepMP5-c, -d, -g and -h) in seven strains and are represented by black bars accompanied by boxes indicating the putative donor sequence; details of RepMP5-c are given in Fig. 3.
(a) MPN142 (185747 – 189403)

(b) RepMP5-c (185747 – 188447) subtype-dependent DNA sequence differences

(c) RepMP5-c subtype-dependent amino acid sequence differences
During the selection of appropriate PCR primer sequences, we found the DNA sequence homology between some of the RepMP5 elements to extend beyond the boundaries that were initially determined for these elements (Himmelreich et al., 1996). Consequently, the borders of some of the elements (RepMP5-d and -e) had to be redefined, so that these elements turned out to be considerably larger than reported previously (Table 1). In addition, for two of the RepMP5 elements (RepMP5-e and -h), it was not possible to design unique PCR primers targeting the exact borders of the elements, as these elements were joined at one side to another repetitive element (a RepMP1 and a RepMP4 element, respectively). The PCR primers targeting elements RepMP5-e and -h were therefore designed in such a way that RepMP5-e would be amplified jointly with the neighbouring RepMP1 element, and RepMP5-h would be amplified together with the adjacent RepMP4 element.

All eight RepMP5 element-specific primer sets gave rise to PCR products of the expected lengths for all 25 investigated strains (Table 2). Thus, a total of 200 PCR products were produced. Subsequently, the PCR products were cloned into E. coli plasmid pBluescript SK− and sequenced using a 'primer walking' strategy. To minimize the occurrence of sequencing errors due to infidelity of the (proofreading) Pfu DNA polymerase, duplicate sequencing reactions were performed on independently produced PCR fragments in case a unique polymorphism was encountered in a single strain.

**Strain subtype-dependent sequence variation among RepMP5 elements**

The sequences of the RepMP5 elements from all 25 strains are represented schematically in Fig. 2. As previously described for the RepMP2/3 and RepMP4 elements within the M. pneumoniae genome (Spuesens et al., 2009), each RepMP5 element was found to have either a subtype 1-specific or a subtype 2-specific sequence. This subtype differentiation corresponded completely with that derived previously from the P1 gene sequences of these strains (Table 2). This signifies that subtype 1 and subtype 2 strains can be distinguished on the basis of SNPs in all RepMP5 elements, as well as by single-nucleotide deletions (in RepMP5-a; Fig. 2a) and a single insertion (in RepMP5-e; Fig. 2e). The most abundant subtype-specific variation was observed in the RepMP5-c element, which is located in the MPN142 gene (Fig. 3). This variation has also been observed in a study in which the RepMP5-c elements from strains M129 and FH were compared (Ruland et al., 1994). The sequence differences between subtype 1 and subtype 2 strains give rise to significant differences in the predicted amino acid sequences of the MPN142-encoded P40 and P90 proteins (Fig. 3c).

**Subtype-independent sequence differences in RepMP5 elements**

The sequences of the RepMP5 elements also differ among strains in a subtype-independent fashion. The observed differences (or mutations) include nucleotide transitions and transversions, as well as insertions and deletions (Supplementary Table S2). Taken together, 83.4% of the mutations were unique to a given RepMP5 element within a strain, whereas 16.6% of the mutations were shared with the corresponding element from at least one other strain. It is interesting to note that differences were also found between strain M129 sequences deposited in GenBank and those determined in this study in four of the eight RepMP5 elements (RepMP5-e to -h). These differences may originate either from sequencing errors or from the accumulation of mutations within the genome of laboratory strain M129 during prolonged culturing. Nevertheless, repeated amplification and sequencing of elements RepMP5-e to -h from strain M129 did not indicate the occurrence of amplification and/or sequencing errors in this study.

Another interesting observation in two of the eight RepMP5 elements was the presence of tandem repeats of varying lengths. Within element RepMP5-b, the sequence 5′-ATC-3′ was found to be present in tandem repeats of either three (in 19 strains), four (in four strains), five (in one strain) or six (in one strain) (Fig. 2b). Another tandem repeat was identified in element RepMP5-e. While a single copy of this 12 bp repeat (5′-CGCCGAGCTAAG-3′) was
Recombination between RepMP5 elements

Within four of the eight RepMP5 elements (RepMP5-c, -d, -g and -h) from seven of the 25 strains, sequence stretches were identified that differed significantly from those of the corresponding elements in other strains. Conversely, these sequence stretches were found to be identical to sequences found in other RepMP5 elements from the same strain. Specifically, the RepMP5-d element from subtype 2 strain Mp2004 contained a 199 bp fragment that was identical to a sequence found in element RepMP5-h (Fig. 2d). In this strain, as well as in three other strains (the subtype 1 strain Mp4817, and subtype 2 strains Mp4599 and Mp5196), the RepMP5-g element was found to contain a 127 bp sequence identical to a sequence from either RepMP5-f or RepMP5-h (these two elements are identical in this sequence; Fig. 2g). Furthermore, in subtype 2 strain Mac, a 61 bp sequence was identified in RepMP5-h that was identical to a sequence from the RepMP5-e element (Fig. 2h). The most significant finding, however, was the identification of a 45 bp aberrant sequence within the MPN142-localized RepMP5-c element of two subtype 1 strains (Mp22 and Mp72); this sequence was found to be identical to a sequence from either RepMP5-a or RepMP5-f (these two elements are identical in this sequence; Fig. 4). Together, these data strongly support the hypothesis that RepMP5 elements may rearrange within the M. pneumoniae genome by homologous DNA recombination. In addition, as the ‘donor’ DNA elements appeared to have remained unaltered by the presumed recombination events, the results also suggested that the rearrangement of RepMP5 sequences may have occurred in a nonreciprocal, unidirectional fashion, as proposed previously for the RepMP2/3 and RepMP4 elements (Kenri et al., 1999; Spuesens et al., 2009). The putative homologous recombination events that have occurred in the investigated M. pneumoniae strains are summarized schematically in Fig. 5.

P40 amino acid sequence variation induced by RepMP5 rearrangements

As the putative recombination events that involved the RepMP5-c elements of strains Mp22 and Mp72 could influence the protein coding capacity of the MPN142 genes of these strains, we compared the predicted amino acid sequences derived from these genes with those derived from the MPN142 gene of reference strain M129. As shown in Fig. 4, the MPN142-encoded amino acid sequence from strains Mp22 and Mp72 differs from that of strain M129 at three positions within a 16 amino acid stretch, between amino acid positions 169 and 184 of the P40 protein. The three mutations include a Gly to Ala substitution at position 169, a Lys to Thr substitution at position 181, and a Gln to Lys substitution at position 184. Importantly, these amino acid residues are located in a surface-exposed region of the P40 protein (Layh-Schmitt & Herrmann, 1992; Sperker et al., 1991). These findings therefore indicate that the rearrangement of RepMP5 elements can result in the modification of surface-exposed parts of the M. pneumoniae P40 protein, which may influence the recognition of P40 by human antibodies.

DISCUSSION

The characterization of the RepMP5 elements from a set of 25 M. pneumoniae strains has led to the following main conclusions: (i) the genetic difference between subtype 1 and subtype 2 strains of M. pneumoniae can be observed in each RepMP5 element within the bacterial genome; (ii) DNA sequences can be transferred from one RepMP5 element to another, most likely by unidirectional (non-reciprocal) homologous DNA recombination; and (iii) RepMP5 rearrangements can lead to amino acid sequence changes in surface-exposed regions of the P40 protein, which may contribute to antigenic variation of M. pneumoniae. Although the actual occurrence of recombination events between RepMP5 elements was not monitored in ‘real-time’ in this study, the above conclusions are justified because the sequences of the RepMP5...
sequences among subtype 1 strains and among subtype 2 strains are relatively stable.

Strain subtype-specific sequence differences among repetitive DNA elements have also been previously reported for each copy of the RepMP2/3 and RepMP4 elements from *M. pneumoniae* (Spuesens et al., 2009). In most RepMP elements, the sequence differences between subtype 1 and subtype 2 strains are represented by SNPs. However, within the RepMP elements located within the P1 operon (i.e. RepMP4-c and RepMP2/3-d within ORF MPN141, and RepMP5-c within MPN142), the sequence differences between the two subtypes are much more pronounced. Nevertheless, it is clear that the differences between these subtypes can already be observed at the single-nucleotide level, a notion that is useful in the design of molecular typing assays (Spuesens et al., 2010a). The genetic differences between the subtypes are, however, not solely restricted to RepMP elements, as differences are also found in ‘regular’ ORFs, such as MPN528a (Sluijter et al., 2010; Spuesens et al., 2010b).

Apart from the subtype-specific differences between strains in the sequences of their RepMP5 elements, several subtype-independent as well as strain-specific differences were found among these elements. The most prominent subtype-independent difference observed was the occurrence of variable numbers of tandem repeats in two of the elements. The significance of (the variable number of) these repeats is not yet known, in particular because the elements in which they reside have not been shown to be translated into protein. However, it is interesting to note that the part of RepMP5-b that contains the trinucleotide tandem repeat from strain M129 has previously been cloned within the RepMP5-c element such as to mimic a putative recombination event within the MPN142 gene (Catrin et al., 2004). The recombinant *M. pneumoniae* strain carrying this modified MPN142 gene was found to be virulent in an animal model and to express both the P40 and the P90 protein (Catrin et al., 2004).

Although the RepMP5 elements that are located outside the P1 operon all harbour one or more ORFs (Fig. 2), RepMP5-c is the only RepMP5 element that is currently known to be protein-coding. In line with current hypotheses, the other seven RepMP5 elements could serve as a source of variant sequences, which may be transferred to other RepMP5 elements (including RepMP5-c) by homologous recombination. Indeed, we were able to identify a total of four different putative recombination events that had occurred in seven of the 25 investigated strains. One of these events, which was found in strains Mp22 and Mp72, involved the RepMP5-c element, and was found to induce amino acid sequence changes in a region of the P40 protein that has been demonstrated to be expressed extracellularly and to carry antigenic determinants (Layh-Schmitt & Herrmann, 1992). It is not known whether the altered sequence of P40 in Mp22 and Mp72 is actually targeted by the humoral immune response because

---

**Fig. 5.** Schematic representation of the structure of the RepMP5 elements in 25 *M. pneumoniae* strains. This representation is based on the scheme published previously for the RepMP2/3 and RepMP4 elements within the *M. pneumoniae* genome (Spuesens et al., 2009). Each box represents the sequence of a single RepMP5 element. Each row of boxes represents all RepMP5 elements as found in one or more strains (subtype 1 strains in the top panel and subtype 2 strains in the bottom panel). The different patterns within the boxes indicate putative recombination events; each pattern refers to the donor element from which a particular sequence may have been derived. The presence of a box with two different patterns indicates that the sequences may have originated from either of two different donor elements.
putative epitopes within this region have not yet been identified. Regardless, this finding is believed to be the first to demonstrate a modification of the MPN142 sequence, as well as the MPN142-encoded amino acid sequence, by virtue of the exchange of RepMP5 sequences.

Earlier, a recombination event was hypothesized to have occurred in the MPN142 gene of *M. pneumoniae* strain FH (Ruland *et al.*, 1994). It was suggested that the difference between strain M129 (a subtype 1 strain) and strain FH (a subtype 2 strain) in their RepMP5-c sequences was caused by the introduction of sequences from a distal RepMP5 element (the donor element) into the RepMP5-c element of strain FH. The donor element was found to be present within the genomes of both strain FH and strain M129 (Ruland *et al.*, 1994). Indeed, we found this putative donor element, which is RepMP5-h, to be present in all investigated *M. pneumoniae* strains. However, because all subtype 2 strains analysed to date have a RepMP5-c sequence similar to that of strain FH (Fig. 3), this putative recombination event is likely to have taken place early in the evolution of *M. pneumoniae*, concurrently with, or immediately following, the divergence of the two main strain lineages of *M. pneumoniae* (subtype 1 and subtype 2). In contrast, the putative recombination events of RepMP5 elements that were identified in this study are likely to have occurred more recently.

As the primary structure of the MPN142 ORF of a given *M. pneumoniae* isolate may have clinical and/or epidemiological consequences, it may be useful to include information on this structure in the classification system that was previously proposed for *M. pneumoniae* strains (Spuesens *et al.*, 2009). This system contains data on both the subtype of a strain as well as the structure of the MPN141 gene (Spuesens *et al.*, 2009, 2010a, b). On the basis of the current knowledge of the organization of the MPN142 gene, we expanded the *M. pneumoniae* classification system to address the overall structure of all three RepMP elements within the P1 operon. In Table 2 this classification system is applied to all strains from this study.

In conclusion, we have found that putative recombination events between RepMP5 elements can result in amino acid sequence changes in surface-exposed parts of the *M. pneumoniae* P40 protein. Because the RepMP5-c element spans ~74% of the MPN142 ORF and overlaps with sequences encoding both P40 and P90, it is likely that recombination events involving RepMP5-c can also lead to amino acid changes in the P90 protein. We hypothesize that such events, as well as other novel inter-RepMP recombination events, will continue to emerge among clinical isolates of *M. pneumoniae*.

**ACKNOWLEDGEMENTS**

A. M. C. V. R. is supported by grants of the European Society for Paediatric Infectious Diseases, ZonMW and the Erasmus MC.

**REFERENCES**


Edited by: J. Renaudin