Analysis of three different repeated DNA elements present in the P1 operon of *Mycoplasma pneumoniae*: size, number and distribution on the genome

Klaus Ruland, Rainer Wenzel and Richard Herrmann
Department of Microbiology, University of Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, FRG

Received July 12, 1990; Revised and Accepted August 11, 1990

ABSTRACT

*Mycoplasma pneumoniae*, a bacterium pathogenic for humans, has a relatively small genome size of 840 kbp. Even though, several repeated DNA elements have been identified in the genome of this prokaryote, particularly within the P1 gene which codes for a major adhesin protein of *M. pneumoniae*. These elements were characterized in detail with respect to size, number and distribution on the genome, represented by an ordered cloned library covering the complete chromosome. Three different repetitive elements were detected in and around the P1 gene designated as RepMP2/3, RepMP4 and RepMP5. The length of these elements varies between 1.1 to 1.5 kbp (RepMP4), 1.8 kbp (RepMP2/3) and 1.9 to 2.2 kbp (RepMP5). They occur at least 8 to 10 times on the chromosome. Possible functions are discussed and a uniform nomenclature for these repeats is proposed.

INTRODUCTION

Amplified DNA sequences are common phenomena in bacterial genomes, where they carry out a variety of different functions. According to size, they can be divided into large (>1 kbp) and small (<50 bp) repeated sequences (1). Examples for large repeats in *E. coli* are multicopy genes like rRNA operons or duplicated genes for chemoreceptors or the elongation factor Tu. Short amplified sequences occur in transposons and insertion elements (for reviews, see 2,3). In *E. coli*, small repetitive DNA elements have been identified which were designated repetitive extragenic palindromic sequences (REP) sharing a 38 bp consensus sequence. They are frequently located downstream of the 3'-end of coding regions, are cotranscribed and seem to prevent mRNA from nucleolytic degradation during the translation process by increasing its 3'-end stability (4). Another important feature of repetitive DNA elements is the mediation of antigenic switching in pathogenic bacteria. Examples are the antigenic variation mechanisms in *Neisseria gonorrhoeae* (5), *Borrelia hermsii* (6) or *Salmonella typhimurium* (7).

In *Mycoplasma*, a genus of bacteria belonging to the class of Mollicutes the members of which lack a cell wall and have comparatively small genomes of 700-1600 kbp (8, 9, 10), numerous repeated elements have been reported, for instance a reiterated DNA sequence common to three Mycoplasma species, namely *M. hyorhinis*, *M. hypopneumoniae* and *M. flocculare* (11). *M. pneumoniae*, the subject of this study, causes a primary atypical pneumonia in humans. Different research groups have reported the occurrence of various repetitive DNA elements in this organism. Initially, two types of repeats were identified, called RepMP1 and RepMP2 (12). They are 300 bp (RepMP1) and 150 bp (RepMP2) long and are present in at least 10 and 8 copies dispersed on the bacterial chromosome. One copy of RepMP2 is part of the P1 gene which codes for the main adhesin of *M. pneumoniae* a protein of 170 kDa molecular weight (13). This surface localized protein clusters at a tip-like structure (13) and mediates adsorption of the bacterium to its host cells. The corresponding gene has been cloned and sequenced independently by Su *et al.* (14) and Inamine *et al.* (15). A recent study (16) indicates that the P1 gene is organized within an operon and flanked by two open reading frames (ORF's) coding for hypothetical proteins of 28 kDa (upstream of P1, ORF 4) and 130 kDa (downstream of P1, ORF 6) molecular weight; yet a polycistronic mRNA has not been identified.

RepMP2 was localized around nucleotide position 4000 within the 4884 bp long P1 gene (12). Additional repetitive DNA sequences forming part of the P1 operon were found by Su *et al.* (17) and Wenzel *et al.* (18). They identified a region of ca. 1.8 kbp at the 5'-end, an imprecisely defined one downstream of the 3'-end of the P1 gene representing the 5'-end of the postulated 130 kDa protein and another also 1.8 kbp long repetitive sequence in the middle of the P1 gene containing the formerly described RepMP 2 (see Fig. 3). In a recent study, Colman *et al.* (19) showed that another sequence at least 400 bp long in the middle of the hypothetical 130 kDa gene (ORF 6) is also repeated and occurs in as many as eight copies. This element was designated CDS1. Up to now, a function could not be assigned to any of these repeated elements.

In this work, a comprehensive analysis of all known repeated DNA elements of the P1 operon was undertaken with respect to size, copy number, distribution on the *M. pneumoniae* genome and conservation within strains. This was facilitated by an ordered clone library consisting of 34 cosmids, 2 λ phages and one plasmid which covers the complete 840 kbp genome of *Mycoplasma pneumoniae* (20, 21).
MATERIALS AND METHODS

DNA extraction from *M. pneumoniae*

*M. pneumoniae* strain M129 in the 18th broth passage was grown at 37°C using Roux-bottles, containing 120 ml of modified Hayflick medium (22). After a color change of the medium from red to yellow, cells were harvested by centrifugation at 12,500×g for 15 min at 4°C, washed twice in PBS buffer (10 mM sodium phosphate [pH 7.2], 150 mM NaCl) and suspended in 1/10 the original volume of PBS buffer.

2–3 ml of the cell suspension (equivalent to 2×10⁹ cells) were centrifuged and resuspended in 2 ml 50 mM Tris-HCl pH 8.0, 50 mM EDTA, 0.3% SDS, 20 μg/ml Proteinase K. This solution was incubated at 50°C for 2 hours, extracted twice with phenol and once with 1:1 phenol-chloroform. The aqueous phase was dialyzed against 10 mM Tris-HCl pH 8.0, 0.1 mM EDTA.

Subcloning

*M. pneumoniae* DNA (cosmid library) was digested by restriction endonucleases (Boehringer and Pharmacia) under conditions as recommended by the manufacturers. The fragments were ligated into plasmid vectors according to (23). For screening plasmid libraries, colonies were transferred to nitrocellulose filters, pretreated with 10% SDS, the DNA denatured with 1.5 mM NaCl, 0.5 M NaOH and neutralized in 1.5 mM NaCl, 0.5 mM Tris-HCl pH 8.0. After drying and baking at 80°C for 2h the nitrocellulose filters were preincubated at 37 °C in 50 mM Tris-HCl pH 8.0, 1 M NaCl, 1 mM EDTA, 0.1% SDS for 1 h and then used for hybridization.

Southern blotting

The restricted DNA was analyzed on agarose gels (0.8% to 2.0%). After electrophoresis, the gels were soaked for 5 min in 0.25 M HCl and then blotted directly onto nylon membranes (Pall, Biodyne B 0.45 μm), using 0.4 N NaOH as transfer solution, according to the manufacturer's instructions. For optimal fixation, the filters were dried at 80°C for 30 minutes.

Hybridization

**DNA probes labelled with [α-^32^P]-dATP.** After prehybridization in 5×SSC (1×SSC: 150 mM NaCl, 15 mM sodium citrate pH 7.2), 10×Denhardt’s solution (1×Denhardt’s solution: 0.02% Ficoll, 0.02% polyvinylpyrrolidione, 0.02% bovine serum albumin), 1% SDS, 50 μg/ml denatured herring sperm DNA at 37°C for at last 4 h, the hybridization step was performed in 50% formamide, 5×SSC, 5×Denhardt’s solution by adding the appropriate ^32^P-labeled DNA probe and incubating at 37°C overnight. The membranes were washed twice with 2×SSC, 0.1% SDS at 60°C for 15 minutes and at least for 30 minutes with 1×SSC, 0.1% SDS. Autoradiographic signals were obtained by exposing X-ray films (Kodak, Amersham) overnight to the membranes.

**DNA probes labelled with digoxigenin-11-dUTP.** This was done as described above except for the addition of 5% blocking reagent (Boehringer Mannheim) to the hybridization solutions.

Radioactive labelling of nucleic acids

DNA was labelled with [α-^32^P]-dATP by nick translation (nick translation kit, Amersham).

Reactions were carried out for 120 minutes at 15°C and stopped by addition of EDTA and phenol. DNA was separated from unincorporated nucleotides by Sephadex G-50 (Pharmacia) chromatography and denatured by boiling for 10 minutes before use.

Nonradioactive labelling of nucleic acids

Labelling of DNA with digoxigenin-11-dUTP was carried out according to the instructions of the supplier (Boehringer Mannheim).

Figure 1. Southern blot analysis of genomic *M. pneumoniae* DNA. The DNA was digested with EcoRI, analysed on a 0.8% agarose gel and transferred to nylon membranes. The filters were cut into several strips and hybridized with 3 different probes specific for ORF 4, ORF 6 and gene PI (130 = ORF 6; P1 = P1 gene; 28 = ORF 4).

Figure 2. Southern blot analysis of the PI gene. The 5.7 kbp EcoRI fragment cloned in pSPTMP1 was digested with EcoRI and HindIII. The resulting fragments (32 bp to 1.3 kbp) were separated on an agarose gel, blotted, cut into strips and hybridized against [α-^32^P]-dATP labelled cosmid clones which were known to contain P1 related sequences. Cosmid D2 was used as a negative control. Sizes are in bp.
RESULTS
Repetitive regions in the PI operon
First of all it was necessary to determine the extent of unique and repeated DNA sequences in the postulated coding regions of the 10 kbp PI operon. For that purpose, EcoRI restriction fragments (Figure 7) representing ORF 4 (1.1 kbp), the PI gene (5.7 kbp), and ORF 6 (2.7 kbp) were subcloned into plasmids and used as probes in hybridization experiments against Southern blots of EcoRI digested M. pneumoniae DNA (Figure 1). ORF 4 was found to be unique whereas the PI- and ORF 6-specific probes revealed a number of cross-reacting fragments in the M. pneumoniae genome. These findings are consistent with the results of Su et al. (17) and Colman et al. (19).

Repeats within the PI gene
The construction of an ordered clone library representing the complete genome of M. pneumoniae in a set of overlapping or adjacent cosmid clones, using restriction endonuclease EcoRI as cloning enzyme was recently described. Therefore, it is possible to align any repeated DNA element to a specific cloned EcoRI fragment. A 5.7 kbp EcoRI fragment (see cosmid clone pcosMP2, Figure 7), subcloned in the plasmid pSPT18 (pSPTMP2), was used as a probe for gene PI specific repeated sequences. This fragment contains the complete PI gene and in addition a 600 bp long region from the 5'-end of ORF 6 (16). This clone was 32P-labelled by nick translation and hybridized against a Southern blot containing the EcoRI-restricted cosmid clones of the gene library. 16 clones reacted positively showing that they contained PI cross-reacting sequences. These clones were 32P-labelled and hybridized against a Southern blot of EcoRI/HinCII-restricted DNA from pSPTMP2 (see Figure 2). These enzymes subdivide the PI gene and ORF 6 into 10 fragments between 32 bp and 1.3 kbp in size. The PI gene contains two repetitive regions, one at its 5'-end represented by

Figure 3. Comprehensive diagram of repeated and single copy regions in and around Pl-gene. Data from Su et al. (17) and Colman et al. (19) (A) were compared to our results (B). The start codon of PI (ATG) is numbered as 1. Hatched boxes represent DNA stretches with homology to different cosmids. Fig. 3B: A, B, D, E, G, H, I and J are HinCII subfragments of the PI gene. C and F are HinCII/EcoRI subclones. The numbers above the hatched boxes indicate the size in bp of the different restriction fragments used for detailed analysis of the repeated elements. Diagonal hatched boxes of cosmid E7 show the order of the PI operon. The interrupted lines between the repetitive elements indicate, that we can exclude, that in any of the analyzed cosmids RepMP4, RepMP2/3 and RepMP5 are arranged as in the PI operon. Light shaded boxes are used for those elements which show length variation. The two light shaded boxes marked with stars give a maximum length estimation of these repetitive elements. In this case a detailed analysis was not feasible because they are located on one cosmid.
the restriction fragments F and A (see Figure 3) and another one which is characterized by the fragments B and E near the 3'-end. The previously mentioned RepMP2 element is located within fragment E. Since both fragments F/A and B/E are always coupled, we decided to define each pair as one repeat. According to our previous nomenclature, these repeats were numbered RepMP4 (F/A) and RepMP2/3 (B/E). All other regions of the P1 gene appeared to be unique under the stringency conditions used for these hybridizations (see Materials and Methods). A third repeated region exists downstream of the 3'-end of P1 represented by fragment C. This repeat is present within ORF 6 and will be discussed in the following chapter. A comprehensive description of repeated and single copy regions in and around P1 is given in Figure 3. This figure also contains the results of a similar investigation performed by Su et al. (17), however without use of cloned fragments outside the P1 operon. It can be seen that the extent of the repetitive P1 sequences is very similar in both analyses.

This raises the question whether all copies of RepMP4 and RepMP2/3 have the same extent of homology to the P1 gene. To solve this problem, all HincII (EcoRI/HincII) fragments of PSPTMPP1 were subcloned into the plasmid pDM1.2 (24) (for the EcoRI/HincII fragments pSPT18 was used). These clones were further digested by several restriction enzymes, separated on agarose gels, blotted and hybridized with the appropriate ^P-labelled cosmids. As an example, the analysis of HincII-fragment A, which is part of RepMP4, is shown in Figure 4. Together with the analysis of HincII-fragment F, these data show that in the case of RepMP4 distinct variations in length ranging from 1.1-1.5 kbp exist between the individual copies of the M. pneumoniae genome (see Figure 4 lane A19) whereas all RepMP2/3 elements seem to have a similar size of about 1.8 kbp (see Figure 3). Differences in signal intensities may be due to varying degrees of homology on DNA level between individual restriction fragments.

Finally, the RepMP elements were correlated to the P1 positive EcoRI restriction fragments of the cosmid library by hybridization of nonradioactive (digoxigenin-11-dUTP) labelled RepMP2/3 and RepMP4 specific probes. Altogether 10 copies of RepMP2/3, two more than recently reported (12), and 8 copies of RepMP4 were identified (Table 1, Figure 7).

Repeats of the putative 130 kDa gene (ORF 6)

Su et al. (17) and Colman et al. (19) have shown that part of ORF 6 is present in multiple copies on the M. pneumoniae chromosome (see Figure 3). Therefore this region was subjected

| Table 1. Correlation of the repetitive elements RepMP2/3 and RepMP4 to the EcoRI restriction fragments in the genome of M. pneumoniae |
|---|---|---|
| **Cosmids** | **EcoRI-fragments (kbp)** | **Cosmids** | **EcoRI-fragments (kbp)** |
| pcosMP GT9 | 25 | pcosMP A5 | 24 |
| pcosMP P1, P2 | 20.1 | pcosMP A19 | 19 |
| pcosMP A19 | 19 | pcosMP C9, R1 | 15.5 |
| pcosMP A65, H10 | 7 | pcosMP C9, A1 | 7.65 |
| pcosMP A1, E7 | 5.7 | pcosMP P2, F4 | 6 |
| pcosMP A1, E7 | 5.7 | pcosMP A1, E7 | 5.7 |
| pcosMP F11 | 5 | pcosMP H8 | 5.2 |
| pcosMP C9 | 2.3 | pcosMP GT9 | 4 |
| pcosMP R1, R2 | 2.3 |

Figure 4. Analysis of P1 gene restriction fragment A (main component of RepMP4, Fig. 3B), using cosmids with RepMP4 related sequences as hybridization probes. The subfragments were generated by the following restriction endonucleases: HindIII/AsuII 112 bp; PvuI/HindII 180 bp; Bam HI/PvuI 254 bp; DraI/BamHI 324 bp; AsuII/DraI 429 bp.

Figure 5. Analysis of RepMP5, represented by the 2.7 kbp EcoRI fragment cloned in pE7BV which also contains the 1.1 kbp long ORF 4 specific EcoRI fragment. The plasmid DNA was digested with EcoRI, SalI and KpnI, blotted and hybridized with RepMP5 positive cosmids. As a positive control, the two cosmids E7, A1 (both containing the P1 operon) and the plasmid pE7BV (= 130) were used.
Table 2. Correlation of RepMP5 to EcoRI restriction fragments in *M. pneumoniae* genome

<table>
<thead>
<tr>
<th>Cosmids</th>
<th>RepMP5</th>
<th>EcoRI-fragments (kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcosMP A5</td>
<td></td>
<td>24</td>
</tr>
<tr>
<td>pcosMP C9, R1</td>
<td></td>
<td>15.5</td>
</tr>
<tr>
<td>pcosMP R1, R2</td>
<td></td>
<td>6.5</td>
</tr>
<tr>
<td>pcosMP F4, P2</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>pcosMP E7</td>
<td></td>
<td>4.4</td>
</tr>
<tr>
<td>pcosMP A1, E7</td>
<td></td>
<td>2.7</td>
</tr>
<tr>
<td>pcosMP H91</td>
<td></td>
<td>1.8</td>
</tr>
<tr>
<td>pcosMP H8, P1</td>
<td></td>
<td>1.65</td>
</tr>
<tr>
<td>pcosMP R1, R2</td>
<td></td>
<td>1.2</td>
</tr>
</tbody>
</table>

Figure 6. Schematic diagram of the cosmid library of *M. pneumoniae*, showing the distribution of RepMP1, RepMP2/3, RepMP4 and RepMP5 on the genome. One line represents one repetitive element. The P1 operon is localized on the cosmids E7 and A1.

Figure 7. EcoRI restriction map of cosmids containing repetitive elements. The construction of the map of the complete genome is described elsewhere (Wenzel et al. in prep.). Numbers indicate size of the EcoRI fragments in kbp.

to a similar analysis as for the PI gene with the aim of determining the extent of these repeated elements. As starting point, the 2.7 kbp EcoRI fragment containing almost the complete 130 kDa gene (see Figure 3) was cloned in the plasmid pDM 1.7 (24) resulting in pE7BV. Hybridization of a Southern blot of the EcoRI digested cosmid library with this digoxigenin-labelled clone as probe revealed 9 cross-reacting EcoRI fragments. These were identical to those already identified by the HincII/EcoRI restriction fragment C (see above). As mentioned, this fragment carries the 5'-end of ORF 6 and is immediately adjacent to the 2.7 kbp EcoRI fragment. ORF 6-positive clones are shown in Table 2. Then the 2.7 kbp EcoRI fragment was further digested and probed with the corresponding digoxigenin-labelled cosmid clones (see Figure 5). The results indicate that in most cases 2.2 kbp of ORF 6 are repeated. Only in the cosmid clones C9, P2 and H91 is this repeat shortened to 1.9 kbp (see Figure 3). Therefore ORF 6 has to be regarded as one repeat and was designated as RepMP5. It has to be mentioned that the C-terminal 370 bp of the 3.7 kbp gene were not analyzed for they were downstream of the 2.7 kbp EcoRI fragment.

**Distribution on the genome**

In order to analyze whether the various repeats of the PI operon appear in a similar organization elsewhere on the chromosome their positions were localized on the EcoRI restriction map of the individual cosmids. The construction of this EcoRI map covering the complete genome is described elsewhere (Wenzel, R. and Herrmann, R., in prep.).

Figures 6 and 7 summarize the distribution of all repetitive DNA elements so far identified on the genome of *M. pneumoniae*, represented by the ordered clone library. The data on RepMP1 — which does not occur in the PI operon — were included in this analysis since we had demonstrated that in three cases RepMP1 and RepMP2 have been linked (12). Adding up the approximate sizes of all
individual repeats, it turns out that about 50 kbp or 6% of the chromosomal DNA is present in multiple copies. Although there is no significant distribution pattern, a certain accumulation of repeats is observed in an area around the P1 operon (Figures 6, 7) and less pronounced in three other regions of the map.

Conservation in other M. pneumoniae strains

Assuming that these repetitive DNA sequences have some important functions we investigated whether these repeats were conserved in other M. pneumoniae species. The following strains had been already characterized in detail by other methods (25): P11428, FH, MAC, M129P3, M129B18 and B176. The first four are individual isolates. M129B18 and B176 are cloned derivatives from later passages of M129P3.

For Southern blot analysis genomic DNA of all six strains was digested with restriction endonuclease EcoRI and analyzed with digoxigenin labelled P1 operon specific probes (Figure 8). Only in M. pneumoniae FH was a restriction fragment length polymorphism (RFLP) observed in two cases: the 4 kbp long EcoRI fragment carrying RepMP4 and cloned in cosmid pcosMPGT9 (Table 1) is slightly larger and the 2.7 kbp long EcoRI fragment containing most of ORF 6 is about 200 nucleotides shorter. No other RFLP was found in the P1 operon region of the analyzed strains.

DISCUSSION

Table 3 summarizes the data on all repeated DNA elements in M. pneumoniae with respect to copy number and size. Copy numbers have to be considered as a minimal estimate, because it can not be excluded that more repeats of the same type are located on one EcoRI restriction fragment. It may also be possible that other classes of repetitive elements exist in M. pneumoniae which have not been discovered yet. We propose to designate the repeats as indicated.

Present models of evolutionary relationships between the members of the rather heterogeneous class of the Mollicutes are based on the assumption that mycoplasmas are subjected to a high selection pressure. The driving force seems to be an increased mutation rate per base pair (26). Since parts of RepMP2/3 and RepMP5 have been sequenced (12, 18) and revealed highly conserved consensus sequences, it is difficult to imagine how the striking DNA homology between individual copies is maintained.

As mentioned repetitive DNA elements have been discovered in many bacterial species. Therefore the detection of amplified DNA sequences in mycoplasmas did not appear to be exceptional. Most surprising is the large size and the relatively high copy number of these elements present in a genome which is considered generally to be among the smallest of an independently replicating organism (8, 27). Since as much as 6% of the DNA of the small genome of 840 kbp is present in conserved multiple copies (as defined by our hybridization conditions, see Materials and Methods), we believe that these elements play an important role.

Su and co-workers (17) have proposed a model in which the repeated elements are considered as 'constant' P1-related multicycle gene domains (PMGD). They are interrupted by single copy (i.e. 'variable') domains, thus producing pseudogenes or gene copies of P1. Expression of these pseudogenes may alter structural and functional properties of P1 by the production of a variety of different proteins.

With respect to this hypothesis, we examined the arrangement of the repetitive elements to each other on the M. pneumoniae genome. For instance, RepMP2/3 and RepMP4 are located in five instances (pcosMPC9, pcosMPP2, pcosMPA5/F11, pcosMPA19, and pcosMPGT9) on the same or an adjacent EcoRI fragment allowing in theory an organization of these two elements as mentioned above. So far, DNA sequence analysis of the region in question of the cosmids pcosMPC9, pcosMPP2, and pcosMPA5/F11 had shown very clearly that in none of these cases RepMP4 and RepMP2/3 were arranged as in the P1 gene. This argues against the model proposed by Su et al. (17). Furthermore, no variation for protein P1 (as far as molecular weight is concerned) has been demonstrated in a survey using different M. pneumoniae isolates and polyclonal sera from pneumonia patients (28).

Nevertheless, it can not be ruled out that minor structural
changes occur in PI and its related proposed 130 kDa protein being caused by the exchange of the conserved repeated elements through homologous recombination. This is also our working hypothesis which has to be confirmed by a detailed analysis. It has to be further investigated if the mentioned RFLPs (see Results) are caused by DNA rearrangement, if copies of repeated DNA sequences are involved in this process and if recombination events exert any effect on protein level.

ACKNOWLEDGEMENT

We are indebted to Elsbeth Pirkl for excellent technical assistance and thank Dr. M.F. Barile for providing a collection of *M. pneumoniae* strains, Dr. K.F. Bott for supplying information prior to publication and Dr. R. Rigg for critical reading of the manuscript. This work was supported by the BMFT (BCT 0381/5).

REFERENCES