Analysis of genetic polymorphisms affecting the four phospholipase C (plc) genes in *Mycobacterium tuberculosis* complex clinical isolates

C. Viana-Niero, P. E. de Haas, D. van Soolingen and S. C. Leão

1Departamento de Microbiologia, Imunologia e Parasitologia, Universidade Federal de São Paulo – Escola Paulista de Medicina (UNIFESP-EPM), Rua Botucatu, 862 3^o^ andar, 04023-062, São Paulo, Brazil

2Diagnostic Laboratory of Infectious Diseases and Perinatal Screening, National Institute of Public Health and the Environment (RIVM), PO Box 1, 3720 BA, The Netherlands

The *Mycobacterium tuberculosis* genome contains four highly related genes which present significant similarity to *Pseudomonas aeruginosa* genes encoding phospholipase C enzymes. Three of these genes, *plcA*, *plcB* and *plcC*, are organized in tandem (locus *plcABC*). The fourth gene, *plcD*, is located in a different region. This study investigates variations in *plcABC* and *plcD* genes in clinical isolates of *M. tuberculosis*, *Mycobacterium africanum* and ‘*Mycobacterium canetti*’. Genetic polymorphisms were examined by PCR, Southern blot hybridization, sequence analysis and RT-PCR. Seven *M. tuberculosis* isolates contain insertions of IS6110 elements within *plcA*, *plcC* or *plcD*. In 19 of 25 *M. tuberculosis* isolates examined, genomic deletions were identified, resulting in loss of parts of genes or complete genes from the *plcABC* and/or *plcD* loci. Partial *plcD* deletion was observed in one *M. africanum* isolate. In each case, deletions were associated with the presence of a copy of the IS6110 element and in all occurrences IS6110 was transposed in the same orientation. A mechanism of deletion resulting from homologous recombination of two copies of IS6110 was recognized in a group of genetically related *M. tuberculosis* isolates. Five *M. tuberculosis* isolates presented major polymorphisms in the *plcABC* and *plcD* regions, along with loss of expression competence that affected all four *plc* genes.

Phospholipase C is a well-known bacterial virulence factor. The precise role of phospholipase C in the pathogenicity of *M. tuberculosis* is unknown, but considering the potential importance that the *plc* genes may have in the virulence of the tubercle bacillus, the study of isolates cultured from patients with active tuberculosis bearing genetic variations affecting these genes may provide insights into the significance of phospholipase C enzymes for tuberculosis pathogenicity.

**INTRODUCTION**

The *Mycobacterium tuberculosis* genome contains four highly related genes encoding phospholipase C enzymes (Cole et al., 1998). Three genes, *plcA*, *plcB* and *plcC*, are organized in tandem (*plcABC* locus). The fourth gene, *plcD*, is located in a different region. In the laboratory strain *M. tuberculosis* H37Rv, *plcD* is truncated by insertion of a copy of IS6110 and by deletion of a 7·9 kb fragment, designated RvD2 (Gordon et al., 1999; Brosch et al., 2002). An intact copy of *plcD* is present in *M. tuberculosis* CDC1551, *Mycobacterium bovis* and *M. bovis* bacille Calmette–Guérin (BCG).

*M. tuberculosis plc* genes present significant similarity with *Pseudomonas aeruginosa* genes *plcH* and *plcN*, which encode haemolytic and non-haemolytic phospholipase C enzymes, respectively (Leão et al., 1995; Ostroff et al., 1990). The β-haemolysis exhibited by *Escherichia coli* extracts over-expressing recombinant PlcA suggests enzymic activity of *M. tuberculosis* Plc proteins (Leão et al., 1995). Moreover, sphingomyelinase and phospholipase C activities were detected in cell extracts from *M. tuberculosis and in Mycobacterium smegmatis* overexpressing recombinant PlcA or PlcB (Johansen et al., 1996). The product of the *plcA* gene was detected in cell extracts from *M. tuberculosis* *grown in vitro* and inside macrophages (Matsui et al., 2000).

Bacterial phospholipase C enzymes are recognized as important virulence factors of *P. aeruginosa* (Berka et al., 1981), *Bacillus cereus* (Gilmore et al., 1989), *Clostridium*
Some members of the *M. tuberculosis* complex, using a different pair of primers to amplify a 152 bp product of *plcA*. They concluded that RD5 is present in most, but not all, isolates of *M. tuberculosis*, *M. africanum* and *M. microti*, and consistently absent from *M. bovis* and *M. bovis* BCG isolates.

**METHODS**

**Bacterial isolates.** Twenty-five *M. tuberculosis* isolates, belonging to a collection of 790 strains from the RIVM (Bilthoven, The Netherlands) isolated from patients with active tuberculosis in 1998, were included in this study. The *mtp40*-associated RFLP was determined in all 790 *M. tuberculosis* isolates and 50 distinct banding patterns were distinguished. At least two isolates from each *mtp40*-RFLP type were subjected to PCR analysis using *mtp40*-specific primers. Isolates from 14 out of the 50 *mtp40*-RFLP types were negative in this PCR, and comprised 54/790 of the strains (6-8%). Ten *mtp40*-PCR-negative and 13 *mtp40*-PCR-positive isolates were selected from the five most prevalent *mtp40*-RFLP types in our collection. In addition, two *mtp40*-RFLP type strains with *mtp40* PCR fragments larger than the normal size (approx. 2 kb) were also included. The most prevalent *mtp40*-RFLP types represented 85% of all 790 strains investigated. *M. tuberculosis* H37Rv and *M. bovis* BCG Pasteur were used as controls (Table 1).

Three *M. africanum* and four *M. canetti* isolates from the National Reference Laboratory for Mycobacteria at Institut Pasteur (Paris, France) culture collection were kindly provided by Dr Véronique Vincent. These strains were characterized extensively using biochemical tests, oxyR-PCR, pncA-PCR, and molecular typing methods, such as IS6110-RFLP typing and spoligotyping (Viana-Niero et al., 2001) (Table 1).

Cultures were grown on Löwenstein–Jensen (LJ) solid medium or in Middlebrook 7H9 (Difco) liquid medium supplemented with albumin, glucose and catalase (Difco) and kept frozen at −70 °C in aliquots of 7H9-ADC with 15% (v/v) glycerol. DNA was prepared as described by van Embden et al. (1993).

**PCR and sequencing.** For primer design, sequences for *plcABC* from *M. tuberculosis* H37Rv and *plcD* from *M. tuberculosis* CDC1551 were obtained from GenBank (http://www.ncbi.nlm.nih.gov/). Primers were designed using Oligo (version 5.0; National Biosciences) and are depicted in Table 2. Each of the primer pairs AD/AR, BD/BR,
Table 1. Isolates included in the study and results of PCR with primers PT1/PT2 (mtp40), AD/AR (plcA), BD/BR (plcB), CD/CR (plcC) and DD/DR (plcD)

+ Positive; −, negative; +*, ∼2·0 kb product; +**, ∼3·0 kb product. IS6110 flanking sequences in the plcABC and plcD loci determined by sequence analysis are shown, with direct repeats in bold. ND, Not determined. X, Not applicable to these strains. Positions of insertions in the plcABC locus were obtained from *M. tuberculosis* H37Rv MTCY98 (GenBank accession no. Z83860); positions of insertions in the plcD locus were obtained from *M. tuberculosis* CDC1551 (GenBank accession no. AE007040).

<table>
<thead>
<tr>
<th>Species</th>
<th>Isolate</th>
<th>PCR</th>
<th>Flanking sequences</th>
<th>Insertion position</th>
<th>PCR</th>
<th>Flanking sequences</th>
<th>Insertion position</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1344</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-286</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1894</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-432</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+**</td>
<td>GGCTGGAAC</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-67</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-365</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-535</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-438</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-969</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-269</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1674</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+**</td>
<td>GGCTGGAAC</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-858</td>
<td>+</td>
<td>**</td>
<td>AGGTCA</td>
<td>TGT</td>
<td>TGT</td>
<td>TGT</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-2433</td>
<td>+</td>
<td>**</td>
<td>CGGGATTA</td>
<td>TGGGTGTAAC</td>
<td>22319 (plcA)</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-742</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-818</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1505</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1289</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1177</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1200</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-803</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-488</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1389</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-279</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-1438</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
<tr>
<td><em>M. tuberculosis</em></td>
<td>97-66</td>
<td>+</td>
<td>**</td>
<td>CGGGAGCCAG</td>
<td>TGA</td>
<td>GCCTGGAAC</td>
<td>X</td>
</tr>
</tbody>
</table>
CD/CR and DD/DR amplified an individual plc gene and amplification specificity was evaluated by restriction analysis of the amplified fragments (data not shown). Internal primers AD2/AR2, BD2/BR2, CD2/CR2 and DD3/DR3, PT1/PT2 primers specific for the mtp40 sequence, and IS1/IS2 primers from IS6110 and PPE1 from Rs2356c (H37Rv) were used to address specific sequencing issues (Table 2). PCR amplicons were sequenced in an automated ABI Prism 377 sequencer (Perkin-Elmer) using the BigDye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems). Sequences were compared using BLAST (http://www.ncbi.nlm.nih.gov/blast).

**Southern blot analysis of plc polymorphisms.** Southern blotting was performed as described by van Embden et al. (1993). Briefly, 2 µg DNA was digested using PvuII and subjected to electrophoresis in 0-8 % (w/v) agarose in 0-5 x TBE pH 8-0, at 2 V cm⁻¹. A mixture of PvuII-digested supercoiled DNA ladder (Invitrogen) and HaeIII-digested φX174 DNA (Invitrogen) was used as internal DNA size marker. DNA was blotted onto nylon membranes (Hybond-N-plus; Amersham Biosciences) and probed, in different experiments, with probes complementary to each plc gene. Probes were prepared by PCR using primers AD/AR, BD/BR, CD/CR or DD/DR. Blots were also probed with the internal size marker. Probes were covalently labelled with peroxidase by glutaraldehyde and detected using the ECL Direct System (Amersham) according to the manufacturer’s instructions. Membranes were exposed to an X-ray film (X-OMAT; Kodak). Fingerprinting were analyzed using GELCOMPAR II version 2.5 software (Applied Maths). Fingerprint autoradiograms were superimposed on internal marker autoradiograms for normalization.

**Cloning and sequencing of a plcC-specific PstI restriction fragment.** DNA from isolate 97-1177 was digested using PstI at 37 °C for 2 h, and subjected to electrophoresis in 0-8 % (w/v) agarose, along with the internal size marker. DNA was blotted onto a nylon membrane and hybridized with probes from plcC, IS6110 and the internal size marker, as described previously. The size of the fragment hybridizing with both the plcC and the IS6110 probe was calculated using GELCOMPAR II. Agarose gel slices around the selected band of PstI restriction fragment were excised from a second gel, purified using the Gel Extraction Kit (Qiagen) and ligated into PstI-digested dephosphorylated (CIAP; Invitrogen) pBluescript SK+ (Stratagene), using T4 DNA ligase (Invitrogen) at 16 °C, overnight. Ligated DNA was transformed into electrocompetent E. coli DH5α cells, which were plated onto LB-agar plates containing 100 µg ampicillin ml⁻¹, 0-25 mM IPTG and 1 mM X-Gal. Colony lifts were carried out on Whatman 541 filters and filters were hybridized with the plcC probe, labelled with [α-³²P]dCTP with Ready-To-Go DNA Labelling Beads (Amersham), according to the protocol described by Maas (1983). Positive colonies visualized by exposure to autoradiographic film were confirmed by PCR with primers CD2/CR2 and by PstI digestion of the plasmid DNA. The recombinant plasmid, isolated using the Qiaprep Spin miniprep kit (Qiagen), was sequenced using M13 primers, as described previously.

**RNA purification and RT-PCR.** Two-week-old cultures in Middlebrook 7H9 ADC medium were used for RNA isolation using FastRNA kit-BLUE (QBiogene) and a FastPrep Instrument (QBiogene), according to the manufacturer’s instructions. DNA was removed by DNase I (Invitrogen). RT-PCR was performed with 1 µg RNA using the Access RT-PCR System (Promega) and internal primers specific for each plc gene. Amplicons were sequenced. Control of RT-PCR was performed using primers from atpB, a gene encoding a key component of the proton channel, which is transcribed during axenic growth and after infection of human macrophages (Graham & Clark-Curtiss, 1999).
RESULTS

PCR revealed polymorphisms in the four plc genes

PCR assays were set up using primer pairs AD/AR, BD/BR, CD/CR and DD/DR, for amplification of a 1599 bp fragment from plcA, a 1611 bp fragment from plcB, a 1610 bp fragment from plcC and a 1478 bp fragment from plcD, respectively (Fig. 1). As expected, the plcD amplicon was not detected in M. tuberculosis H37Rv and the plcA, plcB and plcC amplicons were not detected in M. bovis BCG Pasteur.

Results of PCR are summarized in Table 1. Three of the 25 M. tuberculosis, two of the M. africanum and the four 'M. canettii' isolates produced amplicons of all four plc genes. Products from plcA, plcB and plcC, but not from plcD, were amplified from six M. tuberculosis isolates and one M. africanum isolate. The only product amplified from five isolates of M. tuberculosis was plcD. Four M. tuberculosis isolates did not produce any plc amplicon.

Instead of the normal mtp40 product of 396 bp, two isolates, 97-858 and 97-2433, produced amplicons of approximately 2 kb in size with primers PT1/PT2. plcC products larger than expected, of approximately 3 kb, were generated from

Table 2. Sequences and localization of primers used in this work

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
<th>Positions</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>GCAAGTGAGGCCAAGCCCACTCCT</td>
<td>23 656–23 632*</td>
<td>This work</td>
</tr>
<tr>
<td>AR</td>
<td>GCAATGACGGCAAGGCCGAGATTACC</td>
<td>22 057–22 081*</td>
<td>This work</td>
</tr>
<tr>
<td>BD</td>
<td>CGTGACGGGATAGGCAGGGCCGACAA</td>
<td>21 882–21 858*</td>
<td>This work</td>
</tr>
<tr>
<td>BR</td>
<td>CTGACCACATCGCGTCTCCGCTCCTG</td>
<td>20 271–20 295*</td>
<td>This work</td>
</tr>
<tr>
<td>CD</td>
<td>TAAAGGCCATTGACAGGAGAGAG</td>
<td>20 307–20 284*</td>
<td>This work</td>
</tr>
<tr>
<td>CR</td>
<td>CAACCCGTATGCTGAAACATCC</td>
<td>18 697–18 720*</td>
<td>This work</td>
</tr>
<tr>
<td>DD</td>
<td>TTCGCGGAGGAGGGCAGATCTCCTAG</td>
<td>3 526–3 507†</td>
<td>This work</td>
</tr>
<tr>
<td>DR</td>
<td>TACGTCGTCGGATGACTCCGACACGACAC</td>
<td>2 048–2 067†</td>
<td>This work</td>
</tr>
<tr>
<td>atpb1</td>
<td>TTTGAGGGACGGATACAGTCAG</td>
<td>24 297–24 317‡</td>
<td>This work</td>
</tr>
<tr>
<td>atpb2</td>
<td>GAAGATGTCGCGAGGAAGCTCCTCT</td>
<td>24 632–24 652‡</td>
<td>This work</td>
</tr>
<tr>
<td>AD2</td>
<td>ACCATCTGGCAGCAGTACAC</td>
<td>3 23 123–23 104*</td>
<td>This work</td>
</tr>
<tr>
<td>AR2</td>
<td>GCCGTTATTGCTGATGCCAG</td>
<td>22 869–22 886*</td>
<td>This work</td>
</tr>
<tr>
<td>BD2</td>
<td>AGCTGATCGGCATGCAACAC</td>
<td>21 515–21 495*</td>
<td>This work</td>
</tr>
<tr>
<td>BR2</td>
<td>GACTTTGGCTCTGATGACAC</td>
<td>21 231–21 251*</td>
<td>This work</td>
</tr>
<tr>
<td>CD2</td>
<td>GTGGCCAAACACCCGGCAGTGT</td>
<td>19 823–19 803*</td>
<td>This work</td>
</tr>
<tr>
<td>CR2</td>
<td>ACTCGGGGCAGCTGTTGTGAG</td>
<td>19 374–19 394*</td>
<td>This work</td>
</tr>
<tr>
<td>DD3</td>
<td>GCTTCGCGGCAGCTCCTAG</td>
<td>2 2 696–2 676†</td>
<td>This work</td>
</tr>
<tr>
<td>DR3</td>
<td>CGCCGCGGTAGCAGAAGAC</td>
<td>2 467–2 484†</td>
<td>This work</td>
</tr>
<tr>
<td>IS1</td>
<td>CTCCTGGAGCAGTCAGG</td>
<td>3 183–3 199*</td>
<td>Eisenach et al. (1990)</td>
</tr>
<tr>
<td>IS2</td>
<td>TCTGTCAGGCGCCGCTTCCG</td>
<td>3 227–3 245*</td>
<td>Eisenach et al. (1990)</td>
</tr>
<tr>
<td>PT1</td>
<td>CAAGCCGGCGGTGG</td>
<td>22 183–22 199*</td>
<td>Del Portillo et al. (1991)</td>
</tr>
<tr>
<td>PT2</td>
<td>ACCTGGGGCAGCTGTTGTGAG</td>
<td>22 575–22 560*</td>
<td>Del Portillo et al. (1991)</td>
</tr>
<tr>
<td>PPE1</td>
<td>AAAACCAGTTGTCGCGCTAAG</td>
<td>28 901–28 881*</td>
<td>This work</td>
</tr>
</tbody>
</table>

*Positions in M. tuberculosis H37Rv MTCY98 (GenBank accession no. Z83860).
†Positions in M. tuberculosis CDC1551 (GenBank accession no. AE007040).
‡Positions in M. tuberculosis H37Rv (GenBank accession no. Z73419).

http://mic.sgmjournals.org 971

Fig. 1. Schematic representation of the plcABC and plcD loci. PvuII restriction sites (arrows) were obtained from M. tuberculosis H37Rv MTCY98 (GenBank accession no. Z83860) and M. tuberculosis CDC1551 (GenBank accession no. AE007040). A, B, C, D, fragments generated by amplification using primers AD/AR, BD/BR, CD/CR and DD/DR, respectively. These fragments were used as probes in Southern blot hybridizations. 1–8, Fragments detected by hybridization of Southern blots: 1, 574 bp; 2, 2156 bp; 3, 580 bp; 4, 751 bp; 5, 4657 bp (according to M. tuberculosis H37Rv genome data) or 2983 bp (according to empirical data from this work); 6, ≥4 kb (not determined); 7, 538 bp; 8, 3900 bp.
isolates 97-742, 97-818 and 97-1505. Isolates 97-432 and 97-1389 produced plcD fragments of approximately 3 kb in size. In all cases, re-amplification of these amplicons using primers IS1/IS2 generated products of 123 bp, confirming the insertion of IS6110 elements in the mtp40 region of plcA, and in plcC and plcD, respectively.

To characterize the polymorphisms detected by PCR in the plcABC and plcD genomic regions of these selected isolates, plc-based fingerprints were obtained by the use of restriction endonuclease *Pvu*II and specific DNA probes, homologous to the sequences of each of the four plc genes.

**Southern hybridization and sequence analysis confirmed insertions and deletions**

All isolates were subjected to Southern hybridization analysis. Polymorphisms in plc genes detected by PCR and hybridization were analysed by sequencing of selected amplicons.

Fig. 1 shows the physical map and *Pvu*II restriction sites of plcABC obtained from analysis of *M. tuberculosis* H37Rv cosmid MTCY98 (GenBank accession no. Z83860) and plcD from *M. tuberculosis* CDC1551 cosmid MT1799 (GenBank accession no. AE007040). Fragments A, B, C and D were generated by PCR using primers AD/AR, BD/BR, CD/CR and DD/DR, respectively, and were used as probes in Southern blot hybridization (Fig. 1). Information about *Pvu*II fragment sizes was used to identify specific bands hybridizing with each probe. Empirical results showed that the size of the band located at the 5′ region of plcA, calculated using GELCOMPAR II, was 2983 bp instead of 4657 bp, the expected size according to *M. tuberculosis* H37Rv genome data (Fig. 2).

---

**Fig. 2.** Southern blot hybridization with probes A (plcA), B (plcB), C (plcC) and D (plcD) analysed using GELCOMPAR II version 2.5. (a) Hybridization patterns; (b) specific bands hybridizing with each plc probe. Non-marked fragments are the result of cross-hybridization with fragments from other plc genes. Lanes: 1, *M. tuberculosis* H37Rv; 2, 97-1344, 97-286, 97-1894, 97-432, 97-365, 97-535, 97-438, 97-969, 97-269, 97-1674, 140030063, 140030065, 140030068, 140010059, 140010105, 140010107, 140010110; 3, 97-858; 4, 97-2433; 5, 97-742; 6, 97-818, 97-1505; 7, 97-803, 97-1177, 97-1200, 97-1289; 8, 97-488; 9, 97-1389; 10, 97-279, 97-1438, 97-66, 97-464; 11, *M. bovis* BCG Pasteur; 12, 97-438, 97-535, 97-969, 140030065; 13, 97-365; 14, *M. tuberculosis* H37Rv; 15, 97-279, 97-2438, 97-488, 97-858, 97-1894; 16, *M. bovis* BCG Pasteur; 17, 97-2433; 18, 97-286, 97-66, 140030063, 140030068, 140010059, 140010105, 140010107, 140010110; 19, 97-1389; 20, 97-432; 21, 97-803, 97-818, 97-1505, 97-1289, 97-742, 97-1177, 97-1200, 97-1674; 22, 97-269.
Banding patterns generated by hybridization with the four probes confirmed PCR results showing integrity of all four plc genes in three M. tuberculosis isolates, two M. africanum isolates and all four ‘M. canettii’ isolates. The presence of an intact plcABC locus and the lack of one or more of the plcD-hybridizing bands confirmed PCR results in six M. tuberculosis isolates and one M. africanum isolate (Fig. 2).

In isolates 97-858 and 97-2433, Southern hybridization using probe A was consistent with the presence of a copy of IS6110 in the mtp40 region of plcA (Fig. 2). Sequence analysis of the amplicon from isolate 97-858 revealed the site of insertion at position 22319, which is related to the sequence of cosmid MCTY98 (GenBank accession no. Z83860), and a direct repeat of three nucleotides TGT:TGT at the ends of the insertion element (Table 1 and Fig. 3a). The exact insertion site in isolate 97-2433 was not determined.

In isolates 97-818 and 97-1505, which produced a plcC amplicon of 3·0 kb, hybridization using probe C revealed substitution of the 2156 bp PvuII band by two bands of approximately 1·7 kb in size (Fig. 2). Sequence analysis of the 3·0 kb plcC amplicons from isolates 97-818 and 97-1505 revealed insertion of a copy of IS6110 at position 19849 of plcC (related to GenBank sequence MCTY98) and duplication of four nucleotides at the site of insertion, with one nucleotide change, GCAG:GCAA (Table 1 and Fig. 3c). A similar polymorphism was observed in isolate 97-742, in which the 2156 bp PvuII band was substituted by two bands of approximately 1·5 kb and 1·9 kb in size (Fig. 2). Insertion of a copy of IS6110 was confirmed by sequence analysis at position 19668 of plcC, with a TGA:TGA direct repeat (Table 1 and Fig. 3b).

Six isolates, 97-1289, 97-1177, 97-1200, 97-803, 97-488 and 97-1389, lacked plcA- and plcB-hybridizing bands. Southern hybridization with probe C disclosed shifts in the 2156 bp band. Amplicons of approximately 1 kb in size were produced using primers IS2/CR2. Sequence analysis of these amplicons revealed insertion of IS6110 elements in plcC at position 19849 in four isolates, position 19500 in one isolate and position 19916 in one isolate (Table 1 and Fig. 3d, e, f).

In four isolates, 97-66, 97-279, 97-464 and 97-1438, Southern hybridizations with probes A, B and C revealed a single plcC-hybridizing band of approximately 1·3 kb in size. Generation of a 450 bp product using primers IS2/CR2 and lack of hybridization of the plcC probe with the 574 bp restriction fragment confirmed that these isolates retain only the central part of plcC.

In isolates 97-432 and 97-1389, amplification using primers DD/DR resulted in a fragment of 3·0 kb in size. Hybridization using probe D was suggestive of the presence of an intervening sequence in plcD (Fig. 2). Sequence analysis of the amplicon from isolate 97-432 revealed the insertion of a copy of IS6110, with the right imperfect repeat inserted at position 2611 and the left repeat at position 2698 (relative to the sequence of cosmid MT1799; GenBank accession no. AE007040), and deletion of 87 bp from plcD. No direct repeats were detected at the ends of the IS element (Table 1 and Fig. 4a).
To understand the mechanism involved in the deletion of plc genes in this group of isolates, the sequences flanking the ends of the IS insertion were determined in isolate 97-1177. DNA was digested using PstI and hybridized in Southern blots, using probes complementary to plcC and IS6110 (data not shown). A 4·0 kb PstI DNA fragment, which hybridized with both probes, was cloned and sequenced. The right imperfect repeat of IS6110 was inserted in plcC at position 19 849, and the left repeat was inserted at position 28 509 from MT_CY98 (Table 1 and Fig. 3b). Sequence analysis confirmed the deletion of a segment of 8·6 kb in size, containing the 5’ region of plcC and the complete genes plcB, plcA, PPE38 and PPE39, and its substitution by a copy of the IS6110 element. No direct repeats were detected at the ends of the IS element, suggesting that deletion occurred by homologous recombination between two copies of IS6110. Identical polymorphism was observed in the other three isolates and was corroborated by sequence analysis of a product of 2·0 kb, generated by amplification with primers PPE1 and CR2.

Isolates 97-818 and 97-1505, which carry the IS element in plc without subsequent gene deletion, presented a second copy of IS6110 at position 28 509. Sequence analysis of amplicons obtained by PCR using primers PPE and IS1 confirmed that the second IS element, located between genes PPE39 and PPE40, has the same orientation of the IS6110 element. No direct repeats were detected at the ends of the IS element, suggesting that deletion occurred by homologous recombination between two copies of IS6110. Identical polymorphism was observed in the other three isolates and was corroborated by sequence analysis of a product of 2·0 kb, generated by amplification with primers PPE1 and CR2.

**Fig. 4.** Schematic representation of polymorphisms in plcD, between positions 1 and 14 648 from M. tuberculosis CDC1551 (GenBank accession no. AE007040) and corresponding maps of M. tuberculosis H37Rv and H37Rv. (a) Isolate 97-432; (b) isolates 97-269, 97-803, 97-818, 97-1389 (not shown). A 4·0 kb PstI DNA fragment, which hybridized with both probes, was cloned and sequenced. The right imperfect repeat of IS6110 was inserted in plcC at position 19 849, and the left repeat was inserted at position 28 509 from MTCY98 (Table 1 and Fig. 3b). Sequence analysis confirmed the deletion of a segment of 8·6 kb in size, containing the 5’ region of plcC and the complete genes plcB, plcA, PPE38 and PPE39, and its substitution by a copy of the IS6110 element. No direct repeats were detected at the ends of the IS element, suggesting that deletion occurred by homologous recombination between two copies of IS6110. Identical polymorphism was observed in the other three isolates and was corroborated by sequence analysis of a product of 2·0 kb, generated by amplification with primers PPE1 and CR2.

**Polymorphisms in plc genes were characterized in a group of genetically related isolates**

An interesting polymorphism in the plcABC locus was detected in a group of six genetically related isolates, 97-818, 97-1505, 97-803, 97-1289, 97-1177 and 97-1200. The remaining two isolates retained the genes plcA and plcB. Polymorphism in the plcD genomic region was shared by all six isolates (Fig. 4b).

The hybridization pattern generated by the probe complementary to plcD in nine isolates was indistinguishable (Fig. 2). The expected bands of 538 bp and 3·9 kb were not detected, suggesting deletion of the 5’ region. Sequencing of these amplicons produced by amplification using primers IS2/DR2 confirmed the presence of an IS6110 element at position 2611, relative to the sequence of M. tuberculosis position 2611, relative to the sequence of M. tuberculosis CDC1551 (GenBank accession no. AE007040) (Table 1 and Fig. 4b). The size of the deleted fragment was not determined.

The hybridization pattern generated by the probe complementary to plcD in nine isolates was indistinguishable (Fig. 2). The expected bands of 538 bp and 3·9 kb were not detected, suggesting deletion of the 5’ region. Sequencing of these amplicons produced by amplification using primers IS2/DR2 confirmed the presence of an IS6110 element at position 2611, relative to the sequence of M. tuberculosis CDC1551 (GenBank accession no. AE007040) (Table 1 and Fig. 4b). The size of the deleted fragment was not determined.

**Polymorphisms in plc genes were characterized in a group of genetically related isolates**

An interesting polymorphism in the plcABC locus was detected in a group of six genetically related isolates, 97-818, 97-1505, 97-803, 97-1289, 97-1177 and 97-1200. The remaining two isolates retained the genes plcA and plcB. Polymorphism in the plcD genomic region was shared by all six isolates (Fig. 4b).

The hybridization pattern generated by the probe complementary to plcD in nine isolates was indistinguishable (Fig. 2). The expected bands of 538 bp and 3·9 kb were not detected, suggesting deletion of the 5’ region. Sequencing of these amplicons produced by amplification using primers IS2/DR2 confirmed the presence of an IS6110 element at position 2611, relative to the sequence of M. tuberculosis CDC1551 (GenBank accession no. AE007040) (Table 1 and Fig. 4b). The size of the deleted fragment was not determined.

**Analysis of transcription of plc genes**

We also studied possible disturbances caused by insertions of IS6110 elements in the transcription of plc genes in selected isolates. Disruption of plcA or plcC by an IS6110 element did not impair transcription of adjacent plc genes in isolates 97-858 and 97-818, respectively (Fig. 5). Transcription of plcD was confirmed by RT-PCR in isolates with an intact plcD gene, such as M. bovis BCG Pasteur and isolate 97-279 (Fig. 5). Products of plc genes were not detected by RT-PCR in isolates 97-1289, 97-1177, 97-1200, 97-803 and 97-1389 (not shown).

**DISCUSSION**

The mtp40 sequence, located in the plcA opposite strand, was described before identification and sequencing of plc genes (Parra et al., 1991). It has been used in diagnostic PCR for identification of M. tuberculosis (Del Portillo et al., 1991) and, more recently, to evaluate the occurrence of RD5 deletion in clinical M. tuberculosis isolates or in isolates from...
other members of the *M. tuberculosis* complex (Gordon *et al*., 1999; Parsons *et al*., 2002). There have been conflicting results over the distribution of this sequence in members of the complex, and several researchers have described the existence of *mtp40*-negative *M. tuberculosis* isolates (Liebana *et al*., 1996; Weil *et al*., 1996; Vera-Cabrera *et al*., 1997). These *mtp40*-PCR negative results may be a consequence of RD5/RD7 deletion, other types of deletions, insertions of IS6110 elements or mutations in the primer region, and are indicative of the existence of polymorphisms in this genomic region.

To examine polymorphisms in *plc* genes in members of the *M. tuberculosis* complex that cause disease in humans, we analysed 25 selected *M. tuberculosis* clinical isolates and well-characterized isolates from *M. africanum* and *M. canettii*. This analysis revealed the presence of insertions inside the coding regions of *plcA*, *plcC* and *plcD* in *M. tuberculosis* isolates. No polymorphisms in the *plcABC* region were detected in *M. africanum* and *M. canettii*, but the small number of isolates analysed so far does not allow a definite conclusion. The finding of partial *plcD* deletion in one *M. africanum* isolate suggests that this region is also prone to polymorphisms in this subspecies.

The original report on IS6110 distribution in the genome of *M. tuberculosis* H37Rv indicated that insertions occurred preferentially in non-coding regions (Philipp *et al*., 1996). A recent analysis of data collated from various sources demonstrated that 57 of the 95 analysed insertions (60 %) occurred within coding regions (Sampson *et al*., 2001). The impact of these insertions on the phenotype is unknown, as most of the disrupted genes are multicopy genes present in the genome. Disruption of individual members in a family of genes may have limited impact on phenotype, depending on the activity of the protein encoded.

Insertions of IS6110 elements were detected in *plcA* and *plcC*. RT-PCR analysis provided evidence that disruption of either of these genes does not have a significant polar effect on transcription from neighbouring *plc* genes. Raynaud *et al*., (2002) obtained similar results using mutants that harbour transpon insertion in *plcA*, *plcB* or *plcC*.

In this work, discrete IS6110 insertions in *plcC* at positions 19849 and 19668 were detected. Insertion events at positions 19589, 19645 and 19848, located in *plcC*, were observed in a previous study including 11 isolates (Vera-Cabrera *et al*., 2001). These findings indicate that this domain is a preferential integration region, defined as a chromosomal domain of <500 bp, where several points of IS6110 insertion have been identified in different clinical isolates (Warren *et al*., 2000).

A preferential insertion site was also detected in *plcD*. In 10 isolates, insertion of a copy of IS6110 was detected at position 2611, relative to *M. tuberculosis* CDC1551 cosmide MT1799 (GenBank accession no. AE007040). Insertions at position 2611 were reported by other authors in clones IS540.9 and ISL0480.11 (GenBank accession nos AF126473 and AF077945, respectively) (Sampson *et al*., 1999), confirming that this site represents a hot spot for integration. Ho *et al*., (2000) studied 22 non-related clinical isolates and found 18 discrete IS6110 insertion sites in *plcD* and adjacent genes. Lari *et al*., (2001) detected an IS6110 insertion in *plcD* in one isolate and deletions of the entire RvD2 region in 15 out of 45 clinical *M. tuberculosis* isolates.

Detailed analysis of six genetically related isolates disclosed a genetic mechanism of deletion of a fragment located between two IS6110 elements. Deletion was accompanied by loss of the characteristic three- or four-nucleotide direct repeat flanking IS6110 elements. The observed deletion is most likely explained by the homologous recombination mechanism described by Fang *et al*., (1999), and this explanation agrees with findings of other authors (Ho *et al*., 2000).

Fang *et al*., (1999) reported deletions in a preferential locus for IS6110 insertion, named *ipl*, in clinical *M. tuberculosis*
isolates, and attributed these deletions to homologous recombination between two IS6110 elements. This hypothesis was based on the assumption that a recent related ancestor had two IS6110 insertions, with the same orientation, in this locus. Homologous recombination of these two IS6110 sequences would have led to deletion of the DNA segment between them. To our knowledge, in this study, for the first time this hypothesis was substantiated by the finding of clinical isolates that are genetically related as determined by IS6110 typing and that could represent evolutionary intermediates.

According to Sampson et al. (2003), preferential integration of IS6110 elements apparently triggers IS6110-mediated deletion events, as observed in the DR region and in a 20 kb hypervariable region (Ho et al., 2000). Admitting that plc and PPE genes are preferential insertion regions (Sampson et al., 1999; Warren et al., 2000), it could be speculated that recombination between two insertion sequences, one in a plc gene and the other in the vicinity of a neighbouring PPE gene, may have resulted in additional IS6110-mediated deletion events in plc genes, identified in this work.

Proximity and relative orientation of IS6110 elements are expected to influence recombination events within preferential integration regions, but other as-yet-unknown factors, such as IS6110 transposition activity, may also play a role.

Analysis of polymorphisms in the plcD–cutinase region of H37 strains provides some insights. The genomes of M. tuberculosis H37Rv, M. tuberculosis H37Ra and M. tuberculosis CDC1551 contain a copy of IS6110 in front of the cutinase gene, in the same position and with the same orientation (Fig. 4). In H37Ra, a second copy of IS6110 is present in plcD. In M. tuberculosis H37Rv, a 7–9 kb fragment comprising the segment between positions 2856 in plcD and the cutinase gene was deleted, leaving one copy of IS6110. The absence of flanking direct repeat elements strongly suggests that homologous recombination between two copies of IS6110 caused the deletion event (fragment RvD2). Therefore, the presence of two insertion elements, one in plcD and the other in the cutinase gene, triggered deletion in H37Rv but not in H37Ra, suggesting that IS6110 activity is in some way more pronounced in Rv than in Ra. Differences in IS6110 transposition activity between these two strains were confirmed by analysis of IS6110-RFLP patterns from 18 ATCC H37 variants (Bifani et al., 2000). Nine distinct but similar patterns were detected in 15 H37Rv isolates and a single pattern was detected in three H37Ra isolates.

In five isolates, insertions and/or deletions occurred simultaneously in all four plc genes. Even with extensive deletions in plcABC genes, a truncated fragment of the plcC gene was still detected by Southern blot hybridization. Transcription of this truncated plcC gene was not detected by RT-PCR. Since this is the first description of gene interruptions affecting expression of all members in the same gene family, implications on the fitness of these organisms are unknown at this time. Nevertheless, the fact that these strains were recovered from patients with active tuberculosis demonstrates that they retain the ability to cause disease.

Despite the mobile quality of the IS6110 element, its instability is not sufficiently high to support short-term variations in the plc genes. It is also very unlikely that the variability in the plc genes could be influenced by culture procedures. First of all, cultures are not frequently recultured after primary isolation. Strains with the same genotypes often have the same plc gene composition. In fact, the half-life of IS6110 RFLP is between 3 and 5 years (De Boer et al., 1999). This includes transpositions in all genomic regions. Alterations of the plc genes due to transposition of IS6110 would be much less frequent.

Insertions and deletions were identified in 84% of the M. tuberculosis isolates studied and in one M. africanum isolate, with effects ranging from disruption of individual open reading frames to loss of complete genes. Insertions in nearby locations indicate that these genomic regions may be particularly susceptible to insertion events. These findings suggest that insertions of IS6110 in plc loci result in a biological advantage for the bacteria, which is maintained during subsequent strain diversification. Deletions may be subjected to positive selection when the bacteria no longer require the removed genes. The question whether interruptions of these genes by insertion or deletion may have important effects on the biological properties of clinical isolates still deserves an answer. Differences in the virulence and phospholipase C activity of these isolates are currently under evaluation.

ACKNOWLEDGEMENTS

This work was supported by Centro Argentino Brasileiro de Biotecnologia (CABBIO) grant 480382/01-8 from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), and European Commission RDG (INCO-DEV Programme), project No. ICA4-CT-2001-10087. C.V.-N. is the recipient of a fellowship from FAPESP/00/02525-3), D.S. and S.C.L. are members of RELACTB (Tuberculosis Network for Latin America and The Caribbean).

REFERENCES


