MOLECULAR CHARACTERISATION AND DIAGNOSIS

Molecular cloning of a gene (polA) coding for an unusual DNA polymerase I from Treponema pallidum

BERTA RODES*, HSI LIU, STEVEN JOHNSON, ROBERT GEORGE and BRET STEINER

Centers for Disease Control and Prevention, 1600 Clifton Road, Atlanta, GA 30333, USA

The gene coding for the DNA polymerase I from Treponema pallidum, Nichols strain, was cloned and sequenced. Depending on which of the two alternative initiation codons was used, the protein was either 997 or 1015 amino acids long and the predicted protein had a molecular mass of either 112 or 114 kDa. Sequence comparisons with other polA genes showed that all three domains expected in the DNA polymerase I class of enzymes were present in the protein (5′–3′ exonuclease, 3′–5′ exonuclease and polymerase domains). Additionally, there were four unique insertions of 20–30 amino acids each, not seen in other DNA polymerase I enzymes. Two of the inserts were near the boundary of the two exonuclease domains and the other two interrupted the 3′–5′ exonuclease domain which is involved in proofreading. The predicted amino-acid sequence had an exceptionally high content of cysteine (2.4% compared with <0.05% for most other sequenced DNA polymerase I enzymes). The polA gene was further cloned into pProEX™HTa for expression and purification. The transformants expressed a protein of 115 kDa. Antibodies raised against synthetic peptide fragments of the putative DNA polymerase I recognised the 115-kDa band in Western blot analysis. No DNA synthesis activity could be demonstrated on a primed single-stranded template. Although significant quantities of the protein were produced in the host Escherichia coli carrying the plasmid, it was not capable of complementing a polA- mutant in the replication of a polA-dependent plasmid.

Introduction

Treponema pallidum is an obligate human pathogen with no known natural alternate hosts. Numerous attempts to maintain T. pallidum continuously outside an animal host have been unsuccessful; only in the presence of mammalian cells in tissue culture have a limited number of divisions been demonstrated [1]. Because of the limited ability to manipulate this organism biochemically or genetically, little is known about its physiology.

Previous in-vitro studies demonstrated that T. pallidum was killed rapidly by reactive oxygen species [2]; loss of viability could be delayed but not prevented by the presence of reducing agents in the medium [3]. Neither superoxide dismutase nor catalase activity is present in T. pallidum [2, 4], and this is probably a major factor in its sensitivity to oxygen reduction products. The bacterial chromosome is especially vulnerable to damage caused by reactive oxygen species and can suffer a number of different modifications as well as strand breakage when exposed to these agents [5]. Previous experiments indicated that T. pallidum repaired oxidative damage to its DNA far less thoroughly and much more slowly than a repair-proficient strain of Escherichia coli. This failure to repair DNA damage appeared to correlate with a loss of viability in vitro [6]. Slow and incomplete repair of oxidative lesions may increase double-stranded breaks in the chromosomal DNA, which are highly lethal [7] to a slow-growing bacterium such as T. pallidum [8]. Repair of double-stranded breaks depends on the presence of multiple genome copies [9, 10]. T. pallidum, with a 30-h generation time, would be expected to have little or no genome redundancy and would be expected to have little capacity to repair this type of lesion. The basis for the apparent incomplete repair of oxidative lesions in the genomic DNA is presently unknown. The recently published genome sequence of T. pallidum [11] demonstrated the presence

Received 23 Aug. 1999; accepted 17 Nov. 1999.
Corresponding author: Dr B. Steiner.
*Present address: Infectious Diseases Service, Hospital Carlos III, Madrid, Spain.
of some of the genes involved in DNA repair, but their activity and stability are unknown.

This report describes the cloning and sequencing of the gene for DNA polymerase I of *T. pallidum*, an enzyme of major importance in DNA repair and replication. As this enzyme is responsible for gap re-synthesis in DNA repair, its structure and activity could be of major importance in understanding the above-noted defects in DNA repair. The sequence proved to have a number of unique features which are described here.

**Materials and methods**

**Bacterial strains and plasmids**

Genomic DNA used in constructing the various DNA libraries was obtained from the Nichols strain of *T. pallidum*. Plasmids used in the original cloning and expression of the gene, and recombinant plasmids containing partial or complete sequences of the *polA* gene are given in Table 1; the various recombinants are discussed in more detail in the Results section.

**Sequencing of the *polA* gene from *T. pallidum***

The *polA* gene was originally detected as a partial sequence on a recombinant plasmid, designated Tp615, during immunological screening of a *CiaI* library of *T. pallidum* DNA cloned into the *AccI* site of the plasmid pUC8. This recombinant contained 537 codons of the *polA* gene including most of the highly conserved polymerase domain. A DNA probe of 1594 bp was generated by PCR from the sequence of Tp615 (bp 1701–3295 in the reported sequence), labelled with digoxigenin (Boehringer Mannheim, Indianapolis, IN, USA) and used to screen a second library of treponemal DNA generated with the restriction endonuclease *PstI*. A second recombinant from this library (referred to as *PstI-polA*) was isolated by DNA colony blotting; it contained a further 300 codons from the 5' end of the gene. The sequence of the gene was completed by chromosome walking by the single-specific primer PCR method [13]. The sequencing of the original *CiaI* clone (Tp615) was performed by the method of Sanger et al. [14] with the United States Biochemical (USB, Cleveland, OH, USA) sequencing kit with α-S32P dATP and Sequenase 2.0. All subsequent sequencing (including confirmation of the original Tp615 clone) was performed with the automated ABI PRISM Dye Terminator cycle sequencing Core Kit with ampliTaq DNA polymerase (ABI/Perkin-Elmer Cetus, Foster City, CA, USA). After the gene was cloned from *T. pallidum* genomic DNA by PCR, the intact gene was re-sequenced to confirm the sequence obtained by the strategy given above. Primers for sequencing and PCR were either produced by the Biotechnology Core Facility at CDC or were purchased from Integrated DNA Technologies (Coralville, IA, USA).

**Cloning and expression of the *polA* gene product (DNA polymerase I) from *T. pallidum***

The *polA* gene was cloned from *T. pallidum* genomic DNA by PCR amplification. The known sequence of the gene and surrounding DNA were used to construct primers with artificial restriction sites for cloning the

<table>
<thead>
<tr>
<th>Strains or plasmids</th>
<th>Relevant markers</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. pallidum Nichols strain culture</td>
<td>Not applicable</td>
<td>Laboratory collection</td>
</tr>
<tr>
<td>E. coli DH5α</td>
<td>F−, f008lacZΔM15lacAYZ-argF) U169, deoR, recA1, endA1, adenR, phoA supE44lac−, thi−, gyrA96, relA1</td>
<td>BRL/Gibco</td>
</tr>
<tr>
<td>E. coli DH10B</td>
<td>F− metE ΔnusAΔRMS::metE(t) f008lacZΔM15ΔlacAYZ-argF) deoR recA1 endA1 araD139 Δ(arab, leu)7697 galU galK1::tacN</td>
<td>BRL/Gibco</td>
</tr>
<tr>
<td>E. coli SF800</td>
<td>pol41, gyrA</td>
<td>S. Johnson, CDC</td>
</tr>
<tr>
<td>Plasmids</td>
<td>Pen−, Lac+</td>
<td>Bannisch-Perron [12]</td>
</tr>
<tr>
<td>pUC8</td>
<td>Pen−, replicates independent of DNA polymerase I</td>
<td>S. Johnson, CDC</td>
</tr>
<tr>
<td>pUC8 CiaI recombinant; containing partial polA sequence from T. pallidum</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>pUC8 PstI recombinant; containing partial polA sequence from T. pallidum</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>Tp615</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>Tp61551</td>
<td>This study</td>
<td></td>
</tr>
<tr>
<td>Tp615605</td>
<td>This study</td>
<td></td>
</tr>
</tbody>
</table>

Pen− refers to resistance to penicillin, Lac+ for the plasmids, refers to a-complement DH5α.
gene in frame with the β-galactosidase gene promoter in the expression vector pProEX™HTa (BRL/Gibco, Gaithersburg, MD, USA). As there were two potential initiation codons, two different primers were used at the 5′ end of the gene in separate amplifications. The sequence of the primers was as follows: 5′ primer, ATG start (bp 605 in reported sequence), CCCGAATT CATGCAAGAGAAAGAACGC; GG start (bp 551 in reported sequence), CCCGAATTCGCTGCTGACA TACCCATTC; the primer at the 3′ end of the gene was the same for both amplifications, CCCAGCTTC TAATGAAAATCACC. The primers for the 5′ end of amplicons contained sites for EcoRI; the single primer used for the 3′ end of the amplicons contained a restriction site for HindIII. Amplifications were run for 30 rounds and the amplicons were purified (High Pure PCR Product Purification kit, Boehringer Mannheim) and cut with the appropriate enzymes; the fragments were then ligated into the pProEX™HTa plasmid cut with EcoRI and HindIII. The recombinant plasmids were then transformed into E. coli strain DH10b and bacteria containing recombinant plasmids were isolated by standard methods [15]. Isolated colonies of E. coli strain DH10b containing recombinant plasmids were selected and grown to a density of 0.6–1.0 at an OD600 in Luria broth with penicillin 100 μg/ml; expression of the cloned gene was induced with 0.6 mM isopropyl-β-D-thiogalactopyranoside (IPTG). Whole-cell lysates were prepared from samples taken at 1, 2 and 4 h after induction and the whole-cell proteins were separated by SDS-PAGE on polyacrylamide 7.5% gels. Proteins were transferred from the gel to nitrocellulose membranes by the method of Towbin et al. [16] and the polA gene product was detected with a Ni-NTA HRP (horseradish peroxidase) conjugate which interacts with the hexahistidine leader of the expression vector (Qiagen, Chatsworth, CA, USA).

**Purification of the cloned DNA polymerase I by affinity chromatography**

The recombinant protein from TpPolA 605 clone (ATG start, Table 1) was purified by affinity chromatography. This recombinant E. coli strain (DH10b) contained the polA gene cloned into the pProEX™HTa expression vector; fusion proteins in this vector contain a hexahistidine leader. The recombinant bacteria were induced to express the cloned polA gene with IPTG as described above. After the cells had been induced for 2 h, they were sonicated and the soluble fraction was used to isolate the recombinant protein by affinity chromatography with a Ni-NTA spin column which binds the hexahistidine leader of the recombinant protein with high affinity (Qiagen). The wash buffer contained 10 mM imidazole and 10 mM diithiothreitol (DTT). The imidazole was added to prevent non-target protein binding to the column and the DTT was added to stabilise the protein, as it contained a large number of cysteine residues (see below). The bound fusion protein was eluted from the column by increasing the imidazole concentration to 250 mM, as described in the manufacturer's instructions. The column flow through and imidazole-eluted fractions were dialysed, concentrated and analysed by Coomassie Blue staining of SDS-PAGE 7.5% gels.

**Detection of polA gene product by antibodies against internal peptide sequences of the protein**

Three peptides with sequences unique to the DNA polymerase I of *T. pallidum* were used to produce antibodies in New Zealand White rabbits. The peptides were produced at the BioTechnology Core Facility at CDC. The sequence of the three peptides and their position in the amino-acid sequence of the *T. pallidum* DNA polymerase I are as follows: peptide 1, -Q-E-I-D-T-E-A-T-N-D-T-L-Q-M-T-E-S-S-V-L- (amino acid 310–329); peptide 2, -S-O-V-E-G-R-A-S-T-P-E-V-N-S-V-L-K-S-E-L- (amino acid 341–360); peptide 3, -V-L-K-S-E-L-K-T-S-A-V-S-G-A-I-P-I-E-N-R- (amino acid 355–374). The positions are given from the putative GTG start at bp 551; numbering from the alternate ATG start at bp 605 would result in a numbering sequence 18 amino acids shorter (see below for a discussion of the two alternative initiation codons). All these peptides come from sequences within the inserts found in the 3′–3′ exonuclease domains that have no homologues in other sequenced DNA polymerase I enzymes (see Results). The three peptides were synthesised with a carboxyl terminal cysteine and were coupled through this cysteine to either bovine serum albumin (BSA) or keyhole limpet haemocyanin (KLH). Each conjugated peptide was then mixed with an equal volume of Hunter’s Titermax adjuvant (CytRx, Norcross, GA, USA) and injected into two rabbits (150–500 μg of peptide per rabbit). The rabbits were given booster injections of the same amount of antigen 3–4 weeks later. Individual rabbits were bled before injection, at boosting, and 6 weeks after initial infection, and the serum was collected.

**Reverse transcriptase (PCR RT-PCR) for determination of the presence of polA mRNA**

RT-PCR was used to determine whether the gene for DNA polymerase I was transcribed in *T. pallidum*. mRNA was isolated from freshly harvested *T. pallidum* with the Qiagen RNA-easy midi kit. Reverse transcriptase was used to produce two cDNA products spanning the polymerase domain and the combined exonuclease domains. This was followed by PCR amplifications of the cDNA templates to produce amplicons from the polymerase domain (346 bp) and parts of both the 5′–3′ and 3′–5′ exonuclease domains (1025 bp). The primers used for the amplifications were as follows: reverse transcriptase primer for the polymerase domain, TCC ACG TGT ATA CGC AGC GG (location – bp 3567–3548 opposite to sense strand); reverse transcriptase primer for the exonuclease domains, CAT CGA GGA TAC TAC AAC AGG
(location – bp 2202–2182 opposite to sense strand); PCR primers for amplification of cDNA (polymerase domain), GAC GTA TAC CGA ATC GCT TGC (location – bp 2728–2748 on sense strand), TTG CGA TTC TTC TTA CAG G (location – bp 3083–3062 opposite to sense strand); PCR primers for amplification of the exonuclease domains, CTC AGA TAA AGA TGT ACT TCA GC (location – bp 1012–1034 on sense strand) and GCT GGT GCA TGA CAG CTT GG (location – bp 2036–2017 opposite to sense strand). The reverse transcriptase reaction was performed with a Qiagen Omniscript reverse transcriptase kit. The reaction conditions were as follows: primer concentrations 50 μM, RNAase inhibitor (Boehringer Mannheim) 10 units/ml, Omniscript reverse transcriptase 4 units/ml, 5 mM of each dNTP, 1 μg of isolated mRNA for template, in RT buffer. The reaction was allowed to run for 60 min at 37°C. The PCR amplification of the cDNA product was performed with 1 mM of each primer, Expand polymerase (Boehringer Mannheim) 2 units/reaction, 5 mM of each dNTP and 5 μl of the cDNA template. The PCR reaction was performed with an ABI/PE 9700 thermocycler (Perkin-Elmer Cetus). The PCR reaction was performed for 45 cycles under the following conditions: 94°C for 1 min, 57°C for 1 min and 72°C for 1 min. The resulting amplicons were analysed by electrophoresis on agarose 1% gels and stained with ethidium bromide. Control reactions were run without the addition of reverse transcriptase; the controls were uniformly negative and demonstrated that the mRNA preparation contained no contaminating genomic DNA.

Functional assays for DNA polymerase I activity

Functionality of the DNA polymerase I from *T. pallidum* was examined by determining the ability to synthesise a polA-dependent plasmid in a polA- background and by direct assay for polymerase activity. DNA polymerase I is necessary for the replication of col E1-related plasmids in *E. coli* [17]. The polA gene cloned into the plasmid pProEX<sup>T</sup>HTa was introduced into *E. coli* strain SF800 (polA<sup>1</sup>, gyrA<sup>1</sup>), which does not produce DNA polymerase I. Expression of the cloned gene was induced with IPTG and 100 μl of transformants containing either the cloned polA gene or the polA-independent control plasmid pC8023 [18] were plated on to ampicillin-containing medium. Also, the fusion protein from pProEX<sup>T</sup>HTa was purified by affinity chromatography on a Ni-NTA spin column and the purified protein was assayed for polymerase activity on a primed, single-stranded DNA template. Purification of the enzyme was, as indicated above, after induction of expression of the fusion protein with IPTG for 4 h. All solutions used in purification contained 10 mM DTT. The purified fusion protein was dialysed against 50 mM Tris buffer (pH 7.4) containing 5 mM MgCl<sub>2</sub> and 10 mM DTT to remove imidazole and excess salts. Glycerol was added to a final volume of 20% and the purified protein was stored at −20°C until assayed.

The assay used to determine DNA synthesis on a single-stranded primed template was that described elsewhere for random primer labelling of DNA [19]. The activated template was pUC8 DNA cleaved with *Pvu*I and denatured by heat (95°C for 12 min). Ten μl of the crude extract were used for each 40-μl reaction. The reactions were allowed to run for 60 min at 25°C. The reaction mixtures were then chilled on ice and incorporation was terminated by the addition of 20 μl of herring sperm carrier DNA (10 mg/ml) and ice-cold 10% trichloroacetic acid to a final concentration of 5% w/v. The precipitate was collected on to glass fibre filters and the amount of radioactivity incorporated was determined by liquid scintillation counting. The Klenow fragment of DNA polymerase I from *E. coli* (1 unit) was used in the same reaction volume as a positive control.

Computer methods and GenBank accession number

The sequence for the polA gene and surrounding DNA has been assigned the GenBank accession number TPU577757. The sequence reported here was confirmed by Fraser et al. [11] as part of the sequencing of the total genome of *T. pallidum* and was given the designation Tp105. Comparisons with known gene sequences were made with the GCG (Genetics Computer Group) package from the University of Wisconsin and the MacVector software package (version 5.0) of Eastman Kodak. Sequence alignments were made with the Gap and Bestfit programs in the GCG package. For phylogenetic analysis, the amino-acid sequences of the various DNA polymerase I enzymes were aligned with Clustal W [20]. Phylogenetic trees were constructed by the neighbour-joining method with the PHYLIP software package [21]. Bootstrap resampling (1000 data sets) of the multiple alignment was done to test the statistical robustness of the tree.

**Results**

**Sequence analysis of the polA gene from *T. pallidum***

A 3885-bp fragment of *T. pallidum* genomic DNA was sequenced, which contained the complete polA gene and the terminal 100 codons of an upstream gene which had strong sequence homology to a putative 5′-nucleotidase gene of *Haemophilus influenzae* [22]. A computer-assisted search identified an open reading frame that encoded a protein with significant homology to the DNA polymerase I of *E. coli*. Sequence similarity was 44.8% for the entire reading frame. The carboxyl terminal of the protein (polymerase domain) had the greatest similarity to the *E. coli* enzyme at 56.6%. The gene contained two potential
initiation codons – an ATG at 605 bp in the reported sequence (GeneBank accession number TPUC577757) and a GTG at bp 551. The ATG codon was not preceded by an identifiable ribosomal binding site (RBS). In contrast, the sequence preceding the GTG start codon contained several suitable RBS sequences within 5–13 bp of the start codon [23]. Because of this characteristic, the GTG was considered to be the putative start of translation. Both translational start sites would produce a protein that contained the three complete polymerase I domains.

The complete sequence of the treponemal polA gene indicated that it coded for a protein with a number of unique characteristics not found in other bacterial DNA polymerase I enzymes. A comparison of the predicted amino-acid sequence of the protein with DNA polymerase I enzymes from six different bacterial species (Fig. 1) revealed two unusual features: an abnormally high content of cysteine and four apparent inserts in the 3′–5′ exonuclease domain with no homologues in other sequenced polA genes (amino acids R306–F339, E359–G387, L444–S463 and V489–H510). The high percentage of cysteine residues was unique to DNA polymerase I and was not seen in the predicted coding sequence of other DNA repair enzymes from T. pallidum. For example, when the predicted coding sequences for recA and uvrA from T. pallidum [11] were compared with the same genes from several other bacteria, no excess of cysteine was seen in either enzyme (data not shown).

The four inserts accounted for the large size of the DNA polymerase I as compared with other enzymes of this class, and the low level of similarity found in the 3′–5′ exonuclease domain compared with the other two domains (33.6%). The unique character of the amino-acid sequence for the DNA polymerase from T. pallidum could be seen when it was compared with the sequences for this enzyme from various other micro-organisms. A phylogenetic tree compiled with the PHYLIP software package showed only a distant relationship to the sequence from Borrelia burgdorferi, the only micro-organism with a sequence on the same branch of the tree (data not shown).

Expression of the polA gene product

To characterise the protein further, the entire polA gene was amplified from whole T. pallidum genomic DNA by PCR. Synthetic restriction sites were introduced as indicated in Materials and methods and the amplicon was inserted into the ProEXHTa expression vector. The resulting recombinant plasmid contained the complete polA gene cloned in frame with the Trc promoter. Constructs were made from both the possible starts of translation (ATG codon at position 605 and GTG codon at position 551). Transformants containing plasmids of the proper size were sequenced to confirm the presence of the complete polA gene. Recombinant plasmids containing the cloned polA gene were then induced with IPTG, and the protein product was detected on Western blots with a Ni-NTA-HRP conjugate which binds to the vector-derived polyhistidine leader. Protein products of the expected sizes were produced from both the ATG and GTG starts (113 kDa and 115 kDa, respectively) (Fig. 2). The appearance of these new protein bands of the expected size after induction suggested that they were the products of the polA gene. Substantially more fusion protein was produced from Tp605 (the ATG start) than from Tp551 (the GTG start). At present, the cause of this is unknown. The total protein obtained after induction was further purified with a Ni-NTA column. The eluted protein had a single band which corresponded to the predicted molecular size of 115 kDa (Fig. 3).

Confirmation of the expression of the polA gene by antibodies to internal peptides in the protein sequence

To prove that the putative DNA polymerase I from T. pallidum was being expressed as a fusion protein in pProEXHTa, antiserum was produced against three synthetic peptides whose sequences were derived from the predicted T. pallidum DNA polymerase I sequence. These peptides were derived from regions of the protein that shared no similarity with the enzyme from E. coli. These antiserum would confirm the expressed protein as being the T. pallidum DNA polymerase I, and could be used for identification of the protein expressed from any vector in which the gene was cloned. The three antiserum showed reactivity with the same 115-kDa protein as the Ni-NTA HRP conjugate (clone 551–2) to the polyhistidine leader of the cloned protein (Fig. 4); the best reactivity was seen with peptide 1. In all cases, the antipeptide antiserum produced minimal background staining with non-target proteins. The results were identical whether BSA or KLH was used as the carrier protein for the synthetic peptides. The antipeptide antiserum did not react with the DNA polymerase I from E. coli, indicating that the protein produced by the recombinant plasmid was the T. pallidum polA gene product.

Transcription of the gene and functional assays for DNA polymerase I activity

The study was unable to demonstrate the presence of the protein either in total protein extracts from T. pallidum on Western blots probed with antipeptide antiserum or with hyperimmune serum from T. pallidum-infected rabbits (data not shown). To determine whether this gene is being transcribed, RT-PCR analyses of total treponemal RNA were performed. Two pairs of primers were designed to cover the polymerase domain and combined exonuclease domains. Amplicons of predicted sizes were observed in both cases (Fig. 5).
Fig. 1. Comparison of the amino-acid sequence of the putative DNA polymerase I from *T. pallidum* with those of several sequenced *T. pallidum* proteins. DNA polymerase I amino-acid sequences used in the comparison are from *E. coli* (GenBank accession number [P05852]), *B. burgdorferi* (BB0548), *Thermus aquaticus* (TTHA0191), *Deinococcus radiodurans* (P52027), *Mycobacterium tuberculosis* (2004291B) and *Streptococcus pneumoniae* (A32494).

Conserved motifs in the three domains of the enzyme are indicated by grey boxes and are as follows: (a) motifs A–F are conserved motifs in the 5′–3′ exo nuclease domain (amino acids 1–299 in *E. coli*); (b) motifs I–III are conserved motifs in the 3′–5′ exo nuclease domain (amino acids 299–521 in *E. coli*); (c) motifs indicated 1–5 are in the polymerase domain (amino acids 522–928 in *E. coli*). Note the underlined insertions in the 3′–5′ exo nuclease domain of the *T. pallidum* sequence.
Fig. 1. (Continued).

Fig. 2. Expression of the T. pallidum polA gene in pProEX™HTa from the alternate initiation codons at position 551 and 605. Lanes 1–3, expression of the TpPolA 551 clone (GTG start) at 0, 2 and 4 h after induction with IPTG; 4, expression of the TpPolA 605 clone (ATG start) at 2 h after induction; 5, E. coli DH10B without plasmid. Arrow shows the expressed T. pallidum DNA polymerase I protein (115 kDa).

The fusion protein from the polA gene cloned into pProEX™HTa with the GTG start (clone 551–2) was used to determine whether the gene product was active in the replication of DNA polymerase I-dependent plasmids. This experiment was performed by transforming clone 551–2 into a polA− strain of E. coli (SF800). No ampicillin-resistant recombinants were obtained in strain SF800, whereas transformants were readily obtained in E. coli DH10B, which is polA+ (data not shown). The competency control for strain
DNA polymerase I, the first DNA polymerase to be isolated and purified, is now known to be essential in DNA repair in most bacteria [24]. In *E. coli*, DNA polymerase I has three enzymic functions which are arranged in the following order from the amino terminus of the protein: (a) a 5′–3′ exonuclease which appears to be of central importance in the DNA repair function of the enzyme; (b) a 3′–5′ exonuclease which is the proof-reader and imparts the high fidelity seen in the enzyme; and (c) the polymerase domain itself [22]. This domain structure is highly conserved [25], the only major exception being the absence of the 3′–5′ proof-reading exonuclease activity in some bacteria [26, 27].

The DNA polymerase I homologue of *T. pallidum* contains a number of unique features. The first unique feature of the amino-acid sequence was the high content of the amino-acid cysteine. In *E. coli* proteins, cysteine is the rarest of all the amino acids [28]. This amino acid is also scarce in the majority of DNA polymerase I proteins for which a sequence is known. Only a few of the putative pol I enzymes identified by sequencing show more than three cysteine residues in a protein composed of c. 900 residues. In contrast, the *T. pallidum* sequence indicated the presence of 24 cysteine residues. Of sequences presently available, only a few other obligate parasites have higher than expected levels of cysteine. These include the DNA polymerase I of *Chlamydia trachomatis*, which has 12 cysteine residues [29]. As cysteine is sensitive to the redox state of the micro-organism, the high content of this amino acid may be related to the low steady-state oxygen tension seen in mammalian tissues [30].

Most of the cysteine residues in the DNA polymerase I of *T. pallidum* were located in the two exonuclease domains. Also, two cysteine residues were found in the conserved 3′–5′ exonuclease sequences I and III (Fig. 1). This feature is unique to the DNA polymerase I from *T. pallidum* and is not a general feature of the enzyme from other spirochaetes, as the sequence for the DNA polymerase I from *B. burgdorferi* contains no cysteine residues [31]. This laboratory has developed primers derived from two of these areas where multiple cysteine residues are present and has shown that they do not produce an ampiclon in various other spirochaetes. Preliminary work has shown that these primers may be readily exploited for the development of a PCR-based diagnostic test for syphilis (unpublished observations).

The second unusual feature of the polymerase I sequence was the presence of four unique insertions that code for 15–20 amino acids in the 3′–5′ exonuclease domain. The only DNA polymerase I of similar length to the enzyme from *T. pallidum* is from *Rhizobium leguminosarum* [32]. The *R. leguminosarum* enzyme is almost identical in length to the *T. pallidum* enzyme. The most interesting feature of the two sequences is that the insertions are

**Discussion**

The study also attempted to determine the ability of the DNA polymerase I from clone 551-2 to perform DNA synthesis on a primed, single-stranded DNA template. Purified pol I protein did not show significant incorporation of 32P into trichloroacetic acid-insoluble material above background, whereas the Klenow fragment of DNA polymerase I from *E. coli* had significantly activity (data not shown).
Fig. 4. Detection of the putative *T. pallidum* DNA polymerase I fusion protein with antibodies to synthetic peptides derived from the sequence of the *polA* gene. Lanes 1 and 2, peptide 1: 1, pre-bleed; 2, 4 weeks after immunisation; 3 and 4, peptide 2 (first immunised rabbit): 3, pre-bleed, 4, 4 weeks after immunisation; 5 and 6, peptide 3 (first immunised rabbit): 5, pre-bleed, 6, 4 weeks after immunisation; 7 and 8, peptide 3 (second immunised rabbit): 7, pre-bleed, 8, 4 weeks after immunisation; 9 and 10, peptide 2 (second immunised rabbit): 9, pre-bleed, 10, 4 weeks after immunisation; 11 and 12, Ni-NTA-HP to the hexahistidine leader of the fusion protein. Arrow indicates the 115-kDa fusion protein. At least one of the two rabbits immunised against each peptide showed a strong reaction with the putative DNA polymerase I; no other band appeared after immunisation in more than one rabbit. The sequence of the three synthetic peptides is given in Materials and methods.

almost at the same location; despite the fact that the two polymerases are not particularly related by primary sequence comparison (data not shown). The significance of these insertions is not known at present.

A more detailed study of the enzyme is necessary to determine whether the inserts or the high cysteine content are related to the apparent instability of the enzyme when cloned into *E. coli*. Construction of a phylogenetic tree based on the amino-acid sequence of the enzyme showed that it was most closely related to the gene from the only other spirochaete for which data are available, *B. burgdorferi* [31]. Despite this fact, the phylogenetic tree for the two organisms is deeply forked. Sequences from other spirochaetes are needed to determine whether the *polA* gene of *T. pallidum* is truly unique in this aspect.

A Southern blot of genomic DNA, digested with either *SnaI* or *PstI*, and probed with a 1594-bp digoxigenin-labelled probe (bp 1150–2744 in the reported sequence), indicated that only a single copy of the gene was present in the *T. pallidum* chromosome (data not shown). This is in agreement with the genome sequence of *T. pallidum*, which indicated the presence of a single gene [11]. The present study was able to show by RT-PCR that this gene is transcribed in *T. pallidum* (Fig. 5). This is unlike the situation with another pathogen, *Mycobacterium leprae*, which has resisted cultivation. It has been suggested that the *polA* gene in this organism is not transcribed and does not contain an active promoter [33, 34]. Despite evidence for transcription of the *polA* gene in *T. pallidum*, the present study was unable to detect the enzyme in *T. pallidum* extracts by immunological means. This result suggests that the amount of enzyme is extremely low. Future studies on the role of this enzyme in DNA repair and replication require the determination of suitable vectors or conditions, or both, that will produce an active enzyme.

These data were presented in part at the 13th Meeting of the International Society for Sexually Transmitted Diseases, Abstract no. 119, July 11–14, 1999. We thank Donna Cruse for excellent technical assistance in the immunological screening of the *T. pallidum* library used in this work.
Fig. 5. Demonstration of the presence of mRNA for DNA polymerase I by RT-PCR. MW, 100-bp ladder. Lane 1, the 346-bp amplicon derived from the polymerase domain; 2, 102.5-bp amplicon spanning the two exonuclease domains.

References


