Molecular characterization and analysis of a gene encoding the acidic repeat protein (Arp) of Treponema pallidum

Hsi Liu,1 Berta Rodes,1,2 Robert George1 and Bret Steiner1

1National Center for HIV/AIDS, Viral Hepatitis, STD and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA 30333, USA
2Hospital Carlos III, Madrid, Spain

The acidic repeat protein (arp) genes from three subspecies of the treponeme Treponema pallidum (T. pallidum subsp. pallidum, Nichols strain; T. pallidum subsp. pertenue, CDC-1 and CDC-2 strains; and T. pallidum subsp. endemicum, Bosnia A strain) were cloned and sequenced. The predicted protein sequence contained a high percentage of glutamic acid, hence the name acidic repeat protein, or Arp. The protein had a potential membrane-spanning domain and a signal peptidase I site. The gene from the Nichols strain of T. pallidum subsp. pallidum contained a set of 14 nearly identical repeats of a 60 bp sequence, which occupied ~51 % of the length of the gene. Analyses of arp from laboratory strains showed that the 5′ and 3′ ends of the genes were conserved, but there was considerable heterogeneity in the number of repeats of this 60 bp sequence. Based on amino acid variations, the 14 sequence repeats could be classified into three types, which were named type I, type II and type III repeats. The type II repeat was the most common in the strains examined. The arp gene of the Nichols strain was subsequently cloned into the expression vector pBAD/TOPO ThioFusion. The expressed protein was detected in a Western blot assay using rabbit immune sera produced against T. pallidum, or synthetic peptides derived from the repeat sequences. Using an ELISA, rapid plasma reagin (RPR) test-positive sera reacted with synthetic peptides derived from the repeat region but not with peptides derived from N and C termini of the Arp protein. These results show that the Arp protein is immunogenic and could prove to be a useful target for serological diagnosis of T. pallidum infection.

INTRODUCTION

Treponema pallidum subsp. pallidum causes venereal syphilis, a chronic disease that has a relapsing course with distinct clinical stages (Stamm, 1999; Liu, 2004). Although T. pallidum was first observed by dark-field microscopy almost 100 years ago, the bacterium still cannot be cultivated and maintained on artificial media (Cox et al., 2003). Recently, repeat regions in the genome have become a focus of study. The repeats are either short tandem repeats of 2–4 nt, called microsatellites, or longer repeats encoding 20 amino acids, called minisatellites (Kajava, 2001; Andrade et al., 2001). The repeating units may represent pathogenic mechanisms used by micro-organisms to evade host immune responses. One characteristic of the repeat proteins is that they are highly immunogenic; thus, they may play a role in the pathogenesis of the disease and could also be potential vaccine candidates.

The majority of the T. pallidum proteins characterized to date are lipoproteins without repetitive sequences (Chamberlain et al., 1989; Radolf, 1995; Fraser et al., 1998). However, several proteins with repetitive sequences have been reported, including a leucine-rich repeat protein in the cytoplasmic membrane (Shevchenko et al., 1997). This protein is not particularly immunogenic. The TmpA and TmpB proteins also contain a small number of repeats and have been shown to be immunogenic (Yelton et al., 1991). The significance of these repeat sequences is unknown, but none of the resultant proteins is known to be involved in pathogenesis, and the immune response to these proteins appears not to be protective (Yelton et al., 1991; Shevchenko et al., 1997).

We recently reported a molecular typing system for T. pallidum that is partially based on the size heterogeneity of a gene that we referred to as the arp (acidic repeat protein)
gene (Pillay et al., 1998, 2002; Sutton et al., 2001; Cox et al., 2003; Pope et al., 2005). PCR amplification of the repeat region of the gene showed variations in the size of amplicons obtained from different isolates; the differences in size were consistently multiples of 60 bp, the size of a single repeat. These results suggested that the number of repeats present in this gene varied among different isolates (Pillay et al., 1998; Cox et al., 2003; Liu, 2004). We report here the characterization of the \textit{arp} gene of the three subspecies of \textit{T. pallidum}.

**METHODS**

**Bacterial strains, plasmids and sera.** The plasmid pUC8 and the Nichols strain of \textit{T. pallidum} subsp. \textit{pallidum} were used to construct a genomic library. Subsequently, the \textit{arp} gene was cloned from one strain each of \textit{T. pallidum} subsp. \textit{pallidum} (Nichols strain, reference strain) and \textit{Treponema endemicum} (Bosnia A, reference strain) and two strains of \textit{Treponema pertenue} (CDC-1 and -2). Nichols strain was isolated in 1912 from the cerebrospinal fluid (CSF) of a patient from the Washington, DC, region. CDC-1 and -2 were isolated in November 1980 from individuals in the villages of Densuso (CDC-1) and Akorabo (CDC-2) in North Africa. Bosnia A was originally obtained from the laboratory of Dr Howard Fieldsteel, SRI International. All strains were propagated in New Zealand White rabbits and maintained at the Centers for Disease Control and Prevention (CDC). Infected rabbits were bled at regular intervals and sera were collected and stored at −20 °C for later use. For sequencing analyses, the \textit{arp} gene was cloned into the Topo TA cloning vector (Invitrogen). For gene expression, the complete \textit{arp} gene was cloned into the pBAD/TOPO ThioFusion expression system (Invitrogen). Manufacturer’s instructions were followed in all cloning processes.

Normal human sera [rapid plasma reagin (RPR) and HIV negative] were obtained from donors at the CDC. Syphilitic sera (RPR positive) were from the Georgia Department of Human Resources, Atlanta, GA, and had no patient identifiers. All sera were stored at −20 °C. Freezing and thawing were kept to a minimum.

**Extraction of treponemes from rabbit testis.** Rabbits were inoculated intratesticularly and treponemes were harvested approximately 10 days (Nichols and Bosnia A) to 21 days (CDC-1 and -2) later. Testes were minced, suspended in 5 ml extraction solution (Cox, 1994) and shaken for 30 min on ice using a shaker (150 r.p.m.). The suspension was centrifuged at 400 g for 5 min at 4 °C and the supernatant containing the treponemes was aspirated and transferred into a test tube. In some cases the process was repeated one more time. The presence and the number of treponemes were determined using dark-field microscopy.

**Screening for the \textit{arp} gene from the \textit{T. pallidum} genomic library.** The genomic library was screened for colonies expressing \textit{T. pallidum} proteins by using dot-blot analyses (Rodes et al., 2000). Transformants were transferred to nitrocellulose filters (0.2 μM pore size) and grown on LB plates containing 50 μg ampicillin ml⁻¹ and 0.2 mM IPTG. Colonies that resulted from overnight growth were lysed by exposure to chloroform vapour and placed in lysis buffer (20 mM Tris/HCl, pH 8.0, containing 10 mM EDTA and 10 mg lysozyme ml⁻¹). The lysis buffer was drained from the membranes after 20 min, and the membranes were incubated in 0.2 % Triton X-100 for 30 min. After blocking, the membranes were probed with rabbit immune sera using standard procedures (Sambrook et al., 1989). The blots were then probed with anti-IgG conjugated with horseradish peroxidase (HRP), washed three times with PBS containing 0.05% Tween 20, and developed with 4-chloro-1-naphthol (Sigma).

Positive colonies were selected following Western blot analysis (Sambrook et al., 1989) of soluble proteins. Specifically, cells from these colonies were grown overnight in 50 ml volumes of LB/ampicillin (Amp) medium and disrupted by sonication. Total cellular proteins from these \textit{Escherichia coli} colonies were electrophoresed on 10 % SDS-polyacrylamide gels and then transferred to nitrocellulose membranes. Detection of treponemal recombinant proteins was performed in the same manner as for dot-blot analyses of recombinant colonies.

**PCR cloning using TOPO TA and pBAD/TOPO ThioFusion.** DNA was prepared from the laboratory strains (Nichols, CDC-1 and -2, and Bosnia A) using the QIAamp DNA mini kit (Qiagen). Primers were designed (Table 1), and PCR reactions were performed using High Fidelity PCR Master mix (Roche Applied Science). The PCR reaction was performed with an ABI 9700 thermocycler. The PCR amplification was performed for 45 cycles under the following conditions: 94 °C for 1 min, 57 °C for 1 min and 68 °C for 1 min. The resulting amplicons were then cloned into the TOPO TA vector. DNA sequencing was performed using dRhodamine chemistry with an ABI 310 genetic analyser. In addition, the gene was cloned into the pBAD/TOPO ThioFusion expression vector. The pBAD system produces a fusion protein of the target protein and thioredoxin to facilitate

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cloning into pUC 8</strong></td>
<td></td>
</tr>
<tr>
<td>Forward primer*</td>
<td>GGC GCC ATG GGA TCC GGA ATT CAA AGG CCT ACG TCG ACG AGC– (ATG TTT GTG CCG AGT GAC ATG)</td>
</tr>
<tr>
<td>Reverse primer†</td>
<td>ACA CGG AAG CTT GTG ACC GCA TGC CTC GAC GAT GCA GCC AGC GCC CGC CCA CAA– (TTT CTG ACT CGG ATA GGT C)</td>
</tr>
<tr>
<td><strong>Cloning into pBAD/TOPO ThioFusion</strong></td>
<td></td>
</tr>
<tr>
<td>Forward primer</td>
<td>ATG TTT GTG CGC AGT GAC ATG TTC CCC AAA</td>
</tr>
<tr>
<td>Reverse primer</td>
<td>TCG GCC CTT TTT CAG AAA GGC GTC CTT GA</td>
</tr>
</tbody>
</table>

*The 5’ end of the primer contains the following multiple cloning sites: \textit{Ehel, NcoI, BamHI, EcoRI, Stud and Sall}. The \textit{arp} primer sequences at the 3’ end are in parentheses.

†This primer contains the following multiple cloning sites: \textit{HindIII, KpnI, Spel, XhoI, PstI and NotI}.
protein folding. A hexahistidine tag near the C terminus of the vector facilitates protein purification using a nickel-nitrilotriacetic acid (Ni-NTA) column. Positive clones were screened by examining colonies which grew on LB/Amp plates, followed by sequencing using dRhodamine chemistry.

**Protein expression.** Manufacturer's instructions were followed throughout the experiment. In brief, an overnight growth of *E. coli* containing the fusion protein vector was inoculated into fresh LB medium. When the OD$_{600}$ reached 0.5, l-arabinose was added to a final concentration of 0.2%. The culture was allowed to grow for another 2–4 h and the reaction was terminated by centrifugation when OD$_{600}$ reached 2.0. The pellet was then stored at −20 °C. Total cell (*E. coli*) lysates were prepared by treating the pellet with 6 M guanidine/HCl followed by sonication. A control culture of *E. coli* only (not containing the vector) was treated similarly. The resulting protein solution was then purified using a Ni-NTA column (Qiagen). The lysates and purified protein were examined using Western blot analysis (see below).

**Rabbit infection and peptide immunization.** Rabbits were infected with *T. pallidum* intratracheally in accordance with the CDC animal protocol. Animals were bled before infection and at specific times after infection.

Three peptides (Arp-5, -6 and -7, Table 2), each 20 amino acids long and comprising overlapping sequences taken from the repeat domain, were synthesized with a C-terminal cysteine and were coupled through this cysteine to either BSA or keyhole limpet haemocyanin (KLH). The peptides were produced at the Biotechnology Core Facility at the CDC. The peptides had at least 90 % purity and showed one major peak when analysed using an analytical HPLC equipped with a C18 column (4.6 × 150 mm; System Gold, Beckman-Coulter). Each conjugated peptide was then mixed with an equal volume of Hunter's TiterMax adjuvant (CytRx) and injected subcutaneously into two rabbits (150–500 μl) (KLH). The peptides were produced at the Biotechnology Core Facility at the CDC. The peptides had at least 90 % purity and showed one major peak when analysed using an analytical HPLC equipped with a C18 column (4.6 × 150 mm; System Gold, Beckman-Coulter). Each conjugated peptide was then mixed with an equal volume of Hunter’s TiterMax adjuvant (CytRx) and injected subcutaneously into two rabbits (150–500 μg conjugated peptides per rabbit). The rabbits received a mixture of all three peptides. In the primary immunization, rabbits were injected with BSA-conjugated peptides and were boosted with the same amounts of KLH-conjugated peptides 3–4 weeks later. Individual rabbits were bled before injection (prebled), at boosting and every week after the initial injection. All procedures followed the CDC Animal Care and Use Guideline.

**Western blot.** Western blot analyses were used to probe the lysate or purified proteins with immune rabbit sera raised against *T. pallidum* or the three synthetic peptides (Towbin *et al.*, 1979; Rodes *et al.*, 2000). The whole-cell lysates were separated by SDS-PAGE on 10–20 % gradient gels. Proteins were transferred to nitrocellulose membranes by the method of Towbin *et al.* (1979), and the proteins were detected with immune rabbit sera (against *T. pallidum* or peptides) or anti-thioredoxin antibodies (against thioredoxin of the fusion protein). HRP-conjugated secondary antibodies (either anti-human IgG–HRP or anti-rabbit IgG–HRP) and normal human sera (RPR negative) were used as negative controls.

**ELISA.** ELISA microtitre plates (BD Bioscience, cat. no. 351177) were coated overnight at 4 °C with unconjugated peptides (1 μg ml$^{-1}$) in carbonate/bicarbonate buffer, pH 10. After three washes with TPBS (PBS, pH 7.2, containing 0.05 % Tween 20) and blocking for 2 h with 5 % non-fat dried milk in TPBS, sera were added. After 1 h incubation at room temperature, plates were washed three times with TPBS, and secondary antibody (goat anti-human IgG, H and L chains, 1 : 5000 dilution; Bio-Rad) was added and incubated for 30 min. HRP-conjugated secondary antibody alone and normal human serum (RPR negative) were used as negative controls in each experiment. After three more washes, tetramethyl-benzidine/hydrogen peroxide substrates (BioFX) were added, the reaction was stopped with 1 M H$_2$SO$_4$, and A$_{450}$ was determined with an ELISA plate reader (Anthos Labtec HT3). The background A$_{450}$ value (less than 0.050) was subtracted automatically and negative controls yielded A$_{450}$ <0.150.

**MS.** Gel-purified fusion protein and digested peptides were analysed by MALDI-TOF MS. All mass spectra were acquired in the positive ion mode using a Bruker Reflex IV (Bruker Daltonics) mass spectrometer equipped with delayed extraction and a nitrogen laser. The protein was digested with trypsin (10 : 1, w/w) in 25 mM ammonium bicarbonate at 37 °C overnight. The samples were mixed with matrix (sinapinic acid for protein, or *z*-cyano-4-hydroxyacinamic acid for peptides) and spotted onto a stainless steel MALDI target for MS analysis.

**Sequence analysis and GenBank accession numbers.** The sequences of the *arp* genes were assigned GenBank accession numbers (Table 3). Comparisons with known gene sequences and predictions of protein structure were made using the GCG (Genetics Computer Group) package from the University of Wisconsin, the MacVector software package (version 3.0; Eastman Kodak), and the pSort program (http://www.psort.org/).

**RESULTS AND DISCUSSION**

**Characteristics of the DNA sequences of the *arp* gene**

The complete *arp* gene ORF from the Nichols strain of *T. pallidum* subsp. *pallidum* has 1647 bp (Table 3, AF411124) and is high in G + C content (62.1 %), compared with the mean G + C content of 52.8 % of the *T. pallidum* genome (Fraser *et al.*, 1998). Analyses of the sequences of the *arp* gene of several laboratory strains revealed that the gene is highly conserved, with the 5’ and 3’ ends virtually identical in all strains sequenced to date. The most striking feature of the *arp* gene is a sequence of 14 almost identical 60 bp

---

**Table 2. Peptides used in the study**

The N-terminal cysteines (C) of Arp-5, -6 and -7 are for the conjugation of carrier proteins.

<table>
<thead>
<tr>
<th>Peptide used in the ELISA study</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arp 1</td>
<td>EVEDVPKVKVEPASEGREGER</td>
</tr>
<tr>
<td>Arp 2</td>
<td>EVEDAPKVVKVEPASEGREGER</td>
</tr>
<tr>
<td>Arp 3</td>
<td>PKNTAVEISLEKNAKAVVV</td>
</tr>
<tr>
<td>Arp 4</td>
<td>GAVLENFQRFGKDAFLKKAR</td>
</tr>
</tbody>
</table>

Peptides used in immunizing rabbits

| Arp-5*             | C-LVSPLS REVEDAPKVKVEPAS |
| Arp-6†             | C-REVEDAPKVVKVEPASEREGG |
| Arp-7‡             | C-PKVVEPASEREGEREVEDA |

*Corresponding to amino acids 121–140 in the predicted sequence, Nichols strain.
†Corresponding to amino acids 126–145 in the predicted sequence, Nichols strain.
‡Corresponding to amino acids 133–152 in the predicted sequence, Nichols strain.
repeats. This repeat region of the arp gene makes up approximately 51% of the sequence and is rich in GAGCGTGAGG motifs.

The series of internal repeats found in the Nichols strain (reference strain) was used as the standard to classify the repeats into three types (I, II and III) based on amino acid variations. The nucleotide sequences and corresponding amino acids of the three types of repeats are listed in Fig. 3. Analysis using the type I repeat as a standard revealed that the type II repeat differed from type I by a single pyrimidine base change, resulting in a conserved amino acid change (GCC→GTG; A→V). The type III repeat was different from the type I repeat at three additional locations: an AAG→GGG change, a GAG→GGG change and a CGT→CAT change. These changes represent K→G, E→G and R→H substitutions at the amino acid level. The substitutions are conserved pyrimidine→pyrimidine or purine→purine changes. None of the substitutions involves the third base. Thus, the arp gene of the Nichols strain contains five type I repeats, seven type II repeats and two type III repeats. The arp genes of T. pallidum subsp. pertenue strains CDC-1 and -2 have six and four repeats, respectively, and T. pallidum subsp. endemicum (Bosnia A strain) has eight repeats (Table 3). Among the three subspecies, the subspecies T. pallidum subsp. endemicum (Bosnia A) and T. pallidum subsp. pertenue (CDC-1 and -2) contain only type II repeats. The Nichols strain was the only strain to show variation in repeat type.

Additional features of the DNA sequence were palindromic sequences (CCGGCC) and an imperfect repeat sequence (GAGCGTGAG-GGAGGG-GAGCGTGAG; e.g. nt 421–444) within the repeat region. Both sequences had a high G+C content. This palindromic sequence could increase the probability of the recombination events resulting in a repeating theme. At the protein level, these sequences represent the junctions of two repeats and are also a repeating theme. At the protein level, these sequences represent the junctions of two repeats and are also a repeating theme. At the protein level, these sequences represent the junctions of two repeats and are also a repeating theme. At the protein level, these sequences represent the junctions of two repeats and are also a repeating theme. At the protein level, these sequences represent the junctions of two repeats and are also a repeating theme. At the protein level, these sequences represent the junctions of two repeats and are also a repeating theme.

Table 3. Arp protein characteristics

<table>
<thead>
<tr>
<th>Arp source</th>
<th>GenBank accession no.</th>
<th>Isoelectric point</th>
<th>No. repeats</th>
<th>Glu content (%)</th>
<th>Glu residues/total residues</th>
<th>No. Cys residues</th>
<th>No. Pro residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>T. pallidum subsp. pallidum, Nichols</td>
<td>AF411124 4.63</td>
<td>14</td>
<td>18.1</td>
<td>99/548</td>
<td>1</td>
<td>52</td>
<td></td>
</tr>
<tr>
<td>T. pallidum subsp. endemicum, Bosnia A</td>
<td>AF342807 4.83</td>
<td>8</td>
<td>15.2</td>
<td>65/428</td>
<td>1</td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>T. pallidum subsp. pertenue, CDC-1</td>
<td>AF411126 4.95</td>
<td>6</td>
<td>13.9</td>
<td>55/393</td>
<td>1</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>T. pallidum subsp. pertenue, CDC-2</td>
<td>AF342806 5.22</td>
<td>4</td>
<td>11.8</td>
<td>41/348</td>
<td>1</td>
<td>32</td>
<td></td>
</tr>
</tbody>
</table>

**Characteristics of protein sequences of the arp gene**

The putative Arp protein of the Nichols strain is predicted to be 59.4 kDa and is rich in charged amino acids, in particular 18.1% glutamic acid (99 of 548 amino acids; Table 3). This richness in glutamic acid, the basis of the name ‘acidic repeat protein’, results in a low predicted isoelectric point of 4.63 (Nichols strain). The N terminus of the Arp protein contains a hydrophobic domain (Q26–V60), which may span the cytoplasmic membrane. The end of this domain contains a sequence of four alanines (A45–A48), which may serve as a potential signal peptidase I processing site. In addition, the Arp protein lacks tryptophan residues, which have been shown to anchor proteins to the membrane (Schiffer et al., 1992).

The membrane location of Arp is difficult to determine. T. pallidum has two membrane layers in the cell envelope. The outer membrane is rich in lipids, and current evidence suggests that it contains a very limited amount of protein. The cytoplasmic membrane, on the other hand, contains many inserted proteins (Radolf, 1994, 1995). The single trans-membrane domain, lack of membrane-anchoring residues, and the single potential signal peptidase I site suggest that Arp is secreted. The Arp protein also has four phenylalanines within 15 residues of the C-terminal end of the protein, which is a characteristic of the membrane proteins of Gram-negative bacteria (Struyve et al., 1991). This means that the protein may be either anchored to the inner cytoplasmic membrane or secreted and released between the membranes into the periplasmic space. Alternatively, the protein may be released into the host upon the death of the organism. Studies are ongoing to resolve these questions.

A major feature of this protein sequence is the presence of consecutive and nearly identical repeats of a 20 amino acid sequence. These repeat sequences have an extremely high antigenic index, according to analysis using the Jameson–Wolf method (Jameson & Wolf, 1988). Among the amino acids of the repeat sequence, glutamic acid is 30% (6 of 20) of the total (Table 3).

While the combined tertiary structure of the 20 repeats is difficult to predict, there is a high probability that each repeat would form an extended alpha-helix [according to the Garnier–Robson, and Eisenberg methods (Garnier et al., 1978; Eisenberg et al., 1992)]. A high proportion of charged amino acids and the presence of extended alpha-helix
structures are common features in repeat proteins in pathogenic micro-organisms, and may be related to their functions (Anders et al., 1988; Anderson et al., 1990; Thole et al., 1990; Hood et al., 1996; Kajava, 2001).

Mechanisms of repeat generation

Antigenic repeats can be susceptible to immune selection. When a protein is an important pathogenic or virulence factor, it tends to be conserved, and repeats with variations evolve as a result of immune pressure. It is also known that the length of the repeat sequences can affect immunity. This effect has been shown in studies on group B streptococci, in which the host response to proteins containing small numbers of repeats does not always result in immunity to the same protein with a greater number of repeats (Gravekamp et al., 1996). The immune response toward the Arp protein is significant, as demonstrated by the strong antibody reactivity toward the protein and peptides in the Western blot assay and ELISA, respectively (see below). Because the repeat region of this protein is characterized by almost identical repeats, it suggests that the repeat region has an important function. Clinically, we have observed repeat numbers varying from four to 20 in clinical isolates (Cox et al., 2003).

There are two potential mechanisms used by bacteria to develop multiple repeats in their protein sequences (Andrade et al., 2001; Felger et al., 1997; Kruglyak et al., 1998). The first mechanism, strand slippage, is a process in which the DNA polymerase slips from one repeat region in the gene to another because of the similarities in the sequences (Kajava, 2001). The second mechanism, recombination, usually generates longer repeats, whereas strand slippage usually occurs in shorter repeats (Moxon et al., 1994). The repeat motifs are targets for both recombinational and slip-strand mechanisms of alteration (Deitsch et al., 1997; Moxon et al., 1994), and they can vary considerably from one strain to another, or in some cases within a single strain during the course of an infection. The recombination and mutation-induced changes can include expansion (duplication) or deletion of repeats and alteration in the sequence of the repeats. These genetic mechanisms are also known to allow the introduction of point mutations at an elevated rate (Deitsch et al., 1997; Moxon et al., 1994). For example, the M protein of Streptococcus pyogenes (Hollingshead et al., 1987) shows antigenic variability through both insertion-deletion events and elevated levels of point mutations, and this variation is important in maintaining chronic infection (Harbaugh et al., 1993).

In this study, the arp gene showed considerable heterogeneity in size when different strains of T. pallidum were examined. The structure of the 60 bp repeat suggested an obvious mechanism for the variability in the number of repeats. The repeats were made up of perfect inverted repeats at the amino acid level (EREG-GERE) and imperfect repeats at the nucleic acid level. Because the type II repeat was predominant in the Nichols strain, one could hypothesize that type II would also predominate in other strains. We sequenced the Bosnia A, and CDC-1 and -2 strains, which represent two subspecies, and all contained the type II repeat only. Further experiments are currently under way to determine whether the type II repeat is also predominant in clinical isolates of T. pallidum subsp. pallidum.

Fig. 1. Western blot analysis of expressed clones. arp, cloned into pBAD/TOPO ThioFusion, which also contains a histidine tag, was expressed. The expressed protein ran on a 10–20% gradient gel and was probed with rabbit sera before immunization (lane 2, pre-bleed), 23 days after infection with T. pallidum (lane 3, p.i.), with anti-thioredoxin antibody (lane 4), or with second antibody only (lane 5, no serum). A band corresponding to approximately 110 kDa was seen. Lane 1, molecular mass standard.

Fig. 2. ELISA of Arp peptides. Peptides corresponding to type II repeat (Arp 1, □), type I repeat (Arp 2, ◊), N-terminal protein sequence (Arp 3, ⊙) or C-terminal protein sequence (Arp 4, △) were probed with human syphilitic sera. Results represent the means of duplicate assays. Backgrounds (usually A_450 <0.05) were subtracted automatically. Negative controls (normal donors; RPR negative) yielded A_450 <0.15 (not shown).
Expression of the arp gene and detection of the Arp protein

The arp gene from the Nichols strain of *T. pallidum* subsp. *pallidum* was cloned into the pBAD/TOPO ThioFusion vector, resulting in a fusion protein of Arp and thioredoxin with a polyhistidine tag. Using Western blot analyses, we probed the purified protein with sera collected from rabbits immunized with a polyhistidine tag. The expressed fusion protein had a predicted size of 72 kDa. In the SDS-PAGE gel analyses, however, we observed a band at 110 kDa. When examined using Western blot, this band reacted with anti-thioredoxin antibody (Fig. 1). The identity of this apparent 110 kDa protein was further confirmed through MS analysis. The gel-eluted protein gave a nominal mass of 72 kDa via MALDI-TOF analysis, corresponding to the calculated value for the fusion with thioredoxin. The eluted protein was also subjected to peptide mass fingerprinting through tryptic digestion followed by MALDI-TOF analysis. This technique allows the identification of proteins by matching experimental peptide masses to the theoretical ones generated from a protein or DNA database. More than 13 peptides digested from the gel-eluted protein were found to match the fragments from Arp protein, and 34% sequence coverage was obtained, indicating a protein with an apparent relative molecular mass of 110 kDa, as for Arp. An altered migration pattern on SDS gels has been observed in fibronectin binding proteins (Joh et al., 1994; Rakonjac et al., 1995), and the Arp protein has a potential fibronectin binding site within its multiple repeat region (McGavin et al., 1993).

Detection of anti-Arp antibodies by ELISA

The ability of sera from patients with syphilis to recognize the Arp protein was demonstrated using an ELISA. Four peptides (Table 2) were designed according to potential antigenicity. Peptides Arp 1 and 2 represent repeat types II and I, respectively. Peptides Arp 3 and 4 were from either the N terminus or the C terminus of the Arp protein. RPR-positive human sera showed a strong response against Arp peptides 1 and 2, but not against Arp 3 or 4. A representative experiment is shown in Fig. 2. Sera from normal individuals (RPR negative) reacted at background level (A_{50}<0.15; data not shown). These results suggested that humans infected with *T. pallidum* produce antibodies predominantly against the repeat region, but not against the N or C terminal of the Arp protein. Thus, we demonstrated that the repeat domain of the protein is a specific area of recognition for a syphilis-specific antibody. Because proteins with repeating sequences are often highly immunogenic, the Arp peptides could be useful candidates for the development of a treponemal test for diagnosing syphilis. The functional role of anti-Arp antibody during the relapsing course of untreated syphilis will be an important area for future research.

ACKNOWLEDGEMENTS

We thank Raphel Castillo (Ferguson Fellow Program, CDC) and Cassey Parker [Coordinating Center for Infectious Diseases (CCID), CDC] for assistance in sequencing; and Amy Martin, Keith Levert and Dongxia Wang [Division of Scientific Resources (DSR), CDC] for MS analyses. The Arp protein is covered by US patent no. 7,005,270 and multiple international patents. Use of trade names is for identification only and does not constitute endorsement by the US Department of Health and Human Services, the Public Health Services, or the CDC. The findings and conclusions in this paper are those of the authors and do not necessarily represent the views of the CDC.

REFERENCES


Molecular characterization of Arp protein


