

## The *tprK* Gene Is Heterogeneous among *Treponema pallidum* Strains and Has Multiple Alleles

ARTURO CENTURION-LARA,\* CHARMIE GODORNES, CHRISTA CASTRO,  
WESLEY C. VAN VOORHIS, AND SHEILA A. LUKEHART

Department of Medicine, University of Washington, Seattle, Washington 98195

Received 2 August 1999/Returned for modification 8 September 1999/Accepted 1 November 1999

**We have previously shown that the TprK antigen of *T. pallidum*, Nichols strain, is predominantly expressed in treponemes obtained 10 days after infection and that the hydrophilic domain of TprK is a target of opsonic antibodies and confers significant protection against homologous challenge. The *T. pallidum* genome sequence reported the presence of a single copy of the *tprK* gene in the Nichols strain. In the present study we demonstrate size heterogeneity in the central portions of the TprK hydrophilic domains of 14 treponemal isolates. Sequence analysis of the central domains and the complete open reading frames (ORFs) of the *tprK* genes confirms this heterogeneity. Further, multiple *tprK* sequences were found in the Nichols-defined *tprK* locus in three isolates (Sea 81-4, Bal 7, and Bal 73-1). In contrast, only a single *tprK* sequence could be identified in this locus in the Nichols strain. Alignment of the DNA and deduced amino acid sequences of the whole *tprK* ORFs shows the presence of seven discrete variable domains flanked by highly conserved regions. We hypothesize that these heterogeneous regions may be involved in antigenic heterogeneity and, in particular, evasion of the immune response. The presence of different *tprK* alleles in the *tprK* locus strongly suggests the existence of genetically different subpopulations within treponemal isolates.**

*Treponema pallidum* subsp. *pallidum* (referred to here as simply *T. pallidum*), the etiologic agent of syphilis, causes a lifelong chronic infection in untreated individuals. Syphilis has three distinct clinical stages: the primary localized chancre, the disseminated secondary stage, and the late tertiary phase. The host develops rapid and vigorous humoral and cellular immune responses against *T. pallidum* which eliminate most of the treponemes from primary and secondary lesions. However, a few treponemes evade the immune responses and lead to persistent infection. It has been shown that phagocytosis by macrophages is the major clearance mechanism of *T. pallidum* from early lesions (20, 21). Rabbit macrophages are able to ingest and kill *T. pallidum* in vitro, but efficient phagocytosis requires specific opsonic antibodies (2), presumably directed against surface exposed antigens.

No outer membrane antigens have yet been definitively identified in *T. pallidum* (25). Unlike gram-negative bacteria such as *Escherichia coli*, the outer membrane of *T. pallidum* is very fragile (12) and is easily disrupted by physical manipulation. This characteristic has resulted in misidentification of highly immunogenic periplasmic lipoproteins as surface exposed antigens (24), although these molecules are now thought to be anchored to the periplasmic leaflet of the cytoplasmic membrane (11, 25–27). Freeze fracture electron microscopy studies reveal a very limited number of surface-exposed transmembrane proteins localized in the outer membrane (28, 32). These have been called Tromps (treponemal rare outer membrane proteins) because of their extraordinarily low density, and it has been proposed that the antibody causes the aggregation of Tromps in the intact treponeme (6, 19). Two single-copy genes have been proposed to encode rare outer membrane proteins (3, 5, 10). Tromp1 is homologous to periplasmic binding proteins of ABC transport systems (17), and its outer

membrane location and ion channel activity are in dispute (1, 4, 13, 18). Tromp2 is a 28-kDa protein, but its outer membrane localization has not yet been confirmed independently. Recent studies have reported that the glycerophosphodiester phosphodiesterase (Gpd) of *T. pallidum* is a lipoprotein that binds the Fc fragment of human immunoglobulin G and has immunoprotective capacity against homologous challenge in experimental syphilis, suggesting that this molecule may be surface exposed (7). A second study proposes that Gpd is a periplasmic protein associated with the peptidoglycan-cytoplasmic membrane complex (29). Thus, the identities of surface-exposed molecules in *T. pallidum* are still undetermined.

A new 12-member gene family, termed *tpr*, has been recently identified in the Nichols strain of *T. pallidum* (8, 16, 30). The predicted amino acid sequences of the *tpr* genes (*tprA* through *tprL*) have homology with the major sheath protein (Msp) antigen of *Treponema denticola*, which is reported to be surface exposed, to be involved in cell attachment, and to have porin activity (14, 15). Three *tpr* subfamilies (I, II, and III) can be identified by their predicted amino acid homology (8). Although there is some homology in the amino acid sequences among all Tpr proteins, comparison of the amino acid sequences of the members of subfamilies I and II shows that the NH<sub>2</sub>- and COOH-terminal regions are conserved, whereas the central domains are variable in terms of sequence and length. Subfamily III is composed of five members that are comparatively poorly homologous to each other or to the other Tpr proteins but that still retain small areas of conservation. Using PSORT analysis, three Tpr proteins are predicted to be located in the outer membrane of *T. pallidum*: two from subfamily I (TprF and TprI) and one from subfamily III (TprK) (8). Recent studies have demonstrated that the Nichols TprK antigen, encoded by a single-copy gene in the Nichols strain, is preferentially expressed during infection, is the target of opsonic antibodies, and induces a partially protective immune response (8). We report in this study the existence of multiple, heterogeneous *tprK* alleles in *T. pallidum* isolates other than the

\* Corresponding author. Mailing address: Harborview Medical Center, 325 Ninth Ave., Box 359779, Seattle, WA 98104-2499. Phone: (206) 341-5364. Fax: (206) 341-5363. E-mail: acentur@u.washington.edu.

TABLE 1. *T. pallidum* isolates

Strain	Source, stage <sup>a</sup>	Geographical location	Yr isolated
<i>T. pallidum</i> subsp. <i>pallidum</i>			
Sea 84-2	CSF, 2	Seattle, Wash.	1984
Nichols	CSF, 2	Washington, D.C.	1912
Bal 2	CSF, congenital	Baltimore, Md.	?
Bal 8	CSF, congenital	Baltimore, Md.	?
Bal 3	Blood, congenital	Baltimore, Md.	?
Bal 7	CSF, post-Rx	Washington, D.C.	1976
Bal 73-1	Aqueous humor, congenital	Baltimore, Md.	1973
Sea 81-4	Chancre, 1	Seattle, Wash.	1980
Sea 81-3	CSF, 2	Seattle, Wash.	1981
Chicago	Chancre, 1	Chicago, Ill.	1951
Mexico A	Chancre, 1	Mexico	1953
Sea 83-1	CSF, 1	Seattle, Wash.	1983
<i>T. pallidum</i> subsp. <i>pertenue</i>			
Gauthier	Generalized lesions	Ghana	1982
Haiti B <sup>b</sup>	Abdominal lesions	Haiti	1951

<sup>a</sup> CSF, cerebrospinal fluid; post-Rx, after therapy.

<sup>b</sup> Molecular analysis suggests that the Haiti B isolate may be *T. pallidum* subsp. *pallidum* (9).

Nichols strain, while only a single allele is detectable in the Nichols strain.

#### MATERIALS AND METHODS

**Treponemal strains and DNA extraction.** All *T. pallidum* isolates were propagated in New Zealand White rabbits (22). This project was approved by the University of Washington Animal Care Committee, and animals were handled according to institutional guidelines. The treponemal isolates used in this study were provided by Paul Hardy and Ellen Nell (Johns Hopkins University), James Miller (University of California, Los Angeles), and Peter Perine (Centers for Disease Control and Prevention) or else were isolated at the University of Washington (Table 1). Nichols strain, the standard laboratory strain, has been maintained in rabbits since its isolation in 1912. In contrast, all other isolates have been kept as frozen stocks with limited passage in rabbits.

Suspensions of each treponemal strain were collected, taking careful precautions to avoid cross-contamination, and spun in a microcentrifuge at 12,000 × g for 30 min at 4°C. The pellet was resuspended in 200 μl of 1× lysis buffer (10 mM Tris, pH 8.0; 0.1 M EDTA; 0.5% sodium dodecyl sulfate), and DNA was extracted by using the Qiagen Kit for genomic DNA extraction (Qiagen, Inc., Chatsworth, Calif.) as described elsewhere (9). Each strain was handled separately under stringent PCR-clean conditions.

**Primers and PCR amplification of central domains of *tprK* genes.** The DNA sequences of the Nichols strain *tprK* gene (TP0897) and its flanking regions were obtained from the published *T. pallidum* genome sequence (16). A set of primers, *tprK*-S and *tprK*-As (Table 2 and Fig. 1), was designed to amplify a region of 410 bp (base positions 974778 through 975187) coding for a portion of a putative large hydrophilic domain the TprK antigen (8). PCR amplification of this hydrophilic domain was performed on genomic DNA from 13 more recent isolates and from the Nichols strain maintained in our laboratory (Table 1). PCR amplification of the 14 *T. pallidum* isolates was performed by using a 100-μl reaction

containing 200 μM concentrations of deoxynucleoside triphosphates (Promega, Madison, Wis.), 50 mM Tris-HCl (pH 9.0 at 20°C), 200 mM NH<sub>4</sub>SO<sub>4</sub>, 1 μM concentrations of each primer, 1.5 mM MgCl<sub>2</sub>, and 2.5 U of *Taq* polymerase (Promega). One microliter of purified genomic DNA was used as a template. The cycling conditions were as follows: denaturation at 94°C for 3 min and then 40 cycles of 94°C for 1 min, 64°C for 2 min, and 72°C for 1 min, with a final extension step of 10 min at 72°C. Amplicons of the 14 isolates were then separated in 3% TBE-NuSieve agarose gels. The products of a minimum of two independent PCR reactions per isolate were examined by electrophoresis. Four isolates were then chosen for sequence analysis: Bal 7, Bal 73-1, Sea 81-4, and the Nichols strain maintained in our laboratory. An aliquot from their amplicons was used for direct cloning and sequence analysis as described below.

**Identification of *tprK* alleles in *T. pallidum* strains located in the Nichols *tprK* locus.** To identify the *tprK* sequences of the Bal 7, Bal 73-1, Sea 81-4, and the Nichols isolates that are located in the same locus as the Nichols *tprK* described in the *T. pallidum* genome, we used the primers F5-S and F3-As designed in the 5' and 3' *tprK* flanking regions (Table 2). These oligonucleotides amplify DNA fragments encompassing the 5'-flanking region, the *tprK* open reading frame (ORF) and the 3'-flanking region (Fig. 1). The PCR conditions were the same as described above. PCR products were separated in 1% TBE-agarose gels to confirm the presence of amplicons of the expected molecular weight, and an aliquot was used for direct cloning and sequence analysis.

**Amplification of the 3' end of the Nichols *tprK* gene and the 3'-flanking region.** In order to rule out possible PCR or sequencing artifacts and to confirm our findings, amplicons from independent PCR reactions encompassing the second half of the *tprK* gene and its 3'-flanking region were obtained from the Nichols strain maintained in our laboratory. We used the *tprK*-S and the F3-As primers (Table 2) under the same PCR conditions as those described above. The size of the amplicons was confirmed by agarose gel electrophoresis, and an aliquot was used for direct cloning and sequencing.

**Cloning and sequencing.** After PCR amplification, the products from the different PCR reactions described above were directly cloned into the TOPOII T/A cloning vector (Invitrogen, Carlsbad, Calif.) according to the manufacturer's instructions. Double-stranded plasmid DNA from multiple clones from Bal 7, Bal 73-1, Sea 81-4, and the Nichols isolates containing inserts of *tprK* internal DNA fragments and from clones from amplicons encompassing the 5' and 3' flanking regions plus the *tprK* ORF (Fig. 1) were purified with the Qiagen Plasmid Minikit (Qiagen). Full automated sequencing in both directions was performed on the inserts by the dye terminator method (Perkin-Elmer, Foster City, Calif.) according to the manufacturer instructions but adding molecular-grade dimethyl sulfoxide to a 5% final concentration. The short inserts were sequenced in their full length in both directions with primers pairs M13 (forward) and M13 (reverse) (vector primers), as well as *tprK*-S and *tprK*-As (Table 2 and Fig. 1). The inserts encompassing the 5'- and 3'-flanking region and the *tprK* ORF were sequenced with the M13 forward, M13 reverse, F5-S, F3-As, FW-S, FW-As, 9V-S, 9V-As, *tprK*-S, and *tprK*-As primers (Table 2). The inserts encompassing the 3' end of the *tprK* gene and its 3'-flanking region were sequenced with the M13 forward, M13 reverse, *tprK*-S, *tprK*-As, F3-As, FW-As, and 9V-As primers (Table 2 and Fig. 1). Sequences were assembled by using the CAP sequenced assembly program (<http://cg.tigem.it/ASSEMBLY/assemble.html>).

TABLE 2. Primers for PCR amplification and sequencing

Primer	Sequence
Sense	
M13 forward.....	5'-GTTTTCCCAGTCACGAC
F5-S.....	5'-TCCCCAGTTGCAGCACTAT
FW-S.....	5'-ATGATTGACCCATCTGCCACTTC
9V-S.....	5'-ATATTGAAGGCTATGCGGAGCTC
<i>tprK</i> -S.....	5'-AGTTTGCGTCTAACACCGACTG
Antisense	
M13 reverse.....	5'-CAGGAAACAGCTATGAC
F3-As.....	5'-TCGCGGTAGTCAACAATACCA
FW-As.....	5'-CTACCAAATCAAGCAGCATGCC
9V-As.....	5'-CCTCAAGGAAAGAAGTATCAGG
<i>tprK</i> -As.....	5'-TCGCATGGCCATGTTGAGAAAT

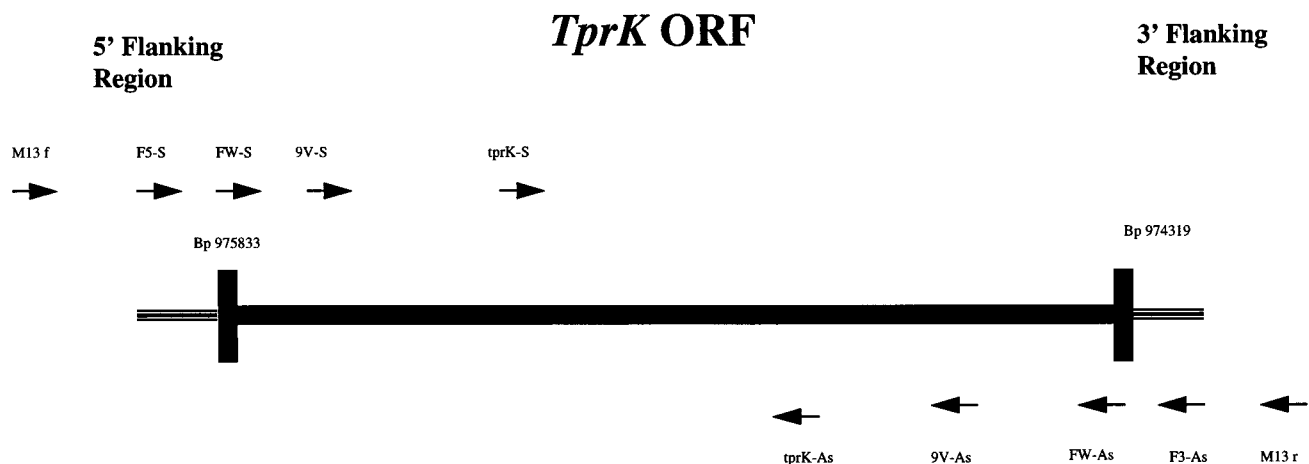


FIG. 1. Orientation and nucleotide position of the different primers used for PCR and sequencing of the central regions of the *tprK* ORF as well as the complete ORF and its flanking regions. Solid lines represent the ORF of the *tprK* gene. Primer positions are as follows: F5-S, base positions 975967 through 975987 upstream of the *tprK* start codon; FW-S, base positions 975809 to 975833; 9V-S, base positions 975629 to 975652 in the *tprK* ORF; *tprK*-S, base positions 975162 to 975187 in the *tprK* ORF; *tprK*-As, base positions 974778 to 974800 in the *tprK* ORF; 9V-As, base positions 974708 to 974730 in the *tprK* ORF; FW-As, base positions 974319 to 974341 in the *tprK* ORF; and F3-As, base positions 973922 to 973943 downstream of the *tprK* stop codon.

**Nucleotide sequence accession numbers.** The sequences determined in this study were deposited in GenBank under accession numbers AF194339 to AF194370.

## RESULTS

**Identification of multiple *tprK* alleles in other isolates of *T. pallidum*.** PCR amplification of the 13 *T. pallidum* isolates (Table 1) and the Nichols strain with the *tprK*-S and *tprK*-As primers resulted in amplicons of different molecular weights and, in some isolates, the presence of two different size amplicons as determined by high-resolution agarose gel electrophoresis (Fig. 2). Cloning and sequencing of this region in three *T. pallidum* isolates (Sea 81-4, Bal 7, and Bal 73-1) resulted in the identification of multiple distinct *tprK* sequences in the more recent isolates: seven sequences in Sea 81-4 and eight each in the Bal 73-1 and Bal 7 isolates. Interestingly, all 14 clones analyzed from the Nichols strain yielded a single *tprK* sequence identical to the corresponding portion of the *T. pallidum* genome sequence. Table 3 shows the number of clones sequenced and the number of different *tprK* alleles identified per isolate.

Alignment of the DNA (not shown) and deduced amino acid sequences (Fig. 3) of these amplicons (primer binding sites are excluded) shows striking regions of heterogeneity flanked by highly conserved domains. There are three very localized regions of heterogeneity which appear to be due to base changes, insertions, and deletions. A few minor changes are seen scattered throughout the conserved domains. It is remarkable that,

despite the high heterogeneity in the variable domains, no stop codons or frameshifts have been introduced in the DNA sequences, giving complete ORFs in the predicted amino acid sequences of these alleles. The Nichols strain and one clone from the Sea 81-4 isolate have the shortest amplicon in this region with 124 predicted amino acids, while the Bal 73-1 isolate has the longest amplicon of 137 predicted amino acids. The middle variable region is the most heterogeneous portion of this region. Comparative analysis of the *tprK* sequences suggests roughly equal variability among and within isolates (not shown).

**Identification of different *tprK* alleles in the Nichols *tprK* locus in street isolates.** The single-copy *tprK* gene (TP0897) of the Nichols strain is located in the *T. pallidum* genome (16) at base positions 975833 through 974319, flanked at its 5' and 3' ends by a putative ATP-dependent nuclease (TP0898) and hypothetical proteins (TP0896 and TP0895, respectively). As a first attempt to elucidate the genetic arrangement of the multiple *tprK* alleles in other isolates, we amplified the *tprK* sequence(s) located in the Nichols *tprK* locus (the 5'- and 3'-flanking regions plus the *tprK* ORF) in other isolates by using the F5-S and F3-As primers (Fig. 1). The amplicons were cloned, sequenced, and compared with the corresponding regions of the *T. pallidum* genome from Nichols strain. Surprisingly, multiple alleles were identified in this locus within individual *T. pallidum* isolates. Table 4 shows the number of clones examined and the number of different *tprK* sequences identified at this locus. Three *tprK* genes were found in five clones

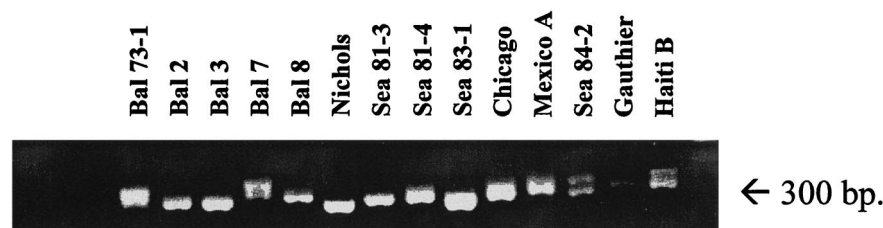


FIG. 2. High-resolution ethidium bromide agarose gel showing amplicons of *T. pallidum* isolates obtained with the short-range primers (*tprK*-S and *tprK*-As) which amplify a central region in a large hydrophilic domain of the TprK antigen. Amplicons vary in the number of bands and sizes.

TABLE 3. Diversity of *tprK* internal region in *T. pallidum* isolates

Isolate	No. of clones sequenced	No. of distinct sequences identified
Nichols	14	1
Bal 7	10	8
Bal 73-1	15	8
Sea 81-4	16	7

TABLE 4. Diversity of whole *tprK* ORFs in the Nichols *tprK* locus

Isolate	No. of clones sequenced	No. of distinct sequences identified
Nichols	5	1
Bal 7	5	3
Bal 73-1	1	1
Sea 81-4	5	2

from Bal 7 isolate, in two sequences in five clones from Sea 81-4, and in one sequence in a single clone from Bal 73-1. Sequencing of five clones from the Nichols strain maintained in our laboratory yielded a single sequence throughout both the whole *tprK* ORF and the 5'- and 3'-flanking regions. Sequence comparison of the deduced protein sequences of the different *tprK* genes from the four isolates revealed seven regions of heterogeneity (V<sub>1</sub> to V<sub>7</sub>) flanked by highly conserved

domains (Fig. 4). Five of the variable domains (V<sub>3</sub> to V<sub>7</sub>) are located in the second half of the TprK proteins, while only two (V<sub>1</sub> and V<sub>2</sub>) are located in the first half of the TprK proteins. As in Fig. 3 above, the heterogeneous regions are due to base changes, insertions, or deletions. Remarkably, all of these changes do not introduce frameshifts or stop codons in the large *tprK* ORFs.

The predicted amino acid sequences of the *tprK* whole

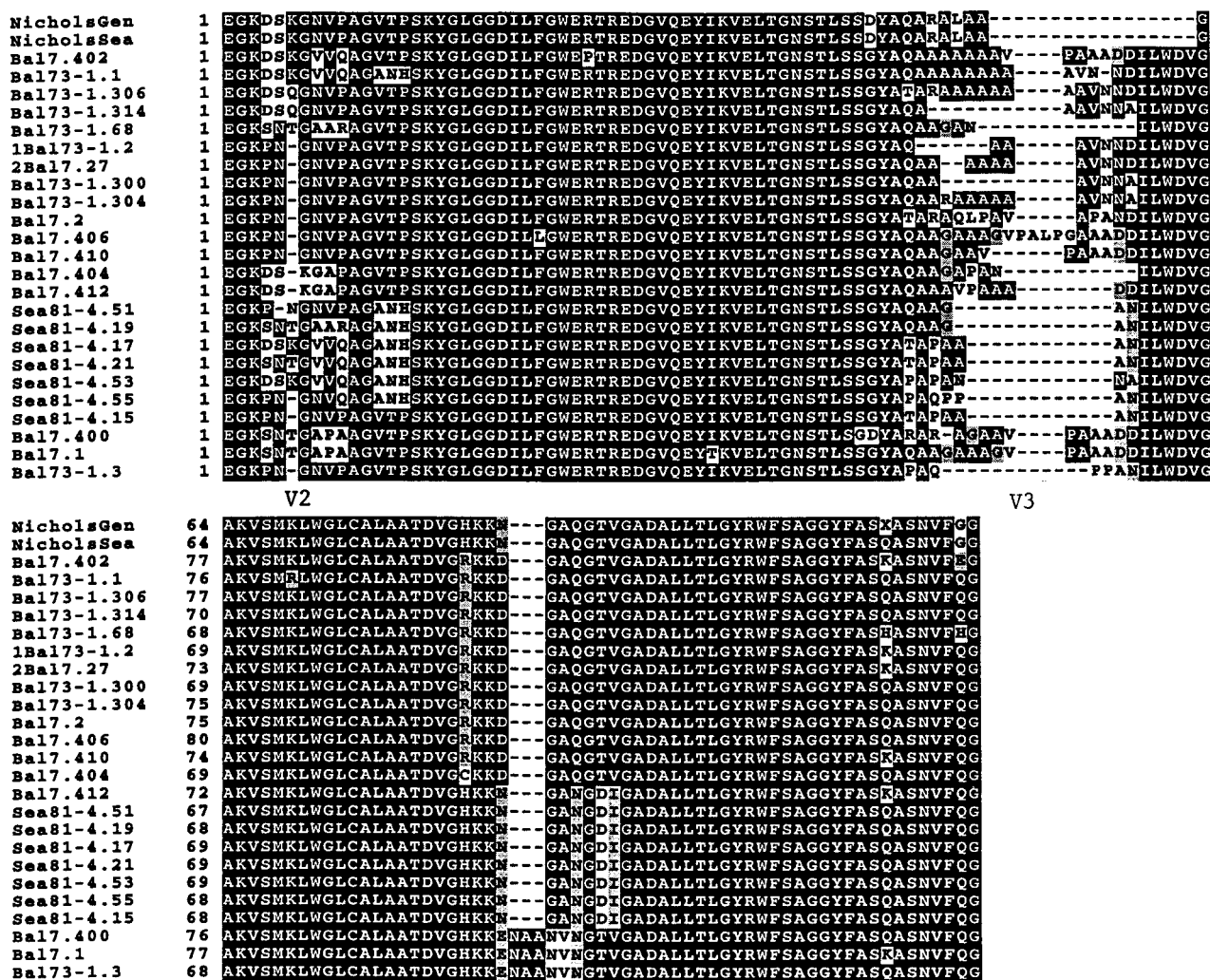


FIG. 3. Alignment of the TprK internal region predicted amino acid sequences of *T. pallidum* isolates obtained with the *tprK*-S and *tprK*-As primers showing highly conserved regions and three variable domains. Shaded areas indicate sequence identity, and broken lines indicate gaps in the alignment. NicholsGen, Nichols strain sequence as reported in the *T. pallidum* genome sequence; NicholsSea, sequence of the Nichols strain maintained in our laboratory. The street isolates shown in this alignment are Bal 7, Bal 73-1, and Sea 81-4 listed in Table 1. The sequences in Fig. 3 to 5 are indicated by isolate name, followed by the clone designation. For example, Bal 73-1.306 is the sequence from clone 306 of the Bal 73-1 isolate.

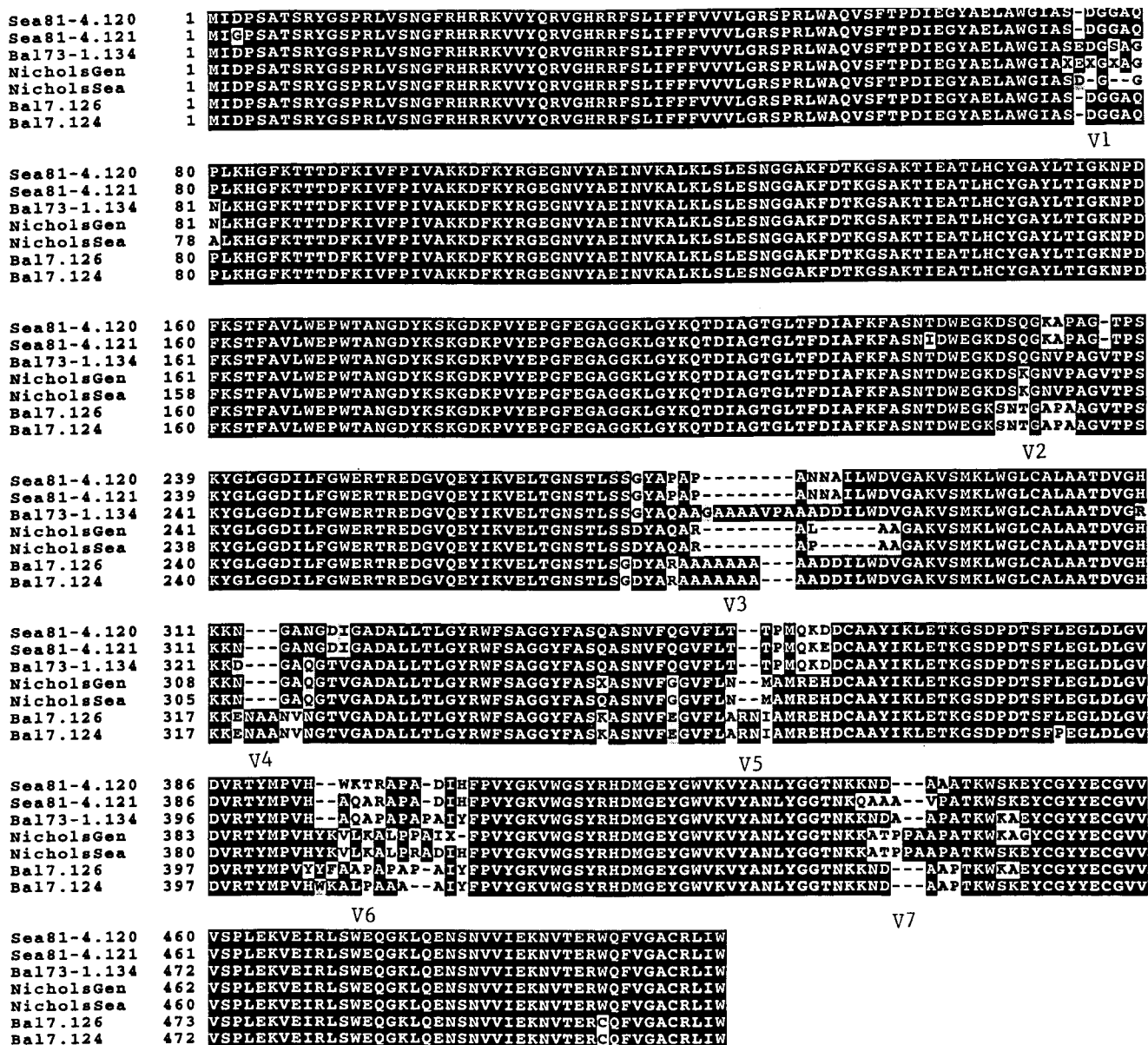


FIG. 4. Alignment of the predicted amino acid sequences of complete ORFs identified in the Nichols *tprK* locus in five *T. pallidum* isolates, showing seven regions of heterogeneity (V<sub>1</sub> to V<sub>7</sub>). Shaded areas indicate sequence identity, and broken line show the gaps in the alignment. NicholsGen, Nichols strain sequence as reported in the *T. pallidum* genome sequence; NicholsSea, sequence of the Nichols strain maintained in our laboratory. Bal 7, Bal 73-1, and Sea 81-4 are street isolates (Table 1). Additional numbers indicate the clone designations from each isolate. All sequences are from amplicons obtained with the F5-S and F3-As primers which bind the *tprK* flanking regions; therefore, no primer binding sites are included in this alignment.

ORFs of the Nichols isolate from the *T. pallidum* genome sequence and from the Nichols strain maintained in our laboratory are almost completely identical. However, they differ slightly in ORF size, giving 505 amino acids for the genome Nichols and 503 amino acids for the Seattle Nichols strain. The products of two independent PCR reactions and sequences of multiple clones consistently yielded the same *tprK* ORF sequence from our Nichols strain. The differences between the two Nichols sequences are located in the V<sub>1</sub> variable region (three amino acid deletions and two amino acid changes in our sequence), the V<sub>3</sub> region (one amino acid change), and the V<sub>6</sub> region (one amino acid deletion and two amino acid changes in the genome se-

quence) and in the V<sub>7</sub> variable domain (three amino acid changes) (Fig. 4).

Analysis of the DNA alignments of the 3'-flanking region of *tprK* in the genome Nichols strain and in Sea 81-4 shows the presence of two putative hairpin structures (Fig. 5) in these strains, beginning 19 and 57 bases downstream of the *tprK* stop codon, respectively, and encompassing distances of 32 and 52 bases. Neither is followed by a run of T (U in RNA) residues. These characteristics are found in Rhodospirillum rubrum dependent putative transcription termination sites. In contrast, the Bal 7 and Nichols (Seattle) isolates have a 67-base deletion beginning 38 nucleotides downstream of the *tprK* stop codon, which causes the loss of portions of these pu-

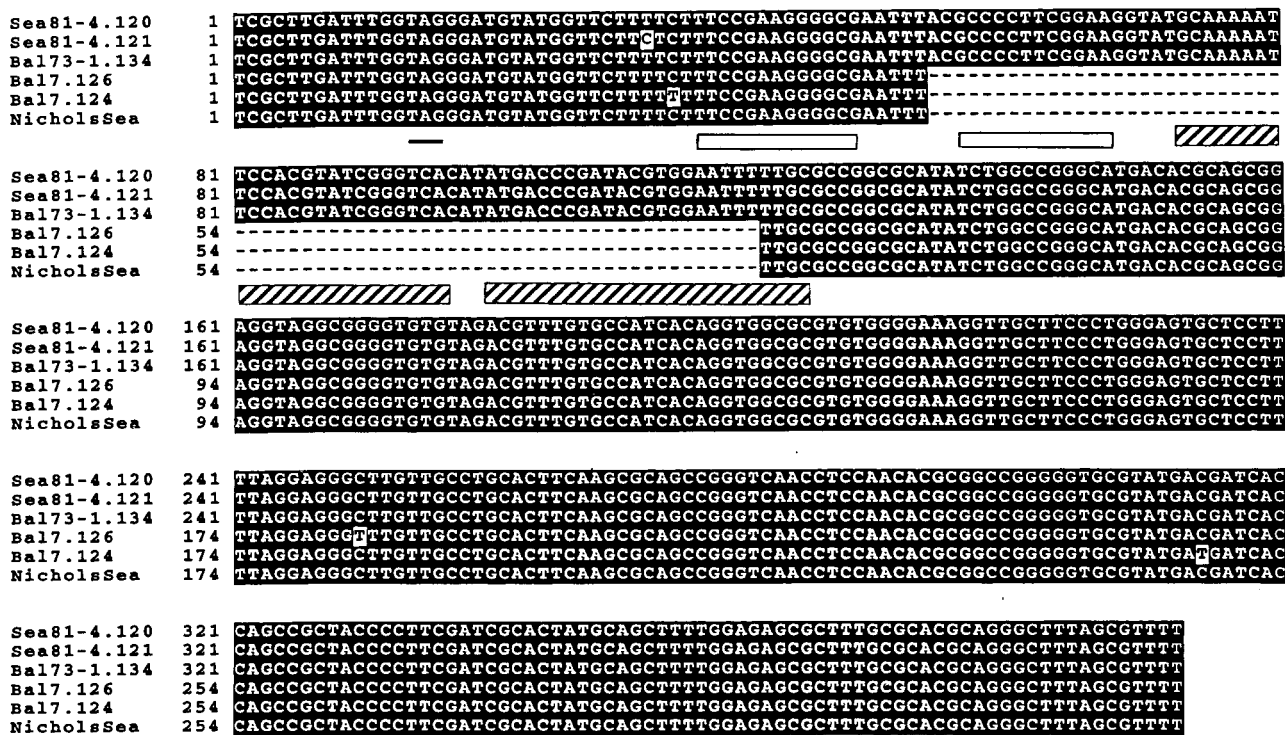


FIG. 5. DNA alignment of the last 16 nucleotides of the *tprK* ORF and its 3'-flanking region from the Bal 7, Bal 73-1, and Sea 81-4 street isolates and from the Nichols strain (NicholsSea, strain maintained in our laboratory; NicholsGen, genome Nichols strain). Single underline indicates the *tprK* stop codon. Open and hatched boxes indicate the two putative transcription termination hairpin structures. Broken lines indicate a sequence deletion of 67 nucleotides in the NicholsSea isolate and the Bal 7 isolate. Sequences shown are the complement of the corresponding regions in the *T. pallidum* genome.

tative hairpin structures. These findings have been confirmed with identical sequences obtained from independent PCR reactions by using a different combination of primers (*tprK*-S and F3-As). Unlike the 3'-flanking regions, the 5'-flanking sequences show almost complete nucleotide identity among all isolates (not shown).

**DISCUSSION**

*T. pallidum* is cleared from early syphilis lesions by macrophage-mediated phagocytosis; this activity is facilitated by the presence of opsonic antibody. We recently showed that antibodies directed against a large hydrophilic region of TprK are opsonic for *T. pallidum* in vitro and that immunization with this recombinant peptide is partially protective against homologous infection (8). Because of the potential importance of TprK in the functional and protective immune responses in syphilis, we have investigated its structure in other *T. pallidum* isolates. This report demonstrates the existence of multiple alleles of *tprK* within isolates and heterogeneity in *tprK* sequences among isolates.

Our DNA sequences are derived from PCR-amplified products, and we recognize that some random sequence artifact may occur during PCR amplification. However, several findings suggest that the heterogeneity that we have identified is not artifactual. First, all 19 clones (derived from four independent PCR amplifications) containing inserts of the Nichols strain yielded identical sequences in this region. Second, identical *tprK* sequences were obtained from independent PCR amplifications of multiple strains (not included in this study). Finally, the observed heterogeneity in *tprK* sequence, including

deletions and insertions, does not introduce frameshifts and stops in the ORF.

Sequence comparison of the *tprK* DNA and their predicted amino acid sequences shows heterogeneous domains flanked by highly conserved regions. Sequence alignments between the *tpr* variable regions and the *T. pallidum* genome failed to identify regions of identity elsewhere in the genome. Close analysis reveals that some *tprK* genes encode nearly identical proteins, such as the internal TprK fragments Sea 81-4.17 and Sea 81-4.21 (only two amino acid differences) and the complete TprK proteins Sea 81-4.120 and Sea 81-4.121 (nine amino acid differences). Furthermore, complete identity in motifs encompassing the whole V<sub>4</sub> region can be seen between and within isolates (Fig. 3). The motif RKKDGAQGTV is shared by 12 sequences from Bal 7 and Bal 73-1 isolates, the motif HKKN GANGDI by 8 sequences from the Sea 81-4 and Bal 7 isolates, and the motif HKKENAANVNGTV by 3 TprK sequences of Bal 7 and Bal 73-1 isolates. On the other hand, some sequences within an isolate are highly divergent, as in the V<sub>4</sub> regions of Bal 73-1.306, Bal 73-1.314, and Bal 73-1.3 sequences (Fig. 3). Roughly equal variability in the TprK sequences is seen between and within isolates.

Multiple *tprK* sequences found within an isolate could arise from multiple *tprK* genes within a single bacterium or from multiple subpopulations within an isolate, with each subpopulation carrying a single *tprK* gene. The presence of genotypically different bacterial organisms within an isolate is demonstrated by the existence of different *tprK* gene sequences located in the same locus of the chromosome. We have shown in this study that there are genotypically different bacterial subpopulations present within *T. pallidum* isolates (Fig. 4). Thus, the diversity of *tprK* alleles in

*T. pallidum* strains may indicate the presence of a large number of different treponemal subpopulations or genetic variants, each carrying a single *tprK* in the same chromosomal locus. In addition to multiple alleles at the recognized *tprK* locus, additional *tprK* sequences could be spread throughout the bacterial chromosome or located in plasmids; this possibility is under investigation. Additional genes could be a reservoir for gene duplication, mutation, or recombination.

Several observations are consistent with the existence of subpopulations in *T. pallidum*. The natural history of syphilis (latent infection interrupted by episodes of active disease), the inability of some *T. pallidum* to be opsonized for phagocytosis (23), and resistance to high titers of specific antibodies in vivo during secondary syphilis may be explained by genetic mechanisms such as phase and antigenic variation of the TprK antigens. Alternatively, tissue tropism, such as neuroinvasive capacity, could originate from the variable regions of surface-exposed TprK antigens. Lastly, cross-immunity studies have demonstrated that infection with a particular *T. pallidum* isolate will induce complete protection only against reinfection with the homologous strain, while protection against other isolates is absent or partial (31) (S. A. Lukehart et al., unpublished data). This observation suggests that there is heterogeneity among the protective antigens of different isolates, which might be explained in part by the variability of TprK.

It is intriguing that there is only a single-copy *tprK* gene in the Nichols strain (16), whereas all other examined isolates from our treponemal strain bank have multiple *tprK* sequences. Because the Nichols strain has been maintained by passage in rabbits for more than 80 years, it may have adapted for survival in rabbit tissues by expressing only a single TprK molecule. Alternatively, multiple intratesticular passages may have led to selection of a single clone from an original population with differing TprK molecules. The other isolates that we examined have been passed in rabbits only a few times relative to the Nichols strain.

It is noteworthy that there are sequence differences in the *tprK* gene at both the DNA and predicted protein level between two Nichols strains maintained in different laboratories. These two Nichols "strains" were both originally obtained from James Miller at the University of California at Los Angeles in 1978 (to La Jolla and then to Houston) and 1979 (to Seattle) and have been propagated separately since that time. Although highly homologous, there are some reproducible differences between the two Nichols strains (Fig. 4), suggesting genetic drift in the *tprK* sequence or the presence of different subpopulations bearing different *tprK* types that may have separately become predominant in the two settings. A particularly striking difference is the finding of a 67-nucleotide deletion in the 3'-flanking region of the Seattle Nichols strain (Fig. 5) compared with the Nichols strain used by the genome project. This finding not only strongly supports the hypothesis that our Nichols strain has diverged from the Houston strain but also may have other implications in terms of gene expression. As described above, there are two putative Rho-dependent transcription termination hairpin structures downstream of the *tprK* stop codon. The deletion in the 3'-flanking region removes portions of these hairpin structures. If either of these hairpins is a Rho-dependent transcription termination site, this may influence *tprK* gene transcription, as well as the transcription of genes 3' of *tprK*. The functions of the genes 3' of *tprK* are unknown, but their transcription could influence the function or expression of *tprK*, as flanking genes do in other operons.

This study describes the identification of multiple, heterogeneous *tprK* genes in non-Nichols *T. pallidum* strains. These

*tprK* genes potentially encode surface-exposed variable proteins as predicted by their very high DNA and amino acid homology with the Nichols TprK antigen, which has been functionally identified as a surface-exposed molecule because it is the target of opsonic antibodies in intact treponemes. These findings invite new research that may lead to a greater understanding of the mechanisms of syphilis pathogenesis and, particularly, immune evasion. The precise functional significance of the multiple variable Tpr K antigens in *T. pallidum* and the molecular mechanisms that generate *tprK* diversity are still unclear.

#### ACKNOWLEDGMENTS

We thank Sally Post for manuscript preparation and Lynn Barrett for confirmatory independent sequencing of *tprK* from the Nichols strain.

This study was supported by Public Health Service grants AI42143 and AI34616 from the National Institutes of Health (S.A.L.) and Sexually Transmitted Diseases Cooperative Research Center New Investigator Award AI31448 from the National Institutes of Health (A.C.-L.).

#### REFERENCES

- Akins, D. R., E. Robinson, D. Shevchenko, C. Elkins, D. L. Cox, and J. D. Radolf. 1997. Tromp1, a putative rare outer membrane protein, is anchored by an uncleaved signal sequence to the *Treponema pallidum* cytoplasmic membrane. *J. Bacteriol.* **179**:5076–5086.
- Baker-Zander, S. A., and S. A. Lukehart. 1992. Macrophage-mediated killing of opsonized *Treponema pallidum*. *J. Infect. Dis.* **165**:69–74.
- Blanco, D. R., C. I. Champion, M. M. Exner, E. S. Shang, J. T. Skare, R. E. Hancock, J. N. Miller, and M. A. Lovett. 1996. Recombinant *Treponema pallidum* rare outer membrane protein 1 (Tromp1) expressed in *Escherichia coli* has porin activity and surface antigenic exposure. *J. Bacteriol.* **178**:6685–6692.
- Blanco, D. R., C. I. Champion, M. A. Lewinski, E. S. Shang, S. G. Simkins, J. N. Miller, and M. A. Lovett. 1999. Immunization with *Treponema pallidum* outer membrane vesicles induces high-titer complement-dependent treponemocidal activity and aggregation of *T. pallidum* rare outer membrane proteins (TROMPs). *J. Immunol.* **163**:2741–2746.
- Blanco, D. R., J. N. Miller, and M. A. Lovett. 1997. Surface antigens of the syphilis spirochete and their potential as virulence determinants. *Emerg. Infect. Dis.* **3**:11–20.
- Blanco, D. R., E. M. Walker, D. A. Haake, C. I. Champion, J. N. Miller, and M. A. Lovett. 1990. Complement activation limits the rate of in vitro treponemocidal activity and correlates with antibody-mediated aggregation of *Treponema pallidum* rare outer membrane protein. *J. Immunol.* **144**:1914–1921.
- Cameron, C. E., C. Castro, S. A. Lukehart, and W. C. Van Voorhis. 1998. Function and protective capacity of *Treponema pallidum* subsp. *pallidum* glycerophosphodiester phosphodiesterase. *Infect. Immun.* **66**:5763–5770.
- Centurion-Lara, A., C. Castro, L. Barrett, C. Cameron, M. Mostowfi, W. C. Van Voorhis, and S. A. Lukehart. 1999. *Treponema pallidum* major sheath protein homologue TprK is a target of opsonic antibody and the protective immune response. *J. Exp. Med.* **189**:647–656.
- Centurion-Lara, A., C. Castro, R. Castillo, J. M. Shaffer, W. C. Van Voorhis, and S. A. Lukehart. 1998. The flanking region sequences of the 15-kDa lipoprotein gene differentiate pathogenic treponemes. *J. Infect. Dis.* **177**:1036–1040.
- Champion, C. I., D. R. Blanco, M. M. Exner, H. Erdjument-Bromage, R. E. Hancock, P. Tempst, J. N. Miller, and M. A. Lovett. 1997. Sequence analysis and recombinant expression of a 28-kilodalton *Treponema pallidum* subsp. *pallidum* rare outer membrane protein (Tromp2). *J. Bacteriol.* **179**:1230–1238.
- Cox, D. L., D. R. Akins, S. F. Porcella, M. V. Norgard, and J. D. Radolf. 1995. *Treponema pallidum* in gel microdroplets: a novel strategy for investigation of treponemal molecular architecture. *Mol. Microbiol.* **15**:1151–1164.
- Cox, D. L., P. Chang, A. W. McDowall, and J. D. Radolf. 1992. The outer membrane, not a coat of host proteins, limits antigenicity of virulent *Treponema pallidum*. *Infect. Immun.* **60**:1076–1083.
- Deka, R. K., Y. H. Lee, K. E. Hagman, D. Shevchenko, C. A. Lingwood, C. A. Hasemann, M. V. Norgard, and J. D. Radolf. 1999. Physicochemical evidence that *Treponema pallidum* TroA is a zinc-containing metalloprotein that lacks porin-like structure. *J. Bacteriol.* **181**:4420–4423.
- Egli, C., W. K. Leung, K. H. Muller, R. E. Hancock, and B. C. McBride. 1993. Pore-forming properties of the major 53-kilodalton surface antigen from the outer sheath of *Treponema denticola*. *Infect. Immun.* **61**:1694–1699.
- Fenno, J. C., K. H. Muller, and B. C. McBride. 1996. Sequence analysis,

- expression, and binding activity of recombinant major outer sheath protein (Msp) of *Treponema denticola*. *J. Bacteriol.* **178**:2489–2497.
16. Fraser, C. M., S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M. D. Cotton, J. C. Venter, et al. 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**:375–388.
  17. Hardham, J. M., L. V. Stamm, S. F. Porcella, J. G. Frye, N. Y. Barnes, J. K. Howell, S. L. Mueller, J. D. Radolf, G. M. Weinstock, and S. J. Norris. 1997. Identification and transcriptional analysis of a *Treponema pallidum* operon encoding a putative ABC transport system, an iron-activated repressor protein homolog, and a glycolytic pathway enzyme homolog. *Gene* **197**:47–64.
  18. Lee, Y. H., R. K. Deka, M. V. Norgard, J. D. Radolf, and C. A. Hasemann. 1999. *Treponema pallidum* TroA is a periplasmic zinc-binding protein with a helical backbone. *Nat. Struct. Biol.* **6**:628–633.
  19. Lewinski, M. A., J. N. Miller, M. A. Lovett, and D. R. Blanco. 1999. Correlation of immunity in experimental syphilis with serum-mediated aggregation of *Treponema pallidum* rare outer membrane proteins. *Infect. Immun.* **67**:3631–3636.
  20. Lukehart, S. A., S. A. Baker-Zander, R. M. Lloyd, and S. Sell. 1980. Characterization of lymphocyte responsiveness in early experimental syphilis. II. Nature of cellular infiltration and *Treponema pallidum* distribution in testicular lesions. *J. Immunol.* **124**:461–467.
  21. Lukehart, S. A., S. A. Baker-Zander, R. M. C. Lloyd, and S. Sell. 1992. Immunology and pathogenesis of syphilis, p. 141–163. In T. C. Quinn, J. I. Gallin, and A. S. Fauci (ed.), *Sexually transmitted diseases*, vol. 8. Raven Press, New York, N.Y.
  22. Lukehart, S. A., S. A. Baker-Zander, and S. Sell. 1980. Characterization of lymphocyte responsiveness in early experimental syphilis. I. In vitro response to mitogens and *Treponema pallidum* antigens. *J. Immunol.* **124**:454–460.
  23. Lukehart, S. A., J. M. Shaffer, and S. A. Baker-Zander. 1992. A subpopulation of *Treponema pallidum* is resistant to phagocytosis: possible mechanism of persistence. *J. Infect. Dis.* **166**:1449–1453.
  24. Norris, S. J., J. F. Alderete, N. H. Axelsen, M. J. Bailey, S. A. Baker-Zander, J. B. Baseman, P. J. Bassford, R. E. Baughn, A. Cockayne, P. A. Hanff, P. Hindersson, S. A. Larsen, M. A. Lovett, S. A. Lukehart, J. N. Miller, M. A. Moskopfidis, F. Muller, M. V. Norgard, C. W. Penn, L. V. Stamm, J. D. van Embden, and K. Wicher. 1987. Identity of *Treponema pallidum* subsp. *pallidum* polypeptides: correlation of sodium dodecyl sulfate-polyacrylamide gel electrophoresis results from different laboratories. *Electrophoresis* **8**:77–92.
  25. Radolf, J. D. 1994. Role of outer membrane architecture in immune evasion by *Treponema pallidum* and *Borrelia burgdorferi*. *Trends Microbiol.* **2**:307–311.
  26. Radolf, J. D. 1995. *Treponema pallidum* and the quest for outer membrane proteins. *Mol. Microbiol.* **16**:1067–1073.
  27. Radolf, J. D., and M. V. Norgard. 1988. Pathogen specificity of *Treponema pallidum* subsp. *pallidum* integral membrane proteins identified by phase partitioning with Triton X-114. *Infect. Immun.* **56**:1825–1828.
  28. Radolf, J. D., M. V. Norgard, and W. W. Schulz. 1989. Outer membrane ultrastructure explains the limited antigenicity of virulent *Treponema pallidum*. *Proc. Natl. Acad. Sci. USA* **86**:2051–2055.
  29. Shevchenko, D. V., T. J. Sellati, D. L. Cox, O. V. Shevchenko, E. J. Robinson, and J. D. Radolf. 1999. Membrane topology and cellular location of the *Treponema pallidum* glycerophosphodiester phosphodiesterase (GlpQ) ortholog. *Infect. Immun.* **67**:2266–2276.
  30. Stamm, L. V., S. R. Greene, H. L. Bergen, J. M. Hardham, and N. Y. Barnes. 1998. Identification and sequence analysis of *Treponema pallidum* tprJ, a member of a polymorphic multigene family. *FEMS Microbiol. Lett.* **169**:155–163.
  31. Turner, T. B., and D. H. Hollander. 1957. *Biology of the treponematoses*. World Health Organization, Geneva, Switzerland.
  32. Walker, E. M., G. A. Zampighi, D. R. Blanco, J. N. Miller, and M. A. Lovett. 1989. Demonstration of rare protein in the outer membrane of *Treponema pallidum* subsp. *pallidum* by freeze-fracture analysis. *J. Bacteriol.* **171**:5005–5011.

Editor: D. L. Burns