JOURNAL OF BACTERIOLOGY, June 1996, p. 3177–3187 Vol. 178, No. 11
0021-9193/96/$04.00+0

Characterization of the Cytoplasmic Filament Protein Gene (*cfpA*) of *Treponema pallidum* subsp. *pallidum*

YUN YOU,[1] STEPHANIE ELMORE,[1] LISA L. COLTON,[1] CHRIS MACKENZIE,[2] JAMES K. STOOPS,[1] GEORGE M. WEINSTOCK,[2,3] AND STEVEN J. NORRIS[1,3]*

Departments of Pathology and Laboratory Medicine,[1] of Biochemistry and Molecular Biology,[2] and of Microbiology and Molecular Genetics,[3] University of Texas Medical School at Houston, Houston, Texas

Received 2 August 1995/Accepted 28 March 1996

Treponema pallidum **and other members of the genera *Treponema*, *Spirochaeta*, and *Leptonema* contain multiple cytoplasmic filaments that run the length of the organism just underneath the cytoplasmic membrane. These cytoplasmic filaments have a ribbon-like profile and consist of a major cytoplasmic filament protein subunit (CfpA, formerly called TpN83) with a relative molecular weight of ~80,000. Degenerate DNA primers based on N-terminal and CNBr cleavage fragment amino acid sequences of *T. pallidum* subsp. *pallidum* (Nichols) CfpA were utilized to amplify a fragment of the encoding gene (*cfpA*). A 6.8-kb *Eco*RI fragment containing all but the 5′ end of *cfpA* was identified by hybridization with the resulting PCR product and cloned into Lambda ZAP II. The 5′ region was obtained by inverse PCR, and the complete gene sequence was determined. The *cfpA* sequence contained a 2,034-nucleotide coding region, a putative promoter with consensus sequences (5′-TTTACA-3′ for −35 and 5′-TACAAT-3′ for −10) similar to the σ^{70} recognition sequence of *Escherichia coli* and other organisms, and a putative ribosome-binding site (5′-AGGAG-3′). The deduced amino acid sequence of CfpA indicated a protein of 678 residues with a calculated molecular mass of 78.5 kDa and an estimated pI of 6.15. No significant homology to known proteins or structural motifs was found among known prokaryotic or eukaryotic sequences. Expression of a LacZ-CfpA fusion protein in *E. coli* was detrimental to survival and growth of the host strain and resulted in the formation of short, irregular filaments suggestive of partial self-assembly of CfpA. The cytoplasmic filaments of *T. pallidum* and other spirochetes appear to represent a unique form of prokaryotic intracytoplasmic inclusions.**

Spirochetes possess a number of unusual structural features, including helical or flat plane-wave shape and periplasmic flagella that tightly entwine the cell body and are encased by the outer membrane (13–16, 19, 31, 41). *T. pallidum* differs from other spirochetes in that it contains relatively few intramembranous protein particles in its outer membrane, as determined by freeze-fracture electron microscopy (33, 40). Seven of eight *Treponema* species examined, some *Spirochaeta* species, and *Leptonema illini* have also been shown to contain cytoplasmic filaments (also called cytoplasmic tubules and cytoplasmic fibrils) (14). Cytoplasmic filaments are ribbon-like structures 7.0 to 7.5 nm wide that run the length of the organism (5, 13–16, 19, 25, 31, 41). An array of 4 to 6 filaments lie in close apposition to the inner membrane and are always localized directly underneath the corresponding group of periplasmic flagella. The function of the cytoplasmic filaments is unknown. Their location indicates that they may be involved in cell motility, although other functions related to helical morphology or cell division are also possible (5).

Masuda and Kawata (25) purified the cytoplasmic filaments of *Treponema denticola* TD-2, *Treponema phagedenis* biotypes Reiter and Kazan, three *Treponema* sp. strains (E-21, E-30, and Y-181) isolated from the oral cavities of patients with pyorrhea, and *L. illini*. In each case, the major constituent of the filaments was an 82-kDa protein. The results of immunoblotting and immunoelectron microscopy suggested that the major polypeptide of the filaments is a genus-specific antigen for *Treponema*. A major 82- to 83-kDa protein of the cytoplasmic filament of *L. illini* did not react with antiserum against the 82-kDa filament protein from strain E-21 by immunoblotting analysis, indicating antigenic divergence.

Previous studies of cytoplasmic filaments involved the characterization of filament structure, cellular location, and protein components. No information has been available regarding the cytoplasmic filament protein gene or the functions of the filament. This study focused on cloning and characterization of the cytoplasmic filament protein gene (*cfpA*) of *Treponema pallidum* subsp. *pallidum* Nichols.

MATERIALS AND METHODS

Bacterial strains and plasmids. *T. pallidum* subsp. *pallidum* Nichols was passaged by intratesticular infection in rabbits. The organisms were extracted from infected rabbit testes and purified by Percoll density gradient centrifugation (6). *T. phagedenis* Kazan 5 was cultured as previously described (23). *Escherichia coli* XL1-Blue MRF′ strain (Stratagene, La Jolla, Calif.), which carries Δ(*mcrA*)*183* Δ(*mcrCB-hsdSMR-mrr*)*173 endA1 supE44 thi-1 recA1 gyrA96 relA1 lac* [F′ *proAB lacI*^{q}*Z*Δ*M15* Tn*10*(Tet^r)], was used for propagation and preparation of Lambda ZAP II (Stratagene) and plasmid DNA. *E. coli* SOLR strain (Stratagene), which carries *e14-*(*mcrA*) Δ(*mcrCB-hsdSMR-mrr*)*171 sbcC recB recJ uvrC umuC*::Tn*5* (Kan^r) *lac gyrA96 relA1 thi-1 endA1 lR* [F′ *proAB lacI*^{q}*Z*Δ*M15*] *Su* (nonsuppressing), was used for excision of the pBluescript SK(−) phagemid from Lambda ZAP II.

Electron microscopy. For visualization of cytoplasmic filaments, freshly extracted *T. pallidum* was centrifuged, resuspended in sterile distilled water, stored at 4°C for ≥24 h, and drawn repeatedly through a 25-gauge needle and syringe. The resulting suspension of lysed organisms was placed on nitrocellulose-coated copper grids and negatively stained by the drop method with 0.25% methylamine tungstate prior to observation with a JEM 1200 EXII electron microscope (JEOL, Peabody, Mass.) at an accelerating voltage of 100 kV. Photographs were taken with an underfocus of ~1 μm and a 1-s exposure on S0163 film (Kodak, Rochester, N.Y.). *E. coli* cells were prepared for electron microscopy of ultrathin sections by a modified version (6) of the method of Lickfield (21). Energy-

* Corresponding author. Mailing address: Dept. of Pathology and Laboratory Medicine, University of Texas Medical School at Houston, P.O. Box 20708, Houston, TX 77225. Phone: (713) 792-5212. Fax: (713) 794-4149. Electronic mail address: (Internet) norr@casper.med.uth.tmc.edu.

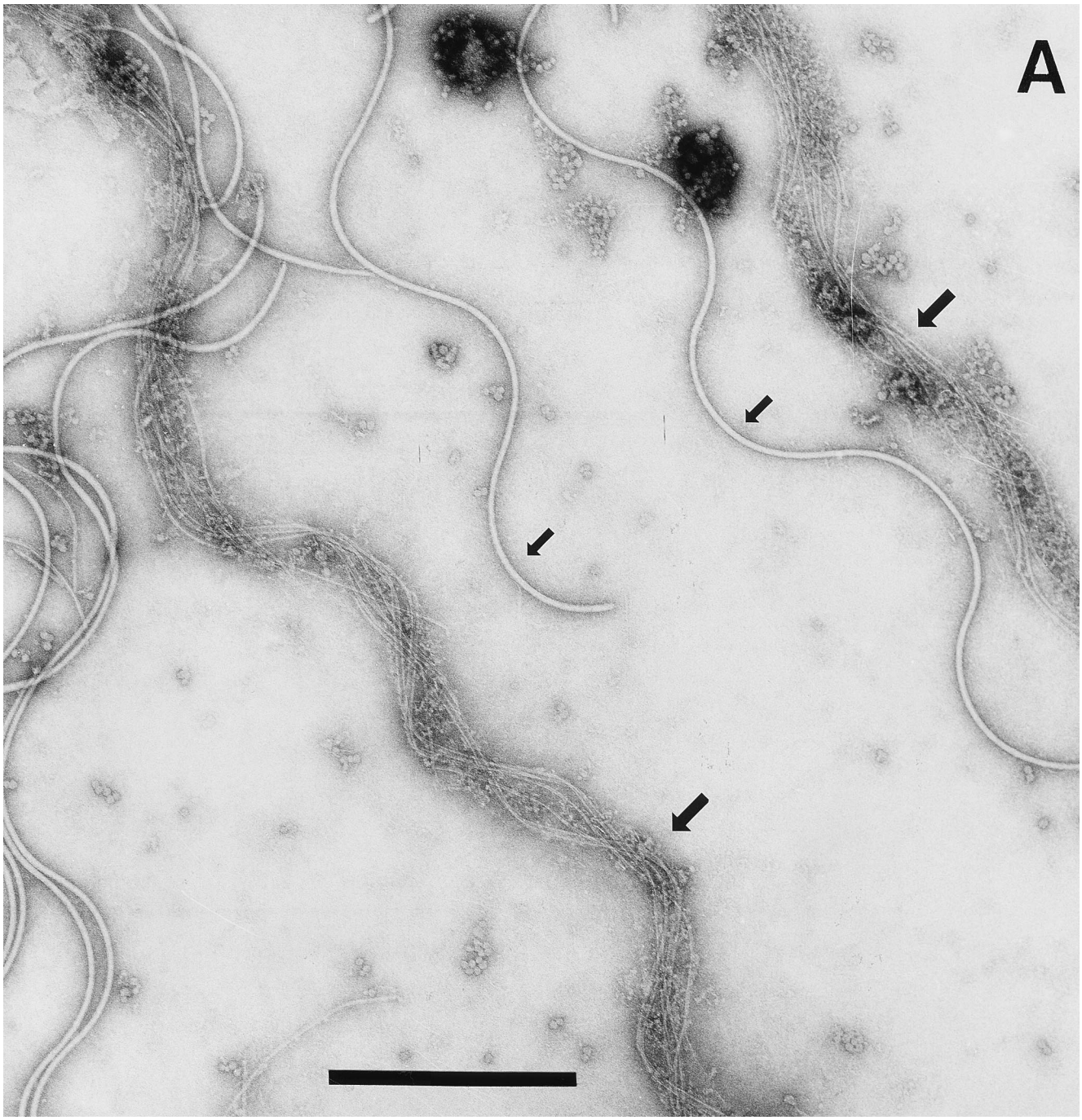

FIG. 1. Electron micrographs of the cytoplasmic filaments of *T. pallidum*, as revealed by negative staining of organisms disrupted in distilled water. (A) Bundles of cytoplasmic filaments (large arrows) exhibit approximately the same periodicity as intact cells (not shown) and flagella (small arrows). Bar = 500 nm. (B) Detailed view of a bundle of six *T. pallidum* cytoplasmic filaments. Individual filaments are twisted along their axes, showing both the wide and narrow dimensions (arrowheads) of their ribbon-like structure. Bar = 100 nm. (C) Cytoplasmic filament fragments (CF) partially purified from *T. phagedenis* Kazan 5 by the CsCl gradient sedimentation method of Masuda and Kawata (25). Note the presence of cosedimenting flagellar basal bodies (BB) and basal plates (BP). Cytoplasmic filaments appeared to be attached to the periphery of the basal plates in some instances (arrow). Bar = 100 nm.

dispersive X-ray spectroscopy of these specimens was kindly performed by Robert Harris in the laboratory of T. J. Beveridge, University of Guelph, Guelph, Ontario, Canada.

Protein sequencing. The cytoplasmic filaments of *T. phagedenis* were purified by the procedure of Masuda and Kawata (25). The major 80-kDa constituent of the *T. phagedenis* cytoplasmic filaments and the corresponding 80-kDa band in *T. pallidum* were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), transferred to a polyvinylidene difluoride membrane (Millipore, Bedford, Mass.), and stained with Coomassie blue R as described by Matsudaira (26). CNBr cleavage fragments of the *T. pallidum* polypeptide were prepared in a similar manner (26, 27). Selected bands were carefully excised from the polyvinylidene difluoride membrane and subjected to Edman degradation. N-terminal sequences were determined by Richard G. Cook at the Baylor College of Medicine Protein Chemistry Core Facility (Houston, Tex.) with a model 477A sequenator (Applied Biosystems, Foster City, Calif.).

Preparation of genomic DNA of *T. pallidum*. *T. pallidum* cells ($\sim 10^{10}$) resus-

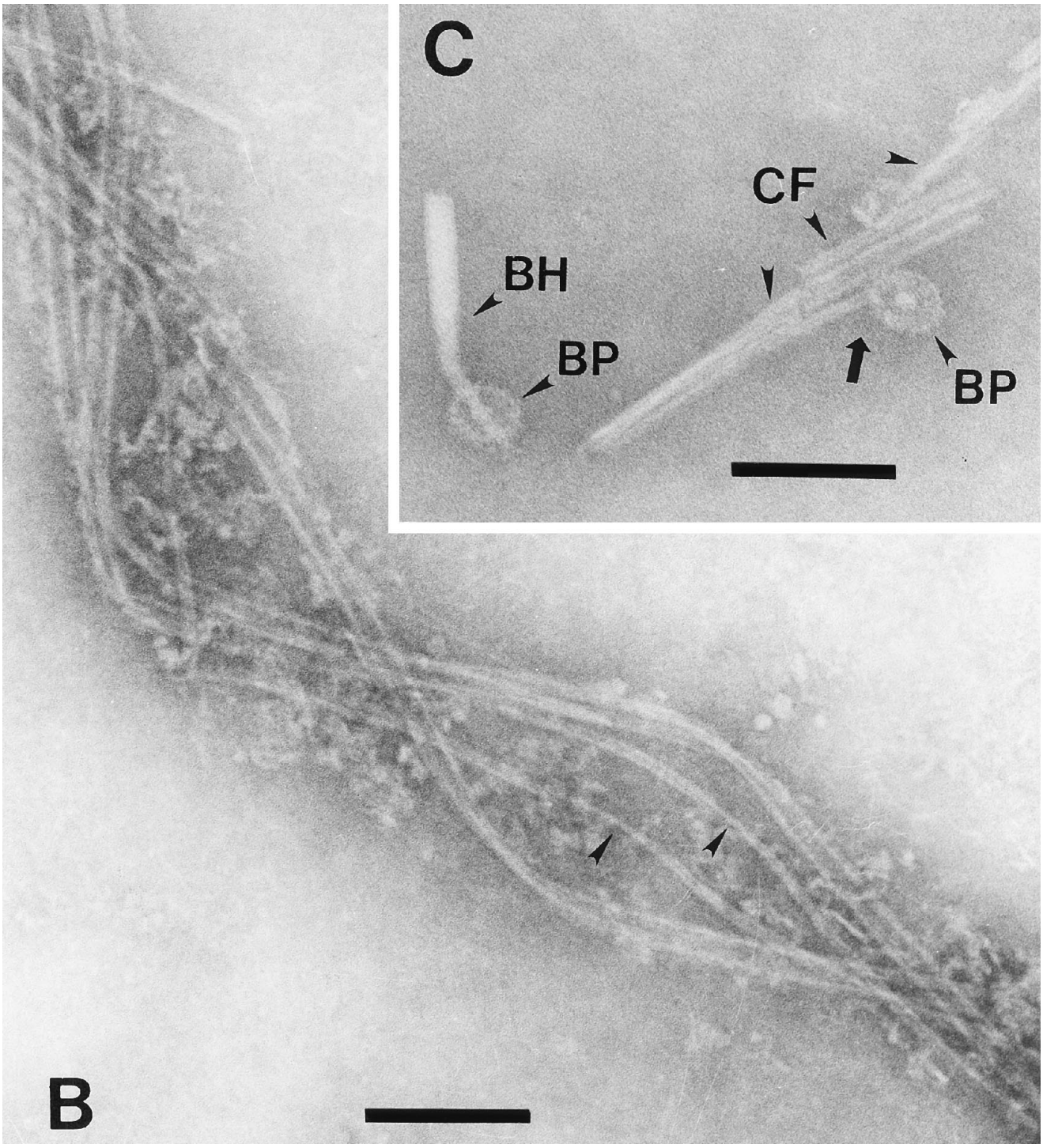

FIG. 1—*Continued.*

pended in TE-SDS solution (0.5% SDS, 9.5 mM Tris-HCl [pH 8.0], 0.095 mM EDTA [pH 8.0]) were incubated at 37°C for 1 to 4 h and then incubated with 80 to 100 μg of proteinase K per ml at 37°C overnight. Samples with 1/20 volume of 8 M ammonium acetate were extracted with phenol-chloroform two or three times and then with chloroform once. The DNA was precipitated by the addition of ethanol and centrifugation. The pellet was resuspended in TE buffer (10 mM Tris-HCl [pH 8.0], 0.1 mM EDTA [pH 8.0]).

PCR. Two degenerate oligonucleotide primers, cfp-N (5′-TTPyCAPyCC NGAPuAAPuCC-3′), and cfp-CNBr (5′-CCPyTGNAPuPuTGNCCPuTAG-3′), were constructed on the basis of amino acid sequences obtained from the CfpA N terminus and CNBr cleavage fragments, respectively. A 424-bp segment of *cfpA* was amplified from *T. pallidum* genomic DNA by using these primers in combination with PCR core reagents and a DNA thermal cycler according to the manufacturer's instructions (Perkin-Elmer Cetus, Norwalk, Conn.). This segment was cloned into the PCR1000 plasmid vector (Invitrogen, San Diego, Calif.) and utilized for partial sequence analysis. A 315-bp amplification product from this region was obtained with forward (5′-AGCAGCAGGAGGTGAAC CAGCTA-3′) and reverse (5′-TCTCCGGTGTTGAACTTATCC-3′) primers and utilized as a probe for identification of genomic DNA segments containing the *cfpA* sequence.

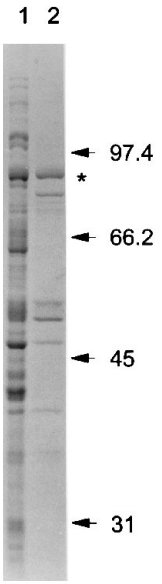

FIG. 2. SDS-PAGE analysis of a partially purified preparation of *T. phagedenis* Kazan 5 cytoplasmic filaments. Lane 1, whole *T. phagedenis* Kazan 5; lane 2, cytoplasmic filament preparation. The relative positions of molecular mass standards (in kilodaltons) are indicated to the right of the gel, and the band with an M_r of 80,000 that represents the major constituent of the cytoplasmic filaments is marked by an asterisk.

Southern blotting. Genomic or plasmid DNA samples digested with a restriction enzyme(s) were electrophoresed on 0.8% agarose gels, transferred onto nylon membranes (Hybond-N+; Amersham, Bedford, Mass.) by capillary transfer with 0.4 N NaOH as the solvent (34, 39). The blots were neutralized in 0.5 M Tris (pH 7.0) for 5 min, rinsed briefly in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), and cross-linked with a Stratalinker UV Crosslinker (Stratagene) under the autocross-link setting. Hybridizations were carried out as described previously (39) utilizing probes labelled with [α-^{32}P]dCTP by random primer extension (Multiprime Kit; Amersham, Arlington Heights, Ill.). Autoradiography was performed with XAR film (Eastman Kodak, New Haven, Conn.) and intensifying screens at −70°C.

Construction and characterization of genomic libraries. Genomic DNA extracted from *T. pallidum* was completely digested with the restriction enzyme *Eco*RI. A library of the unfractionated *Eco*RI fragments was constructed in *Eco*RI-predigested and dephosphorylated Lambda ZAP II vector (Stratagene) and screened by hybridization with a 315-bp PCR product, following the manufacturer's instructions. pBluescript SK(−) phagemids containing *T. pallidum* DNA inserts were obtained from selected Lambda ZAP II clones by using the ExAssist/SOLR in vivo excision system (Stratagene).

A 2.5-kb *Eco*RI-*Sac*I subclone and a 2.0-kb *Eco*RI-*Sal*I subclone in pBluescript SK(−) were utilized for sequence analysis. Separate sets of nested deletion clones were generated from the 5′ end of the *Eco*RI-*Sac*I insert and the 3′ end of the *Eco*RI-*Sal*I insert by unidirectional digestion with exonuclease III (11). The sequences of both strands were determined with an Applied Biosystems model 373A automated sequencer with T3, T7, and (for some regions) internal oligonucleotide primers.

A 2.1-kb DNA fragment containing the promoter region and the additional 71-bp 5′ region of *cfpA* was amplified by inverse PCR as described previously (18, 37) with a circularized 3.1-kb *Kpn*I fragment from genomic DNA as the template. *T. pallidum* genomic DNA was completely digested with *Kpn*I, phenol-chloroform extracted twice, and ethanol precipitated. *Kpn*I-digested DNA was circularized in 100 μl of ligation reaction mixture containing 7 ng of DNA per μl, 7 μl of 10× ligation buffer (0.66 M Tris-Cl [pH 7.5], 50 mM $MgCl_2$, 50 mM dithiothreitol), 1 mM ATP, and 1 U of T4 DNA ligase (Promega). After incubation overnight at room temperature, DNA was precipitated and resuspended in H_2O for inverse PCR. Two primers, cfpE50R (5′-GCTGGTTCACCTC CTGC-3′) and cfpE954F (5′-CAATTCGATTACATCTATCC-3′), were used for PCR. The inverse-PCR mixture contained 700 ng of *Kpn*I-digested genomic DNA, 1× PCR buffer II, 1.5 mM $MgCl_2$, 20 pmol of each primer, 200 μM (each) deoxynucleoside triphosphate, and 2.5 U of *Taq* polymerase in a final volume of 100 μl (reagents from the PCR kit of Perkin-Elmer Cetus). PCR was performed as follows: (i) denaturation at 94°C for 2 min; (ii) 30 cycles, with 1 cycle consisting of denaturation at 94°C for 40 s, annealing at 54°C for 40 s, and extension at 72°C for 3 min; and (iii) a final extension step at 72°C for 10 min. PCR products were analyzed by electrophoresis on agarose gels. The amplified 2.1-kb fragment was purified by using the Magic PCR Prep System (Promega, Madison, Wis.) and eluted in 50 μl of H_2O. A second PCR amplification was performed for 34 cycles by using the same primers and 5 μl of the purified PCR fragment as the template. The resulting PCR product was purified as described above, and the sequence was determined by using the primer cfpE50R. On the basis of this sequence, a primer was designed to obtain the sequence of the complementary strand.

Sequence similarity analyses were performed at the National Center for Biotechnology Information (Bethesda, Md.) using the BLAST network service. Other sequence analysis programs used were part of the Wisconsin Package version 8 (Genetics Computer Group, Madison, Wis.). The COILS program (24) for analyzing coiled-coil structure was accessed through the ExPASy Molecular Biology Server, at the URL http://ulrec3.unil.ch/software/COILS_form.html.

Expression of a LacZ-CfpA fusion protein in *E. coli*. The *cfpA* insert in the ZAPE2-3 clone is in frame of the N-terminal region of the *lacZ* gene in pBluescript SK(−). *E. coli* XL1-Blue MRF′ containing pBluescript SK(−), pZAPE2-3, or pZAPE2-4 were inoculated in Luria-Bertani (LB) broth with 50 μg of ampicillin per ml and incubated overnight at 37°C. Fresh cultures were made by inoculating 100 μl of the cultures grown overnight in 3 ml of LB medium containing ampicillin. After growing at 37°C to log phase (optical density at 600 nm of 0.4 to 1.0), cells were induced by adding 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubating 3 h at 37°C. A portion of each culture was fixed in 1% glutaraldehyde and processed for transmission electron microscopy as described previously (6, 21). For SDS-PAGE, cultures were centrifuged and resuspended in a solution consisting of 50 mM glucose, 25 mM Tris-Cl (pH 8.0), and 10 mM EDTA (pH 8.0). Samples were disrupted by sonication and the addition of an equal volume of solubilization buffer prior to electrophoresis in 8% polyacrylamide gels (29).

Nucleotide sequence accession number. The nucleotide sequence accession number of the *cfpA* region has been deposited in GenBank under accession number U32683.

RESULTS

Structure of cytoplasmic filaments. Cytoplasmic filaments in spirochetes have been described previously as tubules with lumens, cylindrical fibrils, and ribbon-shaped filaments. To determine the structural morphology of the cytoplasmic filaments of *T. pallidum*, organisms were subjected to lysis in distilled water and storage at 4°C for 1 to 7 days. Under these conditions, the membranes and cell wall were disrupted and the cytoplasmic filaments and flagella were released in an intact state. Bundles of cytoplasmic filaments retained the helical conformation which is presumed to occur in intact cells, on the basis of their constant location underneath the helically distributed periplasmic flagella (Fig. 1). The individual filaments had a ribbon-like configuration, as described previously by Eipert and Black (5) for the cytoplasmic filaments of *Treponema refringens*. The *T. pallidum* filaments had a cross-section of ~7.0 nm to 7.5 nm by ~1 nm.

Attempts to purify *T. pallidum* cytoplasmic filaments in sufficient quantities for protein analysis were unsuccessful. However, the cytoplasmic filaments of *T. phagedenis* Kazan 5 were partially purified by the CsCl gradient method of Masuda and Kawata (25). This preparation was greatly enriched for cytoplasmic filaments with the same ribbon-like structure as found with the *T. pallidum* filaments; however, it also contained flagellar basal bodies and basal plates (Fig. 1C). These basal plates are circular structures 40 to 45 nm in diameter that are attached to the end of the flagellar basal bodies. Basal plates

```
a) N-terminal amino acid sequences

      T. pallidum       1 ASLDLPKSPNVFHPEKPSAVG 21
                          ¦ ¦:¦¦.¦¦¦¦¦¦¦¦¦¦¦¦¦
      T. phagedenis     1 AILELPQSPNVFHPEKPSAVG 21

b)  CNBr cleavage fragment sequence

      T. pallidum       1 MYGHLQGHIQRGVNEL 16
```

FIG. 3. Partial amino acid sequences of the major cytoplasmic filament proteins (CfpA) of *T. pallidum* and *T. phagedenis* Kazan 5, as determined by Edman degradation. Identical amino acids are indicated by vertical lines, and similar amino acids are indicated by a colon or period.

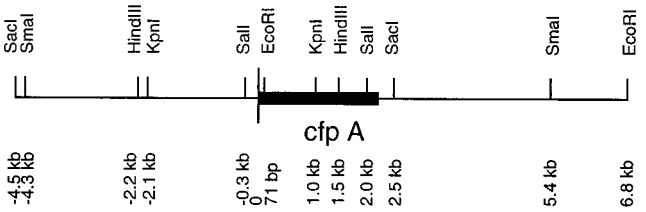

FIG. 4. Restriction map of the *T. pallidum cfpA* gene and the surrounding region. Map locations are relative to the beginning of the open reading frame of *cfpA*.

with cytoplasmic filaments attached to knob-like projections at their periphery (as described previously by Hovind-Hougen [14] and Bermudes et al. [1]) were observed occasionally (Fig. 1C).

The major constituent of the purified *T. phagedenis* filaments was an 80-kDa polypeptide (Fig. 2), similar to cytoplasmic filament subunit proteins described by Masuda and Kawata (25) for *T. phagedenis* and several other spirochetes. The single- and two-dimensional gel migration characteristics of this protein are very similar to those of the *T. pallidum* polypeptide previously described as TpN83, which comprises 3 to 5% of the total cellular protein of *T. pallidum* (29, 32). Other bands present in the *T. phagedenis* cytoplasmic filament preparation may represent degraded or otherwise modified forms of the 80-kDa protein (32), flagellar basal body or basal plate constituents, or other polypeptides associated with the cytoplasmic filaments.

The N-terminal sequences of TpN83 and the 80-kDa protein from purified *T. phagedenis* cytoplasmic filaments were compared (Fig. 3). The first 21 amino acids of the N-terminal sequences of these proteins exhibited 90.5% similarity and 85.7% sequence identity. On the basis of this similarity and morphologic evidence, these proteins were considered the major cytoplasmic filament protein of the two species and were designated CfpA. In addition to the N terminus, the partial sequence of a CNBr cleavage fragment of *T. pallidum* CfpA (Fig. 3) was determined for use in designing oligonucleotides for amplification of a portion of the *cfpA* gene.

Molecular cloning of *cfpA*. Two degenerate oligonucleotide primers based on the amino acid sequences, cfp-N and cfp-CNBr, were utilized for PCR amplification and partial DNA sequence determination. A 424-bp PCR product was amplified from *T. pallidum* Nichols genomic DNA and cloned into the PCR1000 vector. DNA sequence analysis of one of the resulting clones revealed a single open reading frame with a deduced amino acid sequence that matched precisely the *T. pallidum* N-terminal and CNBr amino acid sequences (including those regions not encoded by the PCR primers). In addition, the PCR product contained an *Eco*RI site near the 5′ end. A 315-bp internal amplification product from this region was used in subsequent Southern blot analyses and library screenings.

The estimated size of *cfpA* was approximately 2.0 to 2.2 kb. The 315-bp probe hybridized to a 6.8-kb *Eco*RI fragment, a 3.1-kb *Kpn*I fragment, and a 9.7-kb *Sma*I fragment; on the basis of single and double digestions, a provisional restriction map of the region was obtained (Fig. 4). We were unable to clone any fragments containing the intact *cfpA* gene with a variety of restriction digestions and vectors. The sequence of the PCR product indicated that an *Eco*RI site was located within the open reading frame near the 5′ end of the *cfpA*; thus, the 6.8-kb *Eco*RI fragment contained most of the gene but lacked the promoter and N-terminal coding region. Therefore, a *Eco*RI library of *T. pallidum* genomic DNA in Lambda ZAP II was screened by hybridization with the 315-bp probe. Twelve of ~1,000 plaques screened hybridized with the probe and were selected for in vivo excision of recombinant pBluescript phagemids. The resulting clones contained the 6.8-kb *Eco*RI fragment in either orientation, and representative clones (pZAPE2-3 and pZAPE2-4) were selected for further analysis. A restriction map of the 6.8-kb insert was constructed by restriction enzyme digestion (Fig. 4). On the basis of this map, a 2.5-kb *Eco*RI-*Sac*I subclone and a 2.0-kb *Eco*RI-*Sal*I subclone containing most of *cfpA* were obtained and subjected to complete sequence analysis of both DNA strands by a nested deletion technique.

Attempts to clone restriction fragments containing the 5′-coding region and promoter of *cfpA* were unsuccessful. Southern blot hybridization and sequence analysis indicated that a 3.1-kb *Kpn*I fragment contained the 5′ end of the gene which had not been cloned or sequenced, as well as ~1.0 kb of the sequenced region. Inverse PCR (37) was utilized to amplify this region. *T. pallidum* genomic DNA was digested to completion with *Kpn*I and ligated under conditions favoring the formation of circularized DNA fragments. Forward and reverse primers based on the 1.0 kb of known sequence between the *Eco*RI and *Kpn*I sites within *cfpA* were constructed and utilized to amplify a 2.1-kb segment of the circularized 3.1-kb *Kpn*I fragment in the genomic DNA preparation. The sequence of the 5′ region of *cfpA* was determined by direct sequence analysis of the resulting PCR product.

The complete nucleotide and deduced amino acid sequences of *cfpA* are shown in Fig. 5. The predicted amino acid sequence corresponded precisely with the amino acid sequences obtained from the N terminus and CNBr fragments determined previously. The 2,034-bp open reading frame is preceded by consensus −35 and −10 promoter sequences (5′-TTTACA-3′ and 5′-TACAAT-3′, respectively) resembling those of *E. coli* and other organisms (9, 10, 12, 28), as well as a putative ribosome-binding site (5′-AGGAG-3′) (35). Two predicted inverted repeat structures suggestive of rho-independent transcription terminator sites are located at the 3′ end (Fig. 5). The deduced amino acid sequence is composed of 678 amino acids and has a calculated molecular mass of 78,539 Da, slightly less than the M_r of 80,000 estimated by SDS-PAGE. Comparison with the Edman degradation results indicated that the N-terminal methionine is removed in the mature protein. The estimated pI of the protein is 6.15. The predicted amino acid composition of *T. pallidum* CfpA is similar to the composition of purified *T. phagedenis* cytoplasmic filaments reported by Masuda and Kawata (25); Asp and Glu are the most abundant amino acids by both analyses.

The Chou-Fasman prediction of secondary structure contains 56% α-helix, 19% β-pleated sheet, and 13% turn structure. Approximately 20% identity and 40 to 50% similarity between the predicted CfpA protein sequence and some filament-associated proteins were noted (see Discussion), but no significant homologies with DNA or protein database entries were identified by BLAST and FastA protocols. Utilization of the COILS program (24) indicated that several noncontiguous regions of CfpA possessed the parallel, two-stranded coiled-coil motif. This structure consists of repeating α-helical heptads having hydrophobic residues at the first and fourth positions (24). Coiled-coil structure is found in extended double-helical regions formed from two intertwining α-helices (such as in the tail region of myosin) or in shorter regions of numerous other proteins. In the latter, it is thought that the regions containing the motif may be involved in intra- or intermolecular associations of α-helices (24).

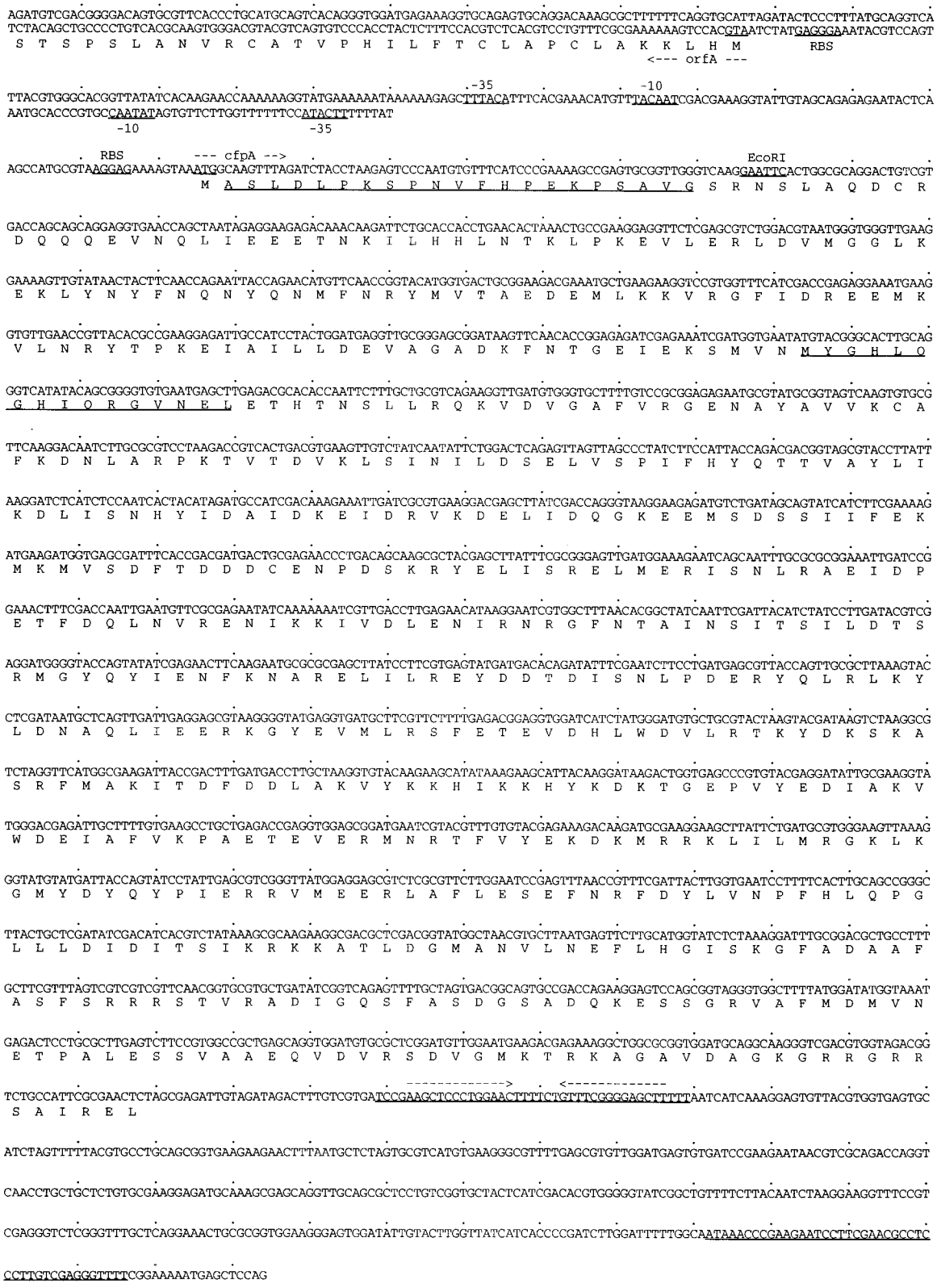
RBS
--- cfpA -->
EcoRI
<--- orfA ---
-35
-10
-10
-35

FIG. 5. Nucleotide sequence and deduced amino acid sequence of the *T. pallidum* subsp. *pallidum* (Nichols) *cfpA* gene. The 2,794-nucleotide sequence includes the promoter region with putative −35 and −10 promoter sequences (underlined), a consensus ribosome-binding site (RBS), and two putative terminators with inverted repeat sequences (underlines and arrows). The underlined regions of the deduced amino acid sequence matched the amino acid sequences determined by Edman degradation. The partial sequence of a putative open reading frame (*orfA*) with associated putative −35 and −10 promoter sequences and RBS are shown at the 5′ end of the sequence.

The 5′ end of the sequenced region also contained the beginning of an open reading frame in the opposite orientation of *cfpA* with putative transcriptional and translational initiation sites (Fig. 5). The 32 amino acids encoded by this open reading frame (*orfA*) lacked detectable sequence homology with existing database entries.

The results of partial sequence analysis of the 4.8-kb region downstream of *cfpA* have indicated the presence of four additional open reading frames, all oriented in the same direction as *cfpA* (data not shown). The predicted protein product of the second of these open reading frames has significant similarity to members of the pyruvate phosphate dikinase family, an intermediary metabolism enzyme found in C_4 plants and protozoa (e.g., accession no. U08399 and P37213). The fourth open reading frame, found at the end of the cloned 6.8-kb DNA insert distal to *cfpA*, encodes a truncated protein product homologous to the first 43 amino acids of the *E. coli* ribosomal protein L27 (accession no. JS0767). Partial sequence analyses of the first and third open reading frames have not demonstrated significant sequence similarity to Genbank and EMBL database entries so far.

cfpA was localized by Southern blot hybridization to the region of the *T. pallidum* genome consisting of the overlap between the *Sfi*I-A, *Not*I-A, *Srf*I-C2, and *Spe*I-A fragments (39; data not shown). *cfpA* and a large motility operon (7, 8, 22; see accession number U36839) are thus located within the same 74-kb region of the chromosome (for further information, see http://utmmg.med.uth.tmc.edu/treponema/tpall.html). The relative locations of *cfpA* and the motility operon within this region are not known, but they are not immediately adjacent to one another on the basis of our sequence data.

Expression of a LacZ-CfpA fusion protein in *E. coli*. One orientation of the 6.8-kb *Eco*RI fragment in pBluescript SK(−) resulted in the in-frame insertion of *cfpA* and the formation of a LacZ-CfpA fusion protein. *E. coli* clones containing this pBluescript construct (pZAPE2-3) grew poorly and were difficult to maintain. Large quantities of the 80-kDa LacZ-CfpA fusion protein were expressed with or without IPTG induction (Fig. 6); additional protein bands with M_rs of 135,000, 96,000, 62,000, and 41,000 were also expressed in this construct. Expression of the CfpA fusion protein in the absence of IPTG was apparently due to leakiness of expression from the *lac* operator in this construct. Electron microscopy of *E. coli* cells expressing the LacZ-CfpA fusion protein (with or without IPTG induction) revealed the presence of large quantities of irregular fibrils that filled the cytoplasm (sparing the nucleoids) and in some cases appeared to penetrate the cell envelope (Fig. 7). Energy-dispersive X-ray spectroscopy revealed a similar spectrum in the samples expressing LacZ-CfpA and controls, indicating that the fibrils were not due to artifactual deposition of heavy metal compounds during staining. Although other proteins expressed by the recombinant construct may be responsible for the formation of these fibrils, our results suggest that the LacZ-CfpA fusion protein may undergo partial self-assembly in *E. coli* or, alternatively, may induce the aggregation of other proteins.

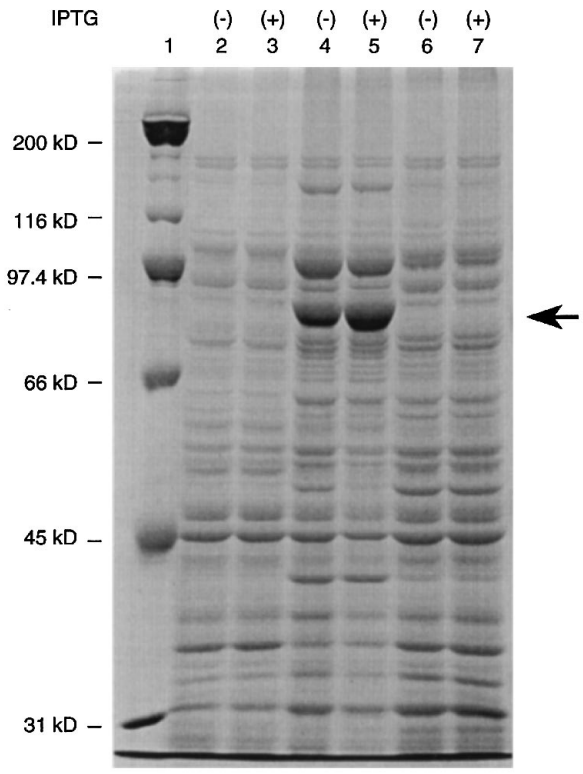

FIG. 6. Expression of a LacZ-CfpA fusion protein in *E. coli*, as demonstrated by SDS-PAGE and Coomassie Blue R staining. Lanes correspond to *E. coli* XL1-Blue MRF′ cells containing pBluescript SK(−) without insert DNA (lanes 2 and 3), pZAPE2-3 plasmid expressing the LacZ-CfpA fusion protein (lanes 4 and 5), and pZAPE2-4 containing *cfpA* in the reverse orientation (lanes 6 and 7). Cells (optical density at 600 nm of 0.4 to 1.0) were incubated in LB medium with (+) or without (−) IPTG at 37°C for 3 h, sonicated, solubilized, and electrophoresed on an SDS–8% polyacrylamide gel. Lane 1 contains protein molecular mass markers. The arrow points to the expected migration position of LacZ-CfpA fusion protein.

DISCUSSION

The cytoplasmic filaments of *T. pallidum* and other spirochetes are highly ordered structures, consisting of ribbon-like elements that extend the length of the cell. Our morphologic observations with *T. pallidum* confirm those obtained by Eipert and Black (5) with *T. refringens*. They determined that the filaments had a highly regular morphology with a rectangular or trapezoidal cross-section. Although negatively stained preparations are more electron-lucent at the edges than in the middle (Fig. 1), the ribbon shape and apparent lack of a lumen in the usual sense indicate that these structures are more properly referred to as filaments or fibrils rather than tubules. The cytoplasmic filaments are difficult to visualize by electron microscopy of ultrathin sections (5, 19, 29, 41), but they appear to run as a bundle, with the wide cross-sectional dimension of each filament in close apposition with that of the neighboring filament (41). This array is closely associated with the inner membrane (perhaps indicating a specific interaction with inner membrane lipids or proteins), and is always located just underneath the cluster of periplasmic flagella. One electron photomicrograph by Hovind-Hougen (1, 14) indicates that the ends of cytoplasmic filaments are bound to the basal plate of a flagellum in a treponemal species (called *T. reiterii* in reference 14 and presumed to be *T. phagedenis* Reiter). We also observed *T. phagedenis* filaments apparently bound to knob-like protrusions on the basal plate (Fig. 1C) but were unable to find examples of intact assemblies of basal bodies, basal plates, and cytoplasmic filaments.

Overall, structural observations to date indicate the following. (i) The assembly and localization of the cytoplasmic filaments is a precise process which must involve highly specific interactions between the protein subunits and with other cellular components. (ii) Their location strongly implies a role of cytoplasmic filaments in the motility of the cell. However,

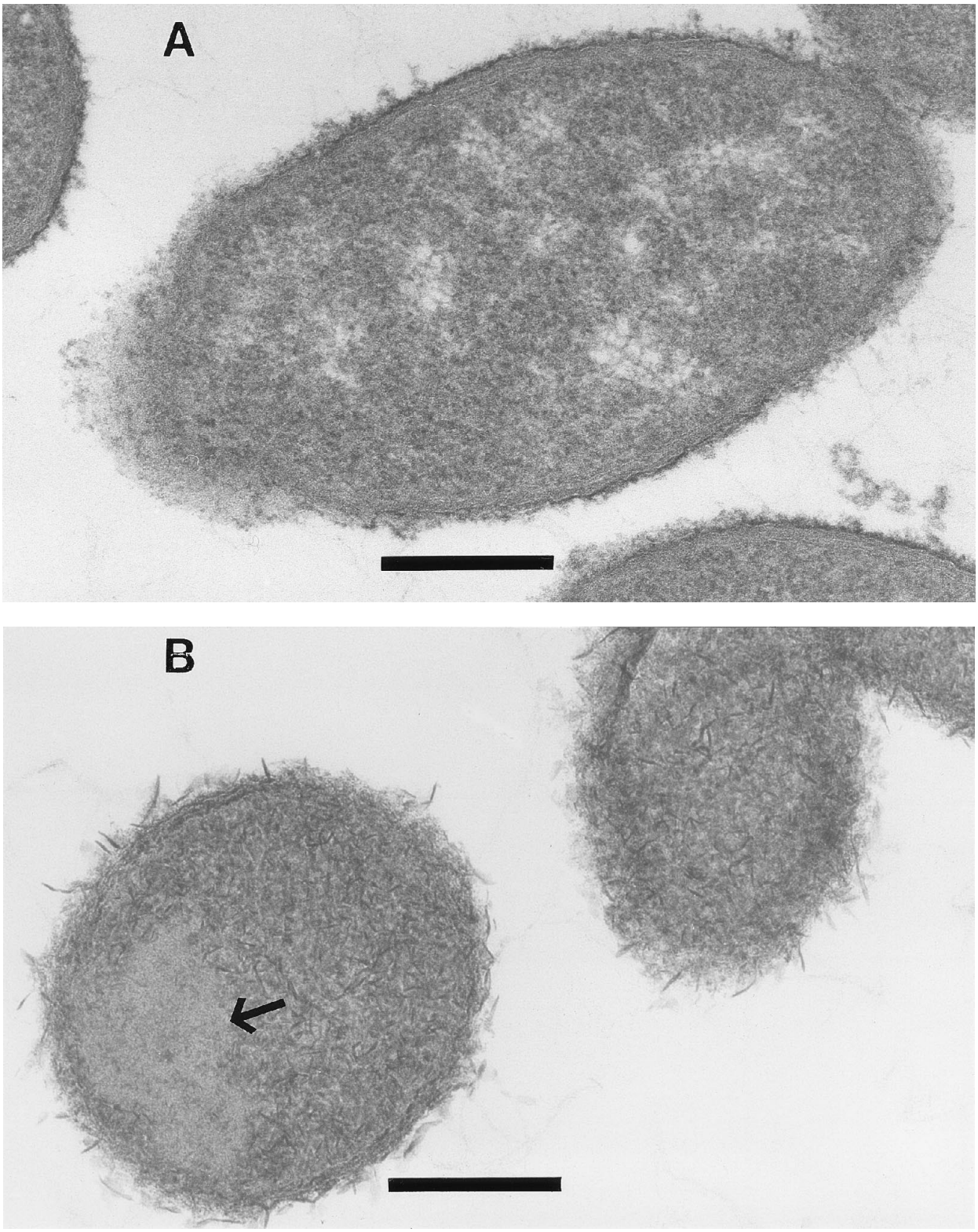

FIG. 7. Electron microscopic appearance of *E. coli* XL1-Blue MRF′ cells expressing the LacZ-CfpA fusion protein. Cells were incubated in the presence of IPTG as described in the legend to Fig. 6 and processed for transmission electron microscopy of ultrathin sections. (A) Control cells containing pBluescript SK(−) plasmid without insert DNA. (B) Cells containing the pZAPE2-3 plasmid expressing the LacZ-CfpA fusion protein. The *E. coli* cells containing pZAPE2-3 were filled with short, irregularly arranged filaments, sparing only the nucleoid regions (arrow). (C) Filaments and debris released from lysed cells, from the same preparation shown in (B). Bars = 200 nm.

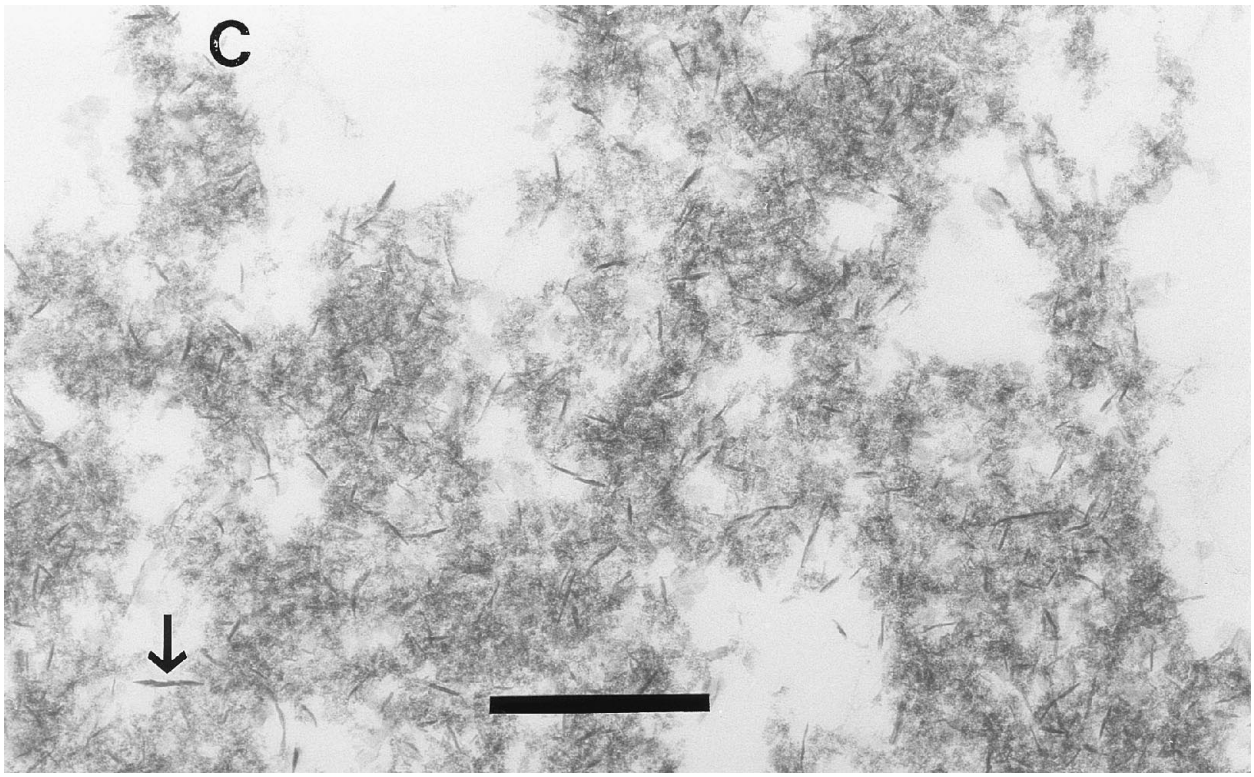

FIG. 7—*Continued.*

other possible functions have been suggested, including maintenance of the helical cell conformation and involvement in DNA replication or segregation (5).

Masuda and Kawata (25) previously purified the cytoplasmic filaments from three *Treponema* strains that were isolated from the oral cavities of patients with pyrrhoea (strains not identified to species level), as well as from *T. phagedenis* and *T. denticola*. The procedure they employed included sonic disruption of the cells, treatment of the lysate with Triton X-100, lysozyme, and *N*-acetylmuramidase to solubilize membranes and peptidoglycan, and (in some separations) CsCl density gradient sedimentation. We were unable to purify the cytoplasmic filaments of *T. pallidum* in significant quantities by these or a variety of other techniques, apparently because of the small amount of starting material available and the lability of the cytoplasmic filament structure. However, the 80-kDa protein that was the major subunit of *T. phagedenis* cytoplasmic filaments was found to have a similar abundance and mobility in single- and two-dimensional gels to a major protein in *T. pallidum*, previously identified as TpN83 (29). Comparison of the N-terminal sequences of these proteins indicated a high degree of sequence identity (85.7%). These similarities in sequence, abundance, and electrophoretic mobility provide evidence that TpN83 represents the major subunit of the cytoplasmic filaments in *T. pallidum*; it is thus referred to as CfpA.

Attempts to clone the intact *cfpA* gene were unsuccessful, most likely because of toxicity of the gene product in *E. coli*. A similar finding was obtained with the *T. pallidum flaA* gene (17, 18), which encodes another abundant gene product, the flagellar sheath protein. Both of these genes have consensus σ^{70} promoters and may be expressed at high levels in *E. coli*, resulting in the accumulation of toxic levels of the proteins and the death of cells containing the intact genes. Because *cfpA* and *flaA* are abundant proteins in *T. pallidum*, their promoters may be particularly adapted to high levels of expression. The mechanism of toxicity in *E. coli* may be different, however; CfpA presumably remains principally in the cytoplasm and may disrupt cellular processes there (e.g., by filament formation), whereas FlaA is secreted into the periplasmic space and may overload the protein export system of the *E. coli* inner membrane.

The complete coding sequence of *cfpA* was determined by combining the sequence obtained from an *Eco*RI fragment lacking the 5′ end of the gene with that of an inverse PCR product containing the putative promoter, ribosome-binding site, and the 5′ end of the coding region. The predicted molecular mass of 78.5 kDa is approximately the same as the M_r of 80,000 determined by SDS-PAGE. Thus, CfpA does not appear to undergo extensive processing, although Edman degradation of the N terminus indicates that the initial methionine residue is removed (Fig. 3 and 5). The Chou-Fasman prediction of secondary structure indicated that an α-helical structure is favored (56%), followed by 19% β-pleated sheet and 13% turn structure. The protein is largely hydrophilic, and contains none of the motifs recognized by the Wisconsin Genetics Computer Group program. Thus, the overall predicted structure of CfpA provides few clues as to its functional and interactive properties.

Although the filamentous proteins contributing to the formation of bacterial flagella and pili have been well characterized, only one other prokaryotic intracytoplasmic filamentous protein gene, *cafA* of *E. coli*, has been cloned and characterized (30, 38). The 51-kDa CafA protein is the major component of cytoplasmic axial filaments. Overproduction of CafA in *E. coli* causes the formation of large bundles of filaments; segregation of daughter cells is prevented, resulting in chained

cells and minicells. Under these conditions, the cytoplasmic axial filaments assemble into a long hexagonal pillar several micrometers long and 0.1 to 0.2 μm in diameter in cross-section. Smaller filamentous structures are presumed to be present during normal cellular growth and may serve cytoskeletal or cytokinetic functions; e.g., they may be involved in chromosome segregation and cell division. CafA displays 34% amino acid similarity with the N-terminal portion of the *E. coli ams* (also called *hmp1*) gene product (2, 3), which in turn exhibits antigenic cross-reactivity (but not significant sequence similarity) with the *Saccharomyces cerevisiae* myosin heavy chain (Myo1) (36). The significance of these similarities is not known.

Optimal alignment of the deduced amino acid sequence of *cfpA* with those of *E. coli cafA*, *E. coli ams* (hmp1), and yeast *myo1* yielded similarities of 42.8, 47.0, and 46.7% and identities of 15.8, 18.4, and 22.3%, respectively. However, this apparent relatedness most likely reflects similarities in amino acid composition, in that comparison of randomized versions of these sequences produced optimized alignments with 42.7 to 45.0% similarity and 18.9 to 20.9% identity. Thus, CfpA does not have a close evolutionary relationship with any known filamentous proteins. It does, however, have limited regions of predicted coiled-coil structure. When present in short segments, this structural motif is thought to represent regions of α-helix that associate with one another at a 20° angle, similar to the relative orientation of the α-helices in myosin, tropomyosin, intermediate filaments, and keratin (24). Such associations may be involved in the intramolecular structure of CfpA subunits or in the association of neighboring subunits in the cytoplasmic filament polymer.

On the basis of our results and previous studies (5, 13–16, 19, 25, 41), the ribbon-like filaments found in treponemal species and other spirochetes are structurally distinct from the tubular structures present in eukaryotes and other prokaryotes and may be a unique structure associated with only a small group of spirochetes. Bermudes et al. (1) recently provided an extensive review of the occurrence of cytoplasmic tubules and filaments in prokaryotes. Proteinaceous tubular structures have been reported in *Azotobacter* species, cyanobacteria, large spirochetes symbiotic in termites, and mollicutes. In addition, the FtsZ protein of *E. coli*, which forms a ring at the point of daughter cell separation during cytokinesis, contains a 7-amino-acid motif that is shared with eukaryotic tubulins and is thought to be involved in GTPase activity and FtsZ polymerization (4). The deduced sequence of CfpA has no significant homology with FtsZ and lacks the tubulin signature sequence or any other recognized nucleotide binding motif. Although we have no firm evidence regarding the relative orientation of CfpA subunits in the cytoplasmic filament structure, a likely arrangement is a linear array of CfpA dimers, each forming one strand of a two-stranded ribbon.

Expression of a LacZ-CfpA fusion protein in *E. coli* resulted in the formation of large quantities of amorphous filaments in *E. coli*, indicating that CfpA is capable of at least partial self-assembly. However, the filaments present in *E. coli* lacked the highly ordered structure of cytoplasmic filaments in *T. pallidum*. This difference may be due to anomalies introduced by the LacZ-CfpA fusion or to a requirement for additional *T. pallidum* components for correct assembly. Little is known about the mechanism of assembly or the possible arrangement of subunits within the cytoplasmic filaments. Filaments from nonpathogenic treponemes are stable in the presence of 2% Triton X-100, 2% sodium deoxycholate, and 2 M urea but dissociate in the presence of 1.3% SDS, 6 M urea, 4 M guanidine hydrochloride, or highly acidic or basic conditions (14, 25). Thus, noncovalent forces are responsible for the integrity of the filament structure. Eipert and Black (5) noted that fine striations were sometimes visible in negatively stained preparations of *T. refringens* cytoplasmic filaments, suggestive of a regular longitudinal arrangement of subunits. They also proposed a complex structure consisting of a rectangular or trapezoidal arrangement of four protein strands.

Thus far, cytoplasmic filaments have been identified only in *Treponema*, *Spirochaeta*, and *Leptonema* species and are not found in *Borrelia*, *Serpulina*, or *Leptospira* species. Thus, they are not required components of spirochete structure. However, it is not known whether cytoplasmic filaments are essential for life of treponemes and other organisms containing them. Although mutational analysis is not practical in *T. pallidum* (because of the inability to culture the bacterium continuously in vitro), insertional mutagenesis of the *flgE* gene has been achieved in *T. denticola* (20). Mutation of *cfpA* in easily cultured spirochetes such as *T. denticola* or *T. phagedenis* may yield information on the possible functions of these unique structural elements.

ACKNOWLEDGMENTS

We thank Robert Harris and Terrance J. Beveridge for providing energy-dispersive X-ray spectroscopy of electron microscopy specimens, and Jerrilyn K. Howell, Monjula Chidambaram, and Mark Pedersen for providing technical assistance in this project.

This work was supported by U.S. Public Health Service grants AI20006 to S.J.N., AI31068 to S.J.N. and G.M.W., and GM46278 to J.K.S.

REFERENCES

1. **Bermudes, D., G. Hinkle, and L. Margulis.** 1994. Do prokaryotes contain microtubules? Microbiol. Rev. **58:**387–400.
2. **Casaregola, S., A. Jacq, D. Laoudj, G. McGurk, S. Margarason, M. Tempete, V. Norris, and I. B. Holland.** 1992. Cloning and analysis of the entire *Escherichia coli* ams gene. ams is identical to hmp1 and encodes a 114 kDa protein that migrates as a 180 kD protein. J. Mol. Biol. **228:**30–40.
3. **Casaregola, S., V. Norris, M. Goldberg, and I. B. Holland.** 1990. Identification of a 180 kD protein in *Escherichia coli* related to a yeast heavy-chain myosin. Mol. Microbiol. **4:**505–511.
4. **de Boer, P., R. Crossley, and L. Rothfield.** 1992. The essential bacterial cell-division protein FtsZ is a GTPase. Nature (London) **359:**254–256.
5. **Eipert, S. R., and S. H. Black.** 1979. Characterization of the cytoplasmic fibrils of *Treponema refringens* (Nichols). Arch. Microbiol. **120:**205–214.
6. **Hanff, P. A., S. J. Norris, M. A. Lovett, and J. N. Miller.** 1984. Purification of *Treponema pallidum*, Nichols strain, by Percoll density gradient centrifugation. Sex. Transm. Dis. **11:**275–286.
7. **Hardham, J. M., J. G. Frye, and L. V. Stamm.** 1995. Identification and sequences of the *Treponema pallidum fliM′*, *fliY*, *fliP*, *fliQ*, *fliR* and *flhB′* genes. Gene **166:**57–64.
8. **Hardham, J. M., J. G. Frye, N. R. Young, and L. V. Stamm.** 1996. Personal communication.
9. **Hawley, D. K., and W. R. McClure.** 1983. Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res. **11:**2237–2255.
10. **Helmann, J. D., and M. J. Chamberlin.** 1988. Structure and function of bacterial sigma factors. Annu. Rev. Biochem. **57:**839–872.
11. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28:**351–359.
12. **Hippel, P. H., D. G. Bear, W. D. Morgan, and J. A. McSwiggen.** 1984. Protein-nucleic acid interactions in transcription: a molecular analysis. Annu. Rev. Biochem. **53:**389–446.
13. **Hovind-Hougen, K.** 1972. Further observations on the ultrastructure of *Treponema pallidum* Nichols. Acta Pathol. Microbiol. Scand. Sect. B **80:**297–304.
14. **Hovind-Hougen, K.** 1976. Determination by means of electron microscopy of morphological criteria of value for classification of some spirochetes, in particular treponemes. Acta. Pathol. Microbiol. Scand. Sect. B Suppl. **255:** 1–41.
15. **Hovind-Hougen, K., and A. Birch-Andersen.** 1971. Electron microscopy of endoflagella and microtubules in *Treponema reiter*. Acta Pathol. Microbiol. Scand. Sect. B **79:**37–50.
16. **Hovind-Hougen, K., A. Birch-Andersen, and H. J. S. Jensen.** 1976. Ultrastructure of cells of *Treponema pertenue* obtained from experimentally infected hamsters. Acta Pathol. Microbiol. Scand. Sect. B **84:**101–108.
17. **Isaacs, R. D., J. H. Hanke, L. M. Guzman-Verduzco, G. Newport, N. Agabian, M. V. Norgard, S. A. Lukehart, and J. D. Radolf.** 1989. Molecular

cloning and DNA sequence analysis of the 37-kilodalton endoflagellar sheath protein gene of *Treponema pallidum*. Infect. Immun. **57:**3403–3411.
18. **Isaacs, R. D., and J. D. Radolf.** 1990. Expression in *Escherichia coli* of the 37-kilodalton endoflagellar sheath protein of *Treponema pallidum* by use of the polymerase chain reaction and a T7 expression system. Infect. Immun. **58:**2025–2034.
19. **Jackson, S., and S. H. Black.** 1971. Ultrastructure of *Treponema pallidum* Nichols following lysis by physical and chemical methods. Arch. Microbiol. **76:**308–324.
20. **Li, H., J. Ruby, N. Charon, and H. Kuramitsu.** 1996. Personal communication.
21. **Lickfield, K. G.** 1976. Transmission electron microscopy of bacteria. Methods Microbiol. **9:**127–176.
22. **Limberger, R. J.** 1996. Personal communication.
23. **Limberger, R. J., and N. W. Charon.** 1986. *Treponema phagedenis* has at least two proteins residing together on its periplasmic flagella. J. Bacteriol. **166:** 105–122.
24. **Lupas, A.** 1996. Prediction and analysis of coiled-coil structures. Methods Enzymol. **266:**513–525.
25. **Masuda, K., and T. Kawata.** 1989. Isolation and characterization of cytoplasmic fibrils from treponemes. Microbiol. Immunol. **33:**619–630.
26. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262:** 10035–10038.
27. **Matsudaira, P.** 1990. Limited N-terminal sequence analysis. Methods Enzymol. **182:**602–613.
28. **McClure, W. R.** 1985. Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. **54:**171–204.
29. **Norris, S. J., and *T. pallidum* Polypeptide Research Group.** 1993. Polypeptides of *Treponema pallidum*: progress toward understanding their structural, functional, and immunologic roles. Microbiol. Rev. **57:**750–779.
30. **Okada, Y., M. Wachi, A. Hirata, K. Suzuki, K. Nagai, and M. Matsuhashi.** 1994. Cytoplasmic axial filaments in *Escherichia coli* cells: possible function in the mechanism of chromosome segregation and cell division. J. Bacteriol. **176:**917–922.
31. **Ovcinnikov, N. M., and V. V. Delektorskij.** 1971. Current concepts of the morphology and biology of *Treponema pallidum* based on electron microscopy. Br. J. Vener. Dis. **47:**315–328.
32. **Penn, C. W.** 1987. Pathogenicity and immunobiology of *Treponema pallidum*. J. Med. Microbiol. **24:**1–9.
33. **Radolf, J. D., M. V. Norgard, and W. W. Shulz.** 1989. Outer membrane ultrastructure explains the limited antigenicity of virulent *Treponema pallidum*. Proc. Natl. Acad. Sci. USA **86:**2051–2055.
34. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
35. **Shine, J., and L. Dalgarno.** 1974. The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA **71:**1342–1346.
36. **Sweeney, F. P., F. Z. Watts, M. J. Podklington, and E. Orr.** 1990. The MYO1 gene from *Saccharomyces cerevisiae*: its complete nucleotide sequence. Nucleic Acids Res. **18:**7147.
37. **Triglia, T., M. G. Peterson, and D. J. Kemp.** 1988. A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. Nucleic Acids Res. **16:**8186.
38. **Wachi, M., M. Doi, T. Ueda, M. Ueki, K. Tsuritani, K. Nagai, and M. Matsuhashi.** 1991. Sequence of the downstream flanking region of the shape-determining genes mreBCD of *Escherichia coli*. Gene **106:**135–136.
39. **Walker, E. M., Y. You, J. K. Howell, A. R. Hoffmaster, G. M. Weinstock, and S. J. Norris.** 1994. Physical map of the genome of *Treponema pallidum* subsp. *pallidum*. J. Bacteriol. **177:**1797–1804.
40. **Walker, E. M., G. A. Zamphigi, 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.
41. **Wiegand, S. E., P. L. Strobel, and L. H. Glassman.** 1972. Electron microscopic anatomy of pathogenic *Treponema pallidum*. J. Invest. Dermatol. **58:**186–204.