INFECTION AND IMMUNITY, Jan. 1989, p. 196–203
0019-9567/89/010196-08$02.00/0

Vol. 57, No. 1

Sequence Analysis of the 47-Kilodalton Major Integral Membrane Immunogen of *Treponema pallidum*

PEI-LING HSU,[1] NEAL R. CHAMBERLAIN,[2] KIM ORTH,[3] CAROLYN R. MOOMAW,[3] LI-QUAN ZHANG,[1] CLIVE A. SLAUGHTER,[3] JUSTIN D. RADOLF,[2,4] STEWART SELL,[1] AND MICHAEL V. NORGARD[2,5]*

Department of Pathology and Laboratory Medicine, The University of Texas Health Science Center at Houston, Houston, Texas 77030,[1] and Departments of Microbiology[2] and Internal Medicine,[4] The Howard Hughes Medical Institute,[3] and The Program in Cell and Molecular Biology,[5] The University of Texas Southwestern Medical Center at Dallas, Dallas, Texas 75235

Received 27 June 1988/Accepted 11 October 1988

The complete primary amino acid sequence for the 47-kilodalton (kDa) major integral membrane immunogen of *Treponema pallidum* subsp. *pallidum* was obtained by using a combined strategy of DNA sequencing (of the cloned gene in *Escherichia coli*) and N-terminal amino acid sequencing of the native (*T. pallidum* subsp. *pallidum*-derived) antigen. An open reading frame believed to encode the 47-kDa antigen comprised 367 amino acid codons, which gave rise to a calculated molecular weight for the corresponding antigen of 40,701. Of the 367 amino acids, 113 (31%) were sequenced by N-terminal amino acid sequencing of trypsin and hydroxylamine cleavage fragments of the native molecule isolated from *T. pallidum* subsp. *pallidum*; amino acid sequence data had a 100% correlation with that of the amino acid sequence predicted from DNA sequencing of the cloned gene in *E. coli*. Although no consensus sequences for the initiation of transcription or translation were readily identifiable immediately 5′ to the putative methionine start codon, a 63-base-pair *PstI* fragment located 159 nucleotides upstream was required for expression of the 47-kDa antigen in *E. coli*. The 47-kDa antigen sequence did not reveal a typical leader sequence. The overall G+C content for the DNA corresponding to the structural gene was 53%. Hydrophilicity analysis identified at least one major hydrophilic domain of the protein near the N terminus of the molecule which potentially represents an immunodominant epitope. No repetitive primary sequence epitopes were found. The combined data provide the molecular basis for further structural and functional studies regarding the role of the antigen in the immunopathogenesis of treponemal disease.

Within the past decade, groups of investigators have begun to characterize on a molecular level the structural components of the bacterium *Treponema pallidum* subsp. *pallidum*, the etiologic agent of venereal syphilis (40). Many of these investigators have sought to identify and to analyze surface-associated or outer membrane proteins of the organism in an attempt to define important immunogens or potential virulence determinants (40, 50). Specifically, Jones et al. (23) reported that monoclonal antibodies directed against a 47-kilodalton (kDa) antigen of *T. pallidum* subsp. *pallidum* possessed treponemicidal activity in the *T. pallidum* immobilization test and in the in vitro-in vivo neutralization test of Bishop and Miller (9, 23). Further work demonstrated that this antigen was abundant in *T. pallidum* subsp. *pallidum* and highly immunogenic in both human and experimental rabbit syphilis (4, 17, 23). Evidence suggests that the 47-kDa antigen may have potential as a serodiagnostic antigen and that monoclonal antibodies directed against the 47-kDa antigen may be used for syphilis diagnosis (4, 20, 28, 30, 37, 39, 48). Infants with congenital syphilis also develop a marked fetal immunoglobulin M response directed specifically against the 47-kDa antigen of *T. pallidum* subsp. *pallidum* (14; P. J. Sanchez, G. H. McCracken, G. D. Wendel, K. Olsen, N. Threlkeld, and M. V. Norgard, J. Infect. Dis., in press). Other pathogenic subspecies of treponemes such as *T. pallidum* subsp. *pertenue*, *T. pallidum* subsp. *endemicum*, and *T. carateum* all apparently possess cognate 47-kDa antigens (5, 6, 23, 27, 29, 30); for example, the predominant serologic response in patients with active pinta was found to be directed against the 47-kDa antigen (15). Immunologic, physicochemical, and genetic data support the pathogen specificity of the 47-kDa antigen (23, 29, 30, 37, 39, 44, 45). Additional work on the 47-kDa antigen of *T. pallidum* subsp. *pallidum* by other investigators recently was reviewed elsewhere (37).

The 47-kDa antigen was originally identified as a surface-associated integral membrane protein on the basis of surface iodination, radioimmunoprecipitation, binding of specific antibody to intact *T. pallidum* subsp. *pallidum*, immunoelectron microscopy, and immunofluorescence experiments (23, 29, 30). However, the putative lability of the outer membrane has hindered drawing firm conclusions about its content(s) on the basis of any one of these techniques alone (30). Penn (41) and Radolf et al. (44) have used detergent fractionation of *T. pallidum* subsp. *pallidum* to identify and to isolate directly the outer membrane of *T. pallidum* subsp. *pallidum*. Penn (41), using Triton X-100, concluded that the 47-kDa antigen was the major protein species present in the putative outer membrane fraction. The 47-kDa antigen also was found to be the predominant moiety present in a detergent extraction procedure containing the minimal amount of Triton X-114 necessary to separate the outer membrane from the protoplasmic cylinder (44). The recombinant DNA-derived form of the antigen also behaves biochemically as an integral membrane protein (12).

We reported previously the cloning and expression of the 47-kDa antigen gene in *E. coli* (37); this has made possible determination of the DNA sequence of the relevant gene. The deduced primary amino acid sequence from the cloned gene was confirmed by direct peptide sequence analysis of

* Corresponding author.

the native *T. pallidum* subsp. *pallidum*-derived antigen. The data will facilitate identification of important antigenic determinants and the elucidation of the structural and functional properties of the molecule.

MATERIALS AND METHODS

Bacterial strains. The virulent Nichols strain of *T. pallidum* subsp. *pallidum* was used as the representative pathogen in this study. It was maintained and cultivated in the testicles of New Zealand White rabbits (without the use of cortisone acetate injections) as previously described (38, 47). Treponemes were isolated in phosphate-buffered saline by differential centrifugation (47) and were enumerated by dark-field microscopy prior to antigen extraction. *E. coli* DH5α (F$^-$ *endA1 hsdR17* [$r_K^- m_K^+$] *supE44 thi-1* λ$^-$ *recA1 gyrA relA1* φ80d*lacZ*ΔM15 Δ[*lacZYA-argF*]U169) (Bethesda Research Laboratories, Inc., Gaithersburg, Md.) was used as the recipient for pUC series plasmid derivatives (54). *E. coli* JM101 ([$r_K^+ m_K^+$] Δ[*lac proAB*] *thi supE*/F′ *traD36 proA*$^+$ *proB*$^+$ *lacI*^{q} *lacZ*ΔM15) (33) was used to harbor M13 derivatives for DNA sequencing analyses.

Plasmids and subcloning into pUC19. Plasmid derivatives were constructed as subclones of pNC81 (12), which originated from plasmid pMN23 (37). Plasmids pPH47.1 (containing *Pst*I fragments A and B) (12) and pPH47.2 (possessing *Pst*I fragments A, B, and D) (12) were generated by inserting the 2.3- and 2.36-kilobase partial *Pst*I fragments of pNC81 (12) into pUC19 vector (54). Plasmid pPH47.4 was made by digesting pPH47.2 with *Kpn*I and recircularization (12). Plasmid pPH47.5 was generated by digesting pPH47.2 with *Hin*dIII and recircularization. Plasmid pPH47.6 was constructed by inserting the 1.35-kb *Eco*RI fragment of pPH47.2 into pUC19. Plasmid pPH47.7 was made by inserting the 1.1-kb *Xho*II-*Eco*RI fragment of pPH47.2 into the *Bam*HI-*Eco*RI sites of pUC19. In this construction, transcription of the 47-kDa antigen mRNA is initiated from the *lac* promoter of pUC19. With the exception of pPH47.7, transcription of 47-kDa antigen mRNA was opposite to the direction of the *lac* promoter in the pUC plasmids. The 47-kDa protein derivatives expressed by pPH47.5, pPH47.6, and pPH47.7 are truncated but contain various numbers of amino acids (i.e., approximately 29, 46, and 8 amino acids, respectively) encoded by the plasmid vector sequence(s). Expression of 47-kDa antigen derivatives by the various plasmids was assessed by immunoblotting expression products with monoclonal antibody 11E3 (23) and rabbit anti-*T. pallidum* subsp. *pallidum* antiserum (12, 23).

Isolation of native 47-kDa antigen from *T. pallidum* subsp. *pallidum* by Triton X-114 phase partitioning. Triton X-114 extraction and phase separation of the 47-kDa *T. pallidum* subsp. *pallidum* protein was performed as described by Bordier (10) as modified by Radolf et al. (44). Detergent-extracted material was processed in either of two ways. (i) The detergent phase (8 ml) was washed five times by repeated dilution to 28 ml with ice-cold phosphate-buffered saline, followed by mixing, rewarming, and centrifugation at 13,000 × *g* for 2 min (20°C). The proteins in the washed detergent phase were then precipitated overnight at −20°C with a 10-fold volume of cold acetone. (ii) Alternatively, for affinity purification of the 47-kDa antigen prior to hydroxylamine cleavage, the Triton X-114 extract was washed three times in 1 ml of 10 mM Tris-HCl (pH 8.0)–5 mM NaCl. The washed detergent phase was diluted from 2 to 1% Triton X-114 in the 10 mM Tris-HCl (pH 8.0)–5 mM NaCl buffer. One milliliter of ReactiGel 6X (Pierce Chemical Co., Rockford, Ill.) containing 2 mg of monoclonal antibody 11E3 per ml of resin was added batchwise to the diluted detergent phase (12). The 47-kDa antigen was then affinity purified (12).

Hydroxylamine cleavage of the native 47-kDa antigen. Purified 47-kDa antigen was dialyzed overnight against 18 liters of distilled H_2O to remove guanidine-HCl. The protein was precipitated overnight with 10 volumes of cold acetone (−20°C). Precipitated protein was collected by centrifugation at 13,000 × *g* for 10 min. The pellet was suspended in 6 M guanidine-HCl–2 M hydroxylamine (HA) (pH 9.0) (11) and was incubated at 45°C for 4 h. The reaction mixture (1 ml) was dialyzed against 1 liter of distilled H_2O overnight (4°C). The protein was lyophilized, and about 100 pmol of HA-cleaved 47-kDa antigen was subjected to sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (22). The HA-cleaved protein was then transferred to polyvinylidene difluoride membrane (Millipore Corp.) (32). The three resulting peptide bands were cut out and subjected to N-terminal amino acid sequencing (below).

Amino acid sequencing of the native 47-kDa antigen. Approximately 100 pmol of the 47-kDa protein was subjected to SDS-PAGE (22) and then transferred to Whatman GF/C glass fiber filter paper derivatized with amino propyl groups by the method of Aebersold et al. (1) as modified by Yuen et al. (55). N-terminal amino acid sequencing was performed on a gas phase sequencer (model 470A; Applied Biosystems) coupled to an on-line high-performance liquid chromatograph (model 120A). Attempts to sequence the N terminus of the intact 47-kDa protein were unsuccessful.

Approximately 500 pmol of the 47-kDa protein was transferred from a 12.5% SDS-PAGE gel to nitrocellulose paper for solid-phase tryptic digestion (2). Peptides were separated by reverse-phase high-performance liquid chromatography on a high-performance liquid chromatograph (model 130A; Applied Biosystems) with a Brownlee RP300 (2.1 × 100 mm) C8 column. Separation was performed in 0.1% trifluoroacetic acid by using a gradient of 0 to 50% acetonitrile over a duration of 120 min at a flow rate of 50 μl/min. Peaks were collected manually onto 1-cm disks of Whatman GF/C paper. Cysteine residues were reduced and alkylated (3). Peptides were then sequenced directly.

DNA sequencing of the 47-kDa antigen gene. Selected DNA fragments were ligated to M13mp18 and used to transfect JM101 cells. Recombinant phages were identified as white plaques on LB plates containing isopropyl-β-D-thiogalactopyranoside (IPTG) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-Gal). The orientations of the inserts were determined by restriction enzyme mapping of the replicative forms of the phage DNA. Single-stranded phage DNAs were purified from the culture supernatants (34). DNA sequences were determined by the dideoxynucleotide chain termination method (49). For most sequencing reactions, the 17-base universal primer (Bethesda Research Laboratories) and the Klenow fragment of DNA polymerase I were used. Two primer oligonucleotides, CATGGTTGACAGCGAGG and CCTCGCTGTCAACCATG, corresponding to nucleotide positions 492 to 508 and 508 to 492 of the 47-kDa antigen gene, respectively, were supplied by Helen Aronovich of the Oligonucleotide Synthesis Laboratory of The University of Texas Southwestern Medical Center.

Computer analyses. The Molecular Biology Information Resource software developed by C. Lawrence of the Baylor College of Medicine, Houston, Tex., was used in conjunction with MicroGenie software (Beckman Instruments, Inc., Palo Alto, Calif.) (43) for DNA sequence analyses.

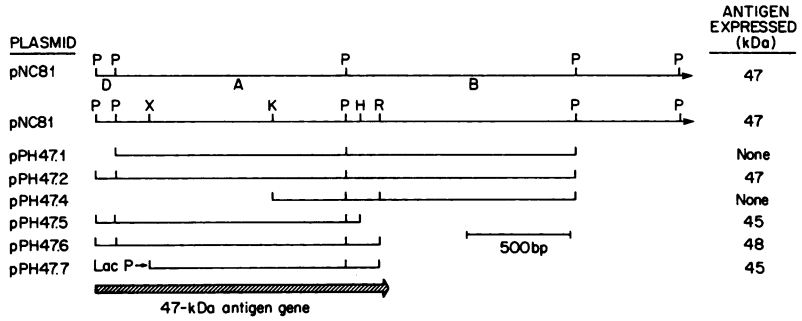

FIG. 1. Partial restriction enzyme maps and relevant expression products of 47-kDa antigen-encoding plasmid derivatives. Plasmid pNC81 (top) has indicated *Pst*I fragments A, B, and D, which collectively constitute a 2.4-kilobase DNA fragment containing the 47-kDa antigen-encoding region. Restriction sites for pNC81 (bottom) are designated as P (*Pst*I), X (*Xho*II), K (*Kpn*I), H (*Hin*dIII), and R (*Eco*RI). The extent of the 47-kDa antigen gene sequence in each subclone is represented by the solid line. The cloning vector for all pPH subclones was pUC19. With the exception of pPH47.7, the direction of transcription of the 47-kDa antigen gene was opposite to that of the *lac* promoter.

RESULTS

Subclones of the 47-kDa antigen gene. The various subclone derivatives of the 47-kDa antigen gene and the relevant expression products of these derivatives are shown in Fig. 1. All subclones originated from pNC81 which contains the entire 47-kDa antigen gene and its regulatory region (12). The first (leftward) *Pst*I site of the D fragment of pNC81 (Fig. 1) is located 5′ to the GC tail used in the original construction of pNC81 (12, 37). Not shown in Fig. 1 is the location of an *Apa*I site (GGGCCC) that presumably was constructed fortuitously as a result of the GC tailing method. Cleavage by *Apa*I cleaves the GC tail attached to the cloning vector. Plasmid constructions lacking the *Pst*I D fragment (e.g., pPH47.1) failed to express any derivative of the 47-kDa antigen. The addition of an active promoter at the *Xho*II site (upstream from the structural gene) could restore expression of some or all of the 47-kDa antigen (e.g., pPH47.7). Thus, as proposed elsewhere (12), the 63-base pair *Pst*I D fragment contains a region that is required for the expression of the 47-kDa antigen gene.

DNA sequencing. The complete nucleotide sequence was obtained by DNA sequencing analysis of subclones shown in Fig. 1. Virtually all of the DNA encoding the structural gene for the 47-kDa antigen was sequenced in both directions. Autoradiographs of sequencing gels were examined by investigators in both Houston and Dallas. By computer analysis, an open reading frame large enough to represent the 47-kDa antigen (Fig. 2) was identified which was compatible with earlier genetic expression data. The sequence contained a TGA stop codon at nucleotide 97, putatively upstream from the 47-kDa structural gene and two additional stop codons (TAG and TAA) at nucleotides 1324 and 1354, respectively. Prior gene expression data established the direction of transcription (12, 37); therefore, the first methionine of the protein positioned downstream from the nucleotide 97 stop codon and in the proper open reading frame was located at nucleotide 223. The nucleotide 97 stop codon did not lie within the 63-base pair *Pst*I fragment required for expression of the 47-kDa antigen gene but rather between the *Pst*I D fragment and the first methionine of the 47-kDa protein. Consensus sequences for −10 Pribnow, −35 (e.g., TTGACA), or −4 to −7 Shine-Dalgarno regions could not be readily identified in the DNA sequence immediately upstream from the first methionine of the protein.

The calculated molecular weight for the protein to the first stop codon is 40,701. The molecular weight calculated on the basis of the second stop codon is about 41,865. The protein contains 10 methionines and one cysteine. There are 54 acidic (Asp, Glu) amino acids, 44 basic (Arg, Lys) amino acids, and the 269 remaining amino acids are neutral; of these, 118 are hydrophobic (Phe, Trp, Tyr, Ile, Leu, Met, Val). The overall G+C content of the DNA was about 53% for the structural gene and about 54% for the entire DNA sequence, consistent with previously published G+C ratios of 52.4 to 53.7% for *T. pallidum* subsp. *pallidum* (Nichols) DNA (35).

Computer analysis revealed at least one perfect inverted repeat and two other areas of the DNA sequence with the potential to form mRNA secondary structures. All are located within the region 5′ to the structural gene. Beginning at nucleotide 87, an 8-base-pair inverted repeat (CGTCT CATCATGAGACG) separated by one C · G base pair exists which extends to nucleotide 103. An additional potential mRNA stem loop structure having homologies between nucleotides 41 to 66 and 141 to 166 (Fig. 3A) also is present. A third potential stem loop-forming region exists between nucleotides 165 to 182 and 211 to 228; this structure contains the methionine start codon at nucleotide 223 (Fig. 3B).

Amino acid sequencing of 47-kDa antigen polypeptide fragments. All attempts to sequence the N terminus of the intact 47-kDa protein were unsuccessful; native 47-kDa antigen preparations isolated either by gel electroelution or by purification with Triton X-114 phase partitioning and monoclonal antibody affinity column chromatography therefore appeared to be blocked to Edman degradation. N-terminal amino acid sequences of unfractionated cyanogen bromide cleavage products, however, were concordant with the predicted amino acid sequence for the 47-kDa protein. N-terminal amino acid sequencing also was carried out on individual trypsin and hydroxylamine cleavage fragments of the 47-kDa antigen; Fig. 2 indicates six trypsin fragments and two hydroxylamine fragments of the 47-kDa antigen that were analyzed by N-terminal amino acid sequencing. The trypsin fragments were located from within 16 amino acids of the N terminus of the molecule to within 20 amino acids of the C terminus of the molecule. Of the 367 amino acids identified from the DNA sequence (Fig. 2), 113 (31%) of the amino acids contained in the native *T. pallidum* subsp. *pallidum* 47-kDa antigen were directly sequenced by N-terminal amino acid sequencing. The sequences of all 113 native amino acids had a 100% correlation with the predicted amino acid sequences derived from DNA sequencing of the cloned gene.

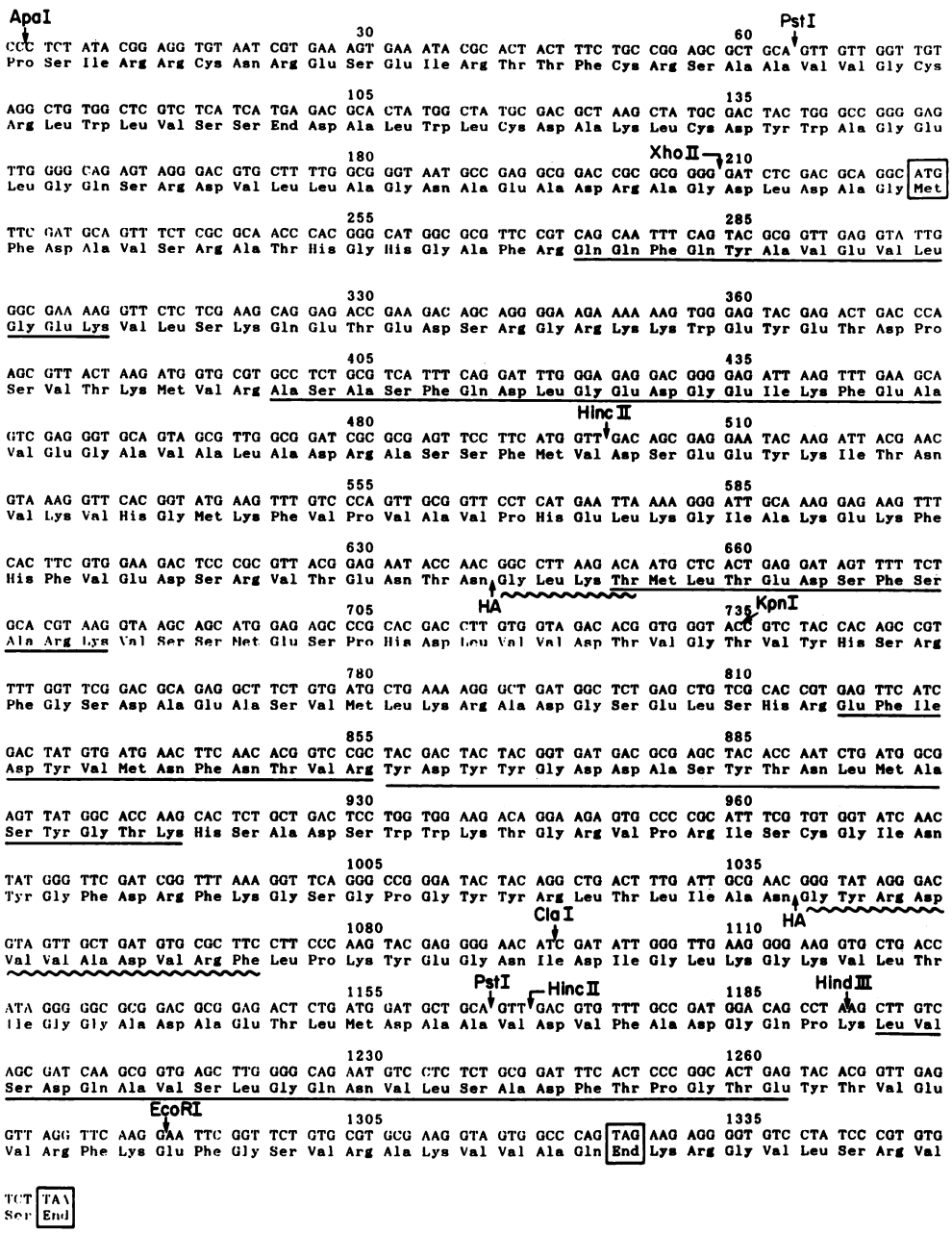

FIG. 2. DNA and corresponding amino acid sequence of the 47-kDa immunogen of *T. pallidum* subsp. *pallidum*. The boxed codons indicate the putative methionine start at the N terminus of the molecule and the first and second stop codons. Arrows above the sequence show DNA cleavage sites for the indicated restriction enzymes. Arrows below the sequence indicate two HA cleavage sites for the protein. ——, Trypsin fragments of the native protein from *T. pallidum* subsp. *pallidum* that were directly N-terminal amino acid sequenced; ∿∿∿, two HA cleavage sites that yielded fragments which also were subjected to N-terminal amino acid sequencing.

Further characterization of the 47-kDa antigen. Hydrophilicity analysis by the algorithm of Hopp and Woods (21) is shown in Fig. 4; a major portion of the protein would appear to be hydrophilic by the given parameters. In particular, one major hydrophilic domain exists within 25 amino acids of the N terminus of the molecule. A hydropathy plot of Kyte and Doolittle (24) also predicts several hydrophobic domains (data not shown).

DNA and protein homology searches were performed by using the National Institutes of Health GenBank (Sept. 1986 version) data bases for bacterial, bacteriophage, and plasmid vector sequences; no significant homologies were identified.

DISCUSSION

A number of laboratories have reported that a 47-kDa polypeptide represents a major, pathogen-specific protein immunogen of *T. pallidum* subsp. *pallidum* (5, 6, 23, 27, 29, 30, 37, 45). This antigen thus is a worthy candidate for detailed structure-function analysis. Future investigations will be facilitated by the nucleotide sequence reported in the present study. This also is the first major treponemal antigen sequenced in which DNA sequencing data has been corroborated by determination of the amino acid sequence for a substantial proportion of the purified native protein.

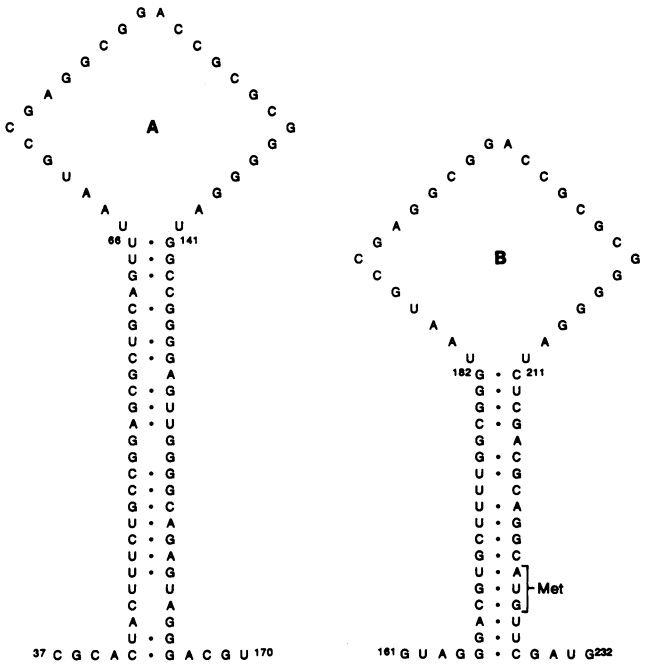

FIG. 3. Potential secondary structures in the mRNA upstream from the 47-kDa antigen gene. Although both structures potentially could form, the structure (B) which traps the AUG methionine (Met) start codon in the stem is energetically favored.

The methionine of position 223 was identified as the presumptive start of translation for the following reasons. First, this methionine initiates an open reading frame large enough to encode a polypeptide of a size compatible with the molecular mass estimated by SDS-PAGE. N-terminal sequence analysis of peptides obtained by both tryptic and hydroxylamine cleavage of the purified, native 47-kDa protein confirmed that this open reading frame did, in fact, encode the 47-kDa antigen. Second, although typical consensus sequences for a −10 Pribnow (TATAAT), a −35 (TTGACA), or a −4 to −7 Shine-Dalgarno (AGGAGG) ribosome-binding site were not readily identified in the DNA sequence immediately 5′ from the methionine codon, previous experiments had shown that the 63-base-pair *Pst*I fragment located 159 nucleotides upstream from position 223 is absolutely essential for expression of the 47-kDa antigen in *E. coli*. Attempts to confirm the assignment of the start codon by N-terminal sequencing of the purified native antigen were unsuccessful, suggesting that the amino terminus of the protein is for some reason blocked to Edman degradation.

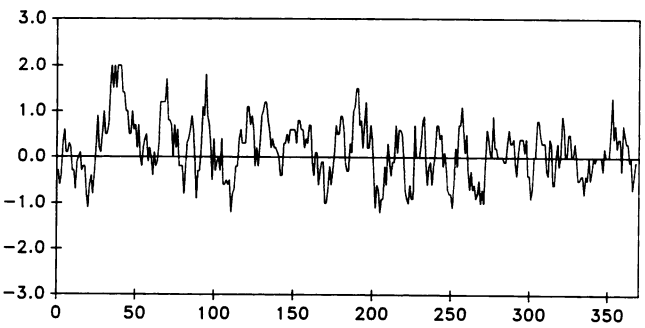

FIG. 4. Hydrophilicity analysis of Hopp and Woods (21). Note the prominent hydrophilic domain near the N terminus of the molecule.

Understanding the initiation of transcription has been puzzling due to the significant distance between the *Pst*I D fragment (required for expression) and the methionine start codon. Hansen et al. (18) reported that expression of the 42-kDa (*tmpA*) and 34-kDa (*tmpB*) antigens of *T. pallidum* subsp. *pallidum* in *E. coli* was poor in the absence of an expression vector but was enhanced significantly for *tmpB* when the p_L promoter of bacteriophage λ was placed some 200 base pairs upstream of the *tmpB* structural gene. The location of one *tmpA* promoter was less than 70 base pairs upstream of the *tmpA* gene, but a second promoter required for transcription was located between 70 and 400 base pairs upstream from the *tmpA* structural gene. Our observation may be analogous to that made regarding the *ompF* and *ompC* genes of *E. coli*, both of which have fairly long untranslated leader regions (80 base pairs for *ompC* and 110 base pairs for *ompF*) that separate the methionine start codon and the −10 Pribnow box (46). Upstream DNA sequences of up to 150 nucleotides or more also have been shown to be essential for full promoter activity in other procaryotic systems (25). Additionally, the 47-kDa antigen gene may be similar to other procaryotic genes with unusual regulatory sequences or which are transcribed by alternative σ factors (19).

Our inability to pinpoint consensus regulatory regions may be consistent with the fact that the expression of this antigen in *E. coli* is relatively poor (12), suggesting weak promoter activity. This may be comparable to the poor expression for treponemal *tmpA* and *tmpB* in *E. coli* (18). Regardless of its activity in *E. coli*, however, the gene probably contains a promoter that is strong for the initiation of mRNA synthesis in *T. pallidum* subsp. *pallidum*.

Computer analysis identified regions of the promoter sequence(s) with the potential to form secondary structures in the mRNA. Formation of mRNA stem loop structures may bring some required regulatory sequence in strategic proximity to another regulatory start signal(s), or the 8-base-pair inverted repeat located at nucleotide 87 may represent some type of control element. In the presence of the *Pst*I D fragment, either of the two stem loop structures shown in Fig. 3 theoretically could form, although the stem loop shown in Fig. 3B may be energetically favored (16). In the absence of the *Pst*I D fragment, only the stem loop shown in Fig. 3B would exist, thereby totally trapping the initiator ATG in the stem which may block the ATG start signal.

A particularly important feature of the 47-kDa structural gene concerns the absence of a typical leader or signal peptide at its amino terminus (7, 42). In fact, a computer-generated hydropathy plot indicated that this portion of the molecule is relatively hydrophilic. This result is consistent with the observation that a product larger than the native 47-kDa antigen could not be identified in a cell-free coupled transcription-translation assay system (unpublished data). Although this molecule was initially thought to be located exclusively in the outer membrane (12, 41), recent data indicate that it may be located in both the cytoplasmic and outer membranes of *T. pallidum* subsp. *pallidum* (44). Moreover, when synthesized at high levels in *E. coli* by using an expression vector system, the cloned 47-kDa antigen partitions into both the inner and outer membranes of that organism as well (12). These data suggest that the mature translation product possesses the necessary structural information for targeting to the cytoplasmic and outer membranes

of both *T. pallidum* subsp. *pallidum* and *E. coli*. *E. coli* outer membrane proteins lacking signal sequences have been described, such as leader peptidase I (53), which can be found in both the cytoplasmic and outer membrane fractions (56). Although virtually nothing is known about the parameters that influence protein export in *T. pallidum* subsp. *pallidum*, Stamm et al. (51), Hansen et al. (18), and Dallas et al. (13) have demonstrated that posttranslational processing of some *T. pallidum* subsp. *pallidum* proteins occurs. A conjectured signal peptide also was noted for the *tmpA* protein of *T. pallidum* subsp. *pallidum* (18).

The 3′ encoding region of the gene contains two stop codons separated by nine amino acid codons. If the first stop codon is used, the discrepancy between the calculated molecular weight (40,701) and that estimated by SDS-PAGE (47,000) (23) is not inordinate. Therefore, we have elected to retain the "47-kDa" nomenclature in order not to cause confusion in the literature (40). The existence of the second TAA stop codon may explain a peculiar phenomenon previously reported; namely, the 47-kDa antigen typically migrates as a 47- to 48-kDa doublet on SDS-PAGE gels (12). Termination may fail occasionally at the first TGA stop codon, thereby allowing the protein to be elongated an additional 10 amino acids. In our hands, the higher-molecular-weight species of the 47- to 48-kDa doublet represents by far the minor component.

An ultimate goal is to relate structure of the 47-kDa antigen with functional and immunologic characteristics. The data obtained in this study provide the molecular basis for elucidation of the respective roles of humoral and cell-mediated immune responses to the immunogenicity of the protein during infection by *T. pallidum* subsp. *pallidum*. Hydrophilicity analysis revealed at least one major hydrophilic domain near the N terminus of the molecule, which represents a primary candidate as an immunodominant epitope (21). In support of this, preliminary epitope mapping experiments showed that a vast majority of mouse monoclonal antibodies raised against the 47-kDa antigen react with the N-terminal hydroxylamine cleavage fragment containing this hydrophilic domain; the same was true when human syphilitic or rabbit anti-*T. pallidum* subsp. *pallidum* sera was examined for antibody reactivity with the hydroxylamine cleavage fragments of the 47-kDa antigen (Chamberlain et al., unpublished data). The protein additionally may contain a domain(s) that serves as a polyclonal activator(s) of B lymphocytes (52). This may partially explain the intense and specific fetal immunoglobulin M response to the 47-kDa antigen in congenital syphilis (14; Sanchez et al., in press). In addition, certain domains may serve as functional T-cell recognition epitopes (8, 31) that promote the activity of cell-mediated immunity and clearance of *T. pallidum* subsp. *pallidum* from primary lesions (50).

There is no doubt that the 47-kDa antigen is an integral membrane protein (12, 44), but the actual basis for the hydrophobic character of the molecule is not readily apparent from its primary sequence. Its characteristic partitioning into the detergent phase upon Triton X-114 extraction (12, 44, 45) substantiates its overall hydrophobic nature. The existence of multiple hydrophilic domains is compatible with the notion that the 47-kDa antigen can reside in an outer membrane, such as in the case of bacterial porins (36).

The availability of the entire sequence for the 47-kDa antigen provides additional practical tools. The entire DNA sequence or selected constituent oligonucleotide portions, including synthetic oligonucleotides, may be used as molecular gene probes for the detection of the organism in various tissues and body fluids or both. Knowledge of the amino acid sequence also allows the testing of strategic synthetic peptides to identify and confirm immunodominant B-cell or T-cell epitopes. Synthetic peptides also may be used as the basis for improved treponemal serologic tests and treponemal synthetic peptide vaccines or both (26).

ACKNOWLEDGMENTS

We thank Mark Swancutt for reviewing the manuscript, Martin Goldberg for excellent technical and computer assistance, and Cindy Baselski for outstanding typing.

This study was partially supported by grant V3-181-64 from the World Health Organization, grant I-940 from the Robert A. Welch Foundation, and Public Health Service grants AI-16692 and AI-17366 from the National Institutes of Allergy and Infectious Diseases to M.V.N. and Public Health Service grant PA01-AI-21290 from the National Institutes of Health to S.S. N.R.C. was a Robert A. Welch Foundation postdoctoral fellow. J.D.R. was a recipient of a Pfizer Pharmaceuticals New Faculty Scholars Award.

LITERATURE CITED

1. **Aebersold, R. H., B. T. David, L. E. Hood, and S. B. H. Kent.** 1986. Electroblotting onto activated glass. J. Biol. Chem. **261:** 4229-4238.
2. **Aebersold, R. H., J. Leavitt, R. A. Saavedra, L. E. Hood, and S. B. H. Kent.** 1987. Internal amino acid sequence analysis of proteins separated by one- or two-dimensional gel electrophoresis by *in situ* protease digestion on nitrocellulose. Proc. Natl. Acad. Sci. USA **84:**6970–6974.
3. **Andrews, P. C., and J. E. Dixon.** 1987. A procedure for *in situ* alkylation of cysteine residues on glass fiber prior to protein microsequence analysis. Anal. Biochem. **161:**524–528.
4. **Baker-Zander, S. A., E. W. Hook, P. Bonin, H. H. Handsfield, and S. A. Lukehart.** 1985. Antigens of *Treponema pallidum* recognized by IgG and IgM antibodies during syphilis in humans. J. Infect. Dis. **151:**264–272.
5. **Baker-Zander, S. A., and S. A. Lukehart.** 1983. Molecular basis of immunological cross-reactivity between *Treponema pallidum* and *Treponema pertenue*. Infect. Immun. **42:**634–638.
6. **Baker-Zander, S. A., and S. A. Lukehart.** 1984. Antigenic cross-reactivity between *Treponema pallidum* and other pathogenic members of the family *Spirochaetaceae*. Infect. Immun. **46:**116–121.
7. **Bankaitis, V. A., E. Altman, and S. D. Emr.** 1987. Export and localization of *Escherichia coli* envelope proteins, p. 75–116. *In* M. Inouye (ed.), Bacterial outer membranes as model systems. John Wiley & Sons, Inc., New York.
8. **Berzofsky, J. A., K. B. Cease, J. L. Cornette, J. L. Spouge, H. Margalit, I. J. Berkower, M. F. Good, L. H. Miller, and C. DeLisi.** 1987. Protein antigenic structures recognized by T cells: potential applications to vaccine design. Immunol. Rev. **98:**9–52.
9. **Bishop, N. H., and J. N. Miller.** 1976. Humoral immunity in experimental syphilis. II. The relationship of neutralizing factors in immune serum to acquired resistance. J. Immunol. **117:** 197–207.
10. **Bordier, C.** 1981. Phase separation of integral membrane proteins in Triton X-114 solution. J. Biol. Chem. **256:**1604–1607.
11. **Bornstein, P., and G. Balain.** 1977. Cleavage at Asn-Gly bonds with hydroxylamine. Methods Enzymol. **47:**132–145.
12. **Chamberlain, N. R., J. D. Radolf, P.-L. Hsu, S. Sell, and M. V. Norgard.** 1987. Genetic and physicochemical characterization of the recombinant DNA-derived 47-kilodalton surface immunogen of *Treponema pallidum* subsp. *pallidum*. Infect. Immun. **56:** 71–78.
13. **Dallas, W. S., P. H. Ray, J. Leong, C. D. Benedict, L. V. Stamm, and P. J. Bassford, Jr.** 1987. Identification and purification of a recombinant *Treponema pallidum* basic membrane protein antigen expressed in *Escherichia coli*. Infect. Immun. **55:**1106–1115.
14. **Dobson, S. R. M., L. H. Taber, and R. E. Baughn.** 1988.

Recognition of *Treponema pallidum* antigens by IgM and IgG antibodies in congenitally infected newborns and their mothers. J. Infect. Dis. **157**:903–910.
15. **Fohn, M. J., F. S. Wignall, S. A. Baker-Zander, and S. A. Lukehart.** 1988. Specificity of antibodies from patients with pinta for antigens of *Treponema pallidum* subspecies *pallidum*. J. Infect. Dis. **157**:32–37.
16. **Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. G. Turner.** 1986. Improved free-energy parameters for prediction of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83**:9373–9377.
17. **Hanff, P. A., T. E. Fehniger, J. N. Miller, and M. A. Lovett.** 1982. Humoral immune response in human syphilis to polypeptides of *Treponema pallidum*. J. Immunol. **129**:1287–1291.
18. **Hansen, E. B., P. E. Pedersen, L. M. Schouls, E. Severin, and J. D. A. van Embden.** 1985. Genetic characterization and partial sequence determination of a *Treponema pallidum* operon expressing two immunogenic membrane proteins in *Escherichia coli*. J. Bacteriol. **162**:1227–1237.
19. **Helmann, J. D., and M. J. Chamberlin.** 1987. DNA sequence analysis suggests that expression of flagellar and chemotaxis genes in *Escherichia coli* and *Salmonella typhimurium* is controlled by an alternative σ factor. Proc. Natl. Acad. Sci. USA **84**:6422–6424.
20. **Hook, E. W., R. E. Roddy, S. A. Lukehart, J. Hom, K. K. Holmes, and M. R. Tam.** 1985. Detection of *Treponema pallidum* in lesion exudate with a pathogen-specific monoclonal antibody. J. Clin. Microbiol. **22**:241–244.
21. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
22. **Hunkapiller, M. W., E. Lujan, F. Ostrander, and L. E. Hood.** 1983. Isolation of microgram quantities of protein from polyacrylamide gels for amino acid sequence analysis. Methods Enzymol. **91**:227–236.
23. **Jones, S. A., K. S. Marchitto, J. N. Miller, and M. V. Norgard.** 1984. Monoclonal antibody with hemagglutination, immobilization, and neutralization activities defines an immunodominant, 47,000 mol wt., surface-exposed immunogen of *Treponema pallidum* (Nichols). J. Exp. Med. **160**:1404–1420.
24. **Kyte, J., and R. F. Doolittle.** 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. **157**:105–132.
25. **Lamond, A. I., and A. A. Travers.** 1983. Requirement for an upstream element for optimal transcription of a bacterial tRNA gene. Nature (London) **305**:249–250.
26. **Lerner, R. A., N. Green, A. Olson, T. Shinnick, and J. G. Sutcliffe.** 1981. The development of synthetic vaccines. Hosp. Pract. **16**:55–62.
27. **Lukehart, S. A., S. A. Baker-Zander, and E. R. Gubish, Jr.** 1982. Identification of *Treponema pallidum* antigens: comparison with a nonpathogenic treponeme. J. Immunol. **129**:833–838.
28. **Lukehart, S. A., M. R. Tam, J. Hom, S. A. Baker-Zander, K. K. Holmes, and R. C. Nowinski.** 1985. Characterization of monoclonal antibodies to *Treponema pallidum*. J. Immunol. **134**:585–592.
29. **Marchitto, K. S., S. A. Jones, R. F. Schell, P. L. Holmans, and M. V. Norgard.** 1984. Monoclonal antibody analysis of specific antigenic similarities among pathogenic *Treponema pallidum* subspecies. Infect. Immun. **45**:660–666.
30. **Marchitto, K. S., C. K. Selland-Grossling, and M. V. Norgard.** 1986. Molecular specificities of monoclonal antibodies directed against virulent *Treponema pallidum*. Infect. Immun. **51**:168–176.
31. **Margalit, H., J. L. Spouge, J. L. Cornette, K. B. Cease, C. DeLisi, and J. A. Berzofsky.** 1987. Prediction of immunodominant helper T cell antigenic sites from the primary sequence. J. Immunol. **138**:2213–2229.
32. **Matsudaira, P.** 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. J. Biol. Chem. **262**:10035–10038.
33. **Messing, J.** 1979. A multipurpose cloning system based on the single-stranded DNA bacteriophage M13. Recomb. DNA Tech. Bull. **2**:43–48.
34. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
35. **Miao, R., and A. H. Fieldsteel.** 1978. Genetics of *Treponema*: relationship between *Treponema pallidum* and five cultivable treponemes. J. Bacteriol. **133**:101–107.
36. **Mizushima, S.** 1987. Assembly of membrane proteins, p. 163–185. *In* M. Inouye (ed.), Bacterial outer membranes as model systems. John Wiley & Sons, Inc., New York.
37. **Norgard, M. V., N. R. Chamberlain, M. A. Swancutt, and M. S. Goldberg.** 1986. Cloning and expression of the major 47-kilodalton surface immunogen of *Treponema pallidum* in *Escherichia coli*. Infect. Immun. **54**:500–506.
38. **Norgard, M. V., and J. N. Miller.** 1983. Cloning and expression of *Treponema pallidum* (Nichols) antigen genes in *Escherichia coli*. Infect. Immun. **42**:435–445.
39. **Norgard, M. V., C. K. Selland, J. R. Kettman, and J. N. Miller.** 1984. Sensitivity and specificity of monoclonal antibodies directed against antigenic determinants of *Treponema pallidum* (Nichols) in the diagnosis of syphilis. J. Clin. Microbiol. **20**:711–717.
40. **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. Moskophidis, F. Müller, M. V. Norgard, C. W. Penn, L. V. Stamm, J. D. van Embden, and K. Wicher.** 1987. Identity of *Treponema pallidum* ssp. *pallidum* polypeptides: correlation of sodium dodecyl sulfate-polyacrylamide gel electrophoresis results from different laboratories. Electrophoresis **8**:77–92.
41. **Penn, C. W., A. Cockayne, and M. J. Bailey.** 1985. The outer membrane of *Treponema pallidum*: biological significance and biochemical properties. J. Gen. Microbiol. **131**:2349–2357.
42. **Politt, S., and M. Inouye.** 1987. Structure and functions of the signal peptide, p. 117–139. *In* M. Inouye (ed.), Bacterial outer membranes as model systems. John Wiley & Sons, Inc., New York.
43. **Queen, C., and L. J. Korn.** 1984. A comprehensive sequence analysis program for the IBM personal computer. Nucleic Acids Res. **12**:581–599.
44. **Radolf, J. D., N. R. Chamberlain, A. Clausell, and M. V. Norgard.** 1988. Identification and localization of integral membrane proteins of virulent *Treponema pallidum* subsp. *pallidum* by phase partitioning with the nonionic detergent Triton X-114. Infect. Immun. **56**:490–498.
45. **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.
46. **Ramakrishnan, G., D. E. Comeau, K. Ikenaka, and M. Inouye.** 1987. Transcriptional control of gene expression: osmoregulation of porin protein synthesis, p. 3–16. *In* M. Inouye (ed.), Bacterial outer membranes as model systems. John Wiley & Sons, Inc., New York.
47. **Robertson, S. M., J. R. Kettman, J. N. Miller, and M. V. Norgard.** 1982. Murine monoclonal antibodies specific for virulent *Treponema pallidum* (Nichols). Infect. Immun. **36**:1076–1085.
48. **Romanowski, B., E. Forsey, E. Prasad, S. Lukehart, M. Tam, and E. W. Hook, III.** 1987. Detection of *Treponema pallidum* by a fluorescent monoclonal antibody test. Sex. Transm. Dis. **14**:156–159.
49. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
50. **Sell, S., and S. J. Norris.** 1983. The biology, pathology, and immunology of syphilis. Int. Rev. Exp. Pathol. **24**:203–276.
51. **Stamm, L. V., T. C. Kerner, Jr., V. A. Bankaitis, and P. J. Bassford, Jr.** 1983. Identification and preliminary characterization of *Treponema pallidum* protein antigens expressed in *Escherichia coli*. Infect. Immun. **41**:709–721.
52. **Vordermeier, H. M., and W. G. Bessler.** 1987. Polyclonal activation of murine B lymphocytes *in vitro* by *Salmonella*

typhimurium porins. Immunobiology **175**:245–251.
53. **Wolfe, P. B., W. Wickner, and J. M. Goodman.** 1983. Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. J. Biol. Chem. **258**:12073–12080.
54. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–119.
55. **Yuen, S., M. W. Hunkapiller, K. J. Wilson, and P. M. Yuan.** 1986. SDS-PAGE electroblotting. Applied Biosystems User Bulletin. **25**:1–15.
56. **Zwizinski, C., T. Date, and W. Wickner.** 1981. Leader peptidase is found in both the inner and outer membranes of *Escherichia coli*. J. Biol. Chem. **256**:3593–3597.