JOURNAL OF BACTERIOLOGY, June 1985, p. 1227–1237
0021-9193/85/061227-11$02.00/0

Vol. 162, No. 3

Genetic Characterization and Partial Sequence Determination of a *Treponema pallidum* Operon Expressing Two Immunogenic Membrane Proteins in *Escherichia coli*

E. B. HANSEN,[1] P. E. PEDERSEN,[1] L. M. SCHOULS,[2] EVA SEVERIN,[1] AND J. D. A. VAN EMBDEN[2]*

Technical University of Denmark, Gensplejsningsgruppen, DK 2800 Lyngby, Denmark, [1] and Rijksinstituut voor Volksgezondheid en Milieuhygiëne, Laboratory for Bacteriology, 3720 BA Bilthoven, The Netherlands[2]

Received 27 November 1984/Accepted 27 February 1985

A detailed physical and genetic map of a previously cloned 5.5-kilobase segment of *Treponema pallidum* DNA is described. This segment expressed two proteins that are cell membrane associated in *Escherichia coli*. The structural genes of these treponemal membrane proteins, *tmpA* and *tmpB*, are coordinately expressed, and transcription in *E. coli* can start from at least two different treponemal promoters. The *tmpA* and *tmpB* proteins are the products of in vivo proteolytic cleavage from precursor proteins which are 2 and 4 kilodaltons larger, respectively, than the mature proteins. Because the sizes of the corresponding proteins produced in *T. pallidum* were identical to those of the mature membrane proteins in *E. coli*, we concluded that a similar proteolytic processing takes place in both *E. coli* and *T. pallidum*. Although *tmpA* and *tmpB* were controlled by the same transcription signals, *tmpB* was expressed to a higher extent than *tmpA*, and only the *tmpB* product could be overproduced by placing the left lambda promoter in front of the structural genes. The nucleotide sequence of the *T. pallidum tmpA* gene was established. This is the first *T. pallidum* gene sequenced. Codon usage and the nature of transcriptional and translational signals are discussed. The deduced amino acid sequence indicated the presence of a sequence that was characteristic for a signal peptide. This sequence information allowed the construction of hybrid genes coding for proteins having β-galactosidase enzyme activity as well as TmpA epitopes. The enzyme-linked antigen was expressed at a high level in *E. coli* when transcriptional and translational signals from coliphage λ were used. In this case the protein produced was a sandwich protein consisting of 21 amino acids of the λ *cro* protein, 204 amino acids of the *T. pallidum* TmpA protein, and 1,020 amino acids of the *E. coli* λ-galactosidase. The potential use of this enzyme-linked antigen for the serodiagnosis of syphilis is discussed.

Treponema pallidum is the causative agent of syphilis. Because the microorganism cannot be grown in vitro, little is known about the factors that contribute to its virulence and the immunologic response during a syphilitic infection. For this reason various groups have cloned genes of *T. pallidum* that encode for immunogenic proteins (17, 24, 27, 30). These recent studies have shown that *T. pallidum* DNA can be expressed well in *Escherichia coli*, and treponemal transcriptional and translational signals appear to be recognized by the *E. coli* machinery for protein synthesis.

Recently, we screened a cosmid library with *T. pallidum* DNA inserts for the production of *T. pallidum* antigens (27). Among the 800 clones screened, one particular phenotype was found most frequently. Of 16 antigen-producing clones, 6 were classified as phenotype B, and these clones produced two immunologically unrelated protein antigens with molecular sizes of 35 and 44 kilodaltons (kd). The latter antigen is of particular interest, because it is one of the major immunogens of *T. pallidum* in syphilitic patients (26; R. V. W. Van Eijk, unpublished data). Furthermore, the 44-kd antigen did not react with antibodies directed against *Treponema phagedenis*, Reiter, suggesting that the 44-kd antigen is serologically specific for *T. pallidum* and, therefore, that it might be a good candidate for use as a specific antigen in the serodiagnosis of syphilis. Both antigens were shown to be incorporated into the cell envelope of *E. coli*, which made them potential candidates for use in vaccines. To construct plasmids that permit inducible expression of the phenotype B antigens in *E. coli* K-12, we initiated this study of the genetic organization of the treponemal DNA involved in expression of the 44- and the 35-kd antigens.

Although the treponemal transcriptional, translational, and posttranslational signals appear to be recognized well by the *E. coli* machinery for protein synthesis, no data have been published on the nature of these expression signals. Therefore, in this study we sequenced one of the genes involved in expression of the phenotype B antigens. Furthermore, the amino acid sequence derived from this DNA sequence could serve for the development of synthetic peptides that might be useful in the serodiagnosis of syphilis or for a vaccine against this infectious disease.

Fusion proteins composed of β-galactosidase and protein antigens might be useful in the serodiagnosis of infectious diseases. Such fusion proteins might be extremely specific probes, because the enzymatic activity is associated only with the antigen of interest, not with other cellular proteins. In this paper we describe, in addition, the construction of a plasmid that allowed the inducible hyperproduction of a hybrid protein composed of β-galactosidase and part of one of the treponemal antigen genes.

MATERIALS AND METHODS

Bacterial strains and plasmids. The *E. coli* K-12 derivatives used throughout this study were SC181 (*leu thi thr hsdR hsdM supE*), obtained from E. Lederberg; P678-54 (*thr leu lacY minA gal minB thi str*), obtained from S. N. Cohen; K-12.H1 (*trp* Sm *lacZ*(Am) λ *bio-uvrB trpEa2* λ *N*am7 *N*am53 *c*I857.H1), obtained from E. Remaut (19); and

* Corresponding author.

TABLE 1. Plasmids used and derived in this study

Plasmid	Derivation or relevant properties	Origin or reference
pPLc236	Carries the p_L promoter of lambda	12
pHC79	Cosmid cloning vector	4
pRIT4616	*Hpa*I deletion of pRIT4694	This study
pRIT4618	*Pvu*II deletion of pRIT4694	This study
pRIT4619	*Kpn*I deletion of pRIT4694	This study
pRIT4620	*Bgl*II deletion of pRIT4694	This study
pRIT4622	*Xho*I deletion of pRIT4694	This study
pRIT4623	*Xho*I deletion of pRIT4694	This study
pRIT4624	*Xho*I deletion of pRIT4694	This study
pRIT4626	*Xho*I deletion of pRIT4694	This study
pRIT4627	*Sst*I deletion of pRIT4694	This study
pRIT4628	Insertion of *Eco*RI-*Sal*I fragment of pRIT4694 into large *Eco*RI-*Sal*I fragment of pPLc236	This study
pRIT4632	*Xho*I deletion of pRIT4694	This study
pRIT4633	Insertion of *Hin*dIII-*Sal*I fragment of pRIT4694 into large *Hin*dIII-*Sal*I fragment of pPLc236	This study
pRIT4637	*Hin*dIII deletion of pRIT4694	This study
pRIT4638	*Sph*I deletion of pRIT4694	This study
pRIT4639	*Sma*I-*Pvu*II deletion of pRIT4694	This study
pRIT4640	*Hpa*I-*Pvu*II deletion of pRIT4694	This study
pRIT4641	*Sma*I-*Hpa*I deletion of pRIT4694	This study
pRIT4642	*Hin*dII deletion of pRIT4694	This study
pRIT4645	*Xho*I deletion of pRIT4695	This study
pRIT4646	*Cla*I-*Hpa*I deletion of pRIT4628	This study
pRIT4650	*Cla*I-*Sma*I deletion of pRIT4628	This study
pRIT4694	*Cla*I-deleted derivative of pRIT4600 expressing *tmpA* and *tmpB*	15
pRIT4695	Spontanous mutant of pRIT4694	15
pEH253	*lacZ*-containing pBR322 derivative	E. B. Hansen
pTG402	Insertion of *Pst*I fragment of pRIT4694 into pEH253	This study
pTG428	Inducible *tmpA-lacZ* fusion plasmid	This study

MC1000 [*thi* Δ (*ara-leu-7679*) *araD139 lac*Δ*74 glaU galK rpsL*] (2). *T. pallidum* (Nichols) was maintained by serial passage in rabbits, and this strain served as a source of *T. pallidum* protein.

Plasmids and derivatives of pRIT4694 (27) are listed in Table 1. Unless otherwise stated, SC181 was used as the host for pRIT4694 and its deletion derivatives. MC1000 was used as the host for pEH253 and the pTG plasmids. pEH253 is a pBR322-derived plasmid bearing the complete *lacZ* gene of *E. coli* (E. B. Hansen, Ph.D. thesis, Technical University of Denmark, Copenhagen, 1984). A polylinker was introduced in the N terminus of the *lacZ* gene of pEH253 giving rise to a frameshift mutation. Therefore, the phenotype of pEH253 was LacZ$^-$. The sequence of the N terminus of *lacZ* with the polylinker was as follows:

Pst wild-type sequence (10)
ATGACCATGATTACGAATTGCTGCAGGTCGACGGATCCGGGGCACTGGCC-
amino acids of wild-type *lacZ* protein: Leu ala
(8) (9)

pTG402 was constructed from pRIT4694 and pEH253 as follows. Partial cleavage of pRIT4694 with *Pst*I allowed a fragment, extending from the *Pst*I site in the *bla* gene to the second *Pst*I site in the *tmpA* gene, to be purified. This fragment was ligated in molar excess to completely *Pst*I-cleaved pEH253 and transformed into strain MC1000. The majority of the ampicillin-resistant transformants showed a weak LacZ$^+$ phenotype on medium containing X-Gal (Sigma Chemical Co., St. Louis, Mo.). The plasmids of 12 of those colonies were analyzed, and all were found to contain the desired plasmid constructions. One of these was designated pTG402 (see Fig. 6).

Deletions extending unidirectionally from the *Eco*RI site of pTG402 were constructed essentially as described by Frischauf et al. (6). Briefly, pTG402 was treated with DNase I in a sufficient amount to linearize half of the molecules, cleaved with *Eco*RI, treated with DNA polymerase I in the presence of all four dNTPs to obtain blunt ends, and size fractionated by agarose gel electrophoresis. Each fraction was ligated in the presence of *Eco*RI linkers and transformed into competent cells of strain MC1000.

pTG428 was used as a source of λ DNA, carrying the *c*I857 gene, p_r promoter, and first 21 codons of the *cro* gene. It was constructed by ligating the *Cla*I-*Bgl*II fragment of λ *c*I857 *S7* (base pairs 36966 to 38103 [22]) into *Cla*I-*Bgl*II-cleaved pGT402, followed by the introduction of an *Eco*RI linker into the *Bgl*II site. As the *Cla*I site is adjacent to an *Eco*RI site, derived from pBR322, plasmid pGT402 has an *Eco*RI fragment of 1,160 base pairs containing the λ repressor and the λ p_r promoter.

Media and reagents. Unless otherwise stated, N-broth or N-agar (13) was used for growth of *E. coli* K-12 cells. Antibiotic supplements were as previously described (27).

The medium used for MC1000 was NY medium (8 g of N-Z amine [Humko, Sheffield, England], 4 g of yeast extract [Lab M, London], 5 g of NaCl, distilled H_2O to 1 liter [pH 7.2 with HCl]). NY agar is NY medium plus 1.5% agar (29), and 20 μg of X-gal (Sigma) per ml was added for identification of β-galactosidase-producing colonies. Unless otherwise stated, the cultivation temperature was 37°C.

Minimal label medium (20) was composed of 80 mM NaCl, 20 mM KCl, 20 mM NH_4Cl, 1 mM $MgCl_2$, 1.5 mM $CaCl_2$, 20 mM Na_2SO_4, 120 mM Tris hydrochloride, 1 mM KH_2PO_4, vitamin B_1 (2 μg/ml), and threonine (50 μg/ml) (pH 7.5).

All enzymes used were from New England Biolabs, Beverly, Mass., or Boehringer Mannheim Biochemicals, Indianapolis, Ind., and were used as recommended by the manufacturer. Nucleotides were from New England Nuclear Corp., Boston, Mass.

^{14}C-labeled amino acids were obtained from Radiochemical Centre, Amersham, England. Anti-*T. pallidum* rabbit antiserum and monoclonal antibodies were used as previously described (27).

DNA technology. Standard procedures were used for preparation of phage λ DNA and plasmid DNA, ligation, and transformation (13). DNA sequencing was performed as described by Maxam and Gilbert (14). End labelings were done either with [γ-^{32}P]ATP and polynucleotide kinase or with the appropriate [α-^{32}P]dNTP and the large fragment of DNA polymerase I (Klenow fragment).

Detection of treponemal antigens in *E. coli* K-12. The expression of treponemal antigens in *E. coli* K-12 was determined by the binding of antibodies to sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (SDS-PAGE)-fractionated proteins on nitrocellulose. This so-called Western blotting was basically done by the method of Burnette (1) with some modifications as described previously (27). As antiserum probe we used an anti-*T. pallidum*

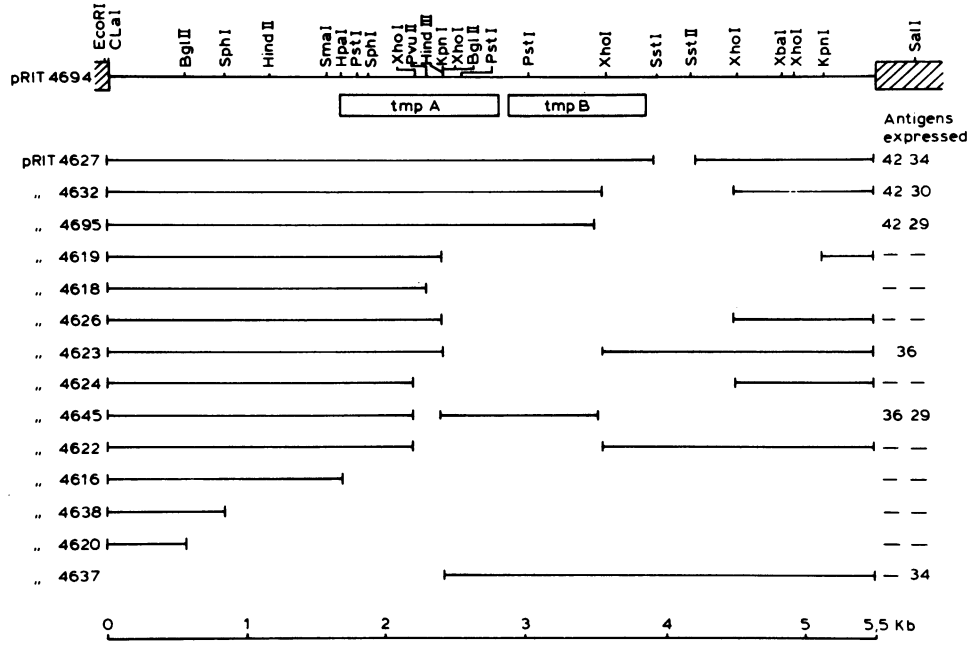

FIG. 1. Physical map and antigen expression of the treponemal DNA segment in pRIT4694 and its derivatives. The crosshatched part represents the vector DNA. The molecular size (in kd) of the expressed proteins that bind anti-*T. pallidum* antibodies as detected by Western blots is indicated on the right. *tmpA* and *tmpB* denote the genes encoding for the 42- and the 34-kd-membrane proteins, respectively.

antiserum that was prepared by intravenous immunization of density gradient-purified *T. pallidum* cells as previously described (27). Serum (1 volume) was absorbed with 1/2 volume of packed *E. coli* K-12 cells that had been ultrasonicated. The activity of β-galactosidase was assayed by the method of Miller (15).

Protein labeling of *E. coli* cells. Plasmid-specified protein synthesis was analyzed in minicells derived from strain P678-54 carrying various plasmids. Minicells were purified from 500-ml overnight cultures grown at 37°C by two cycles of sucrose gradient centrifugation (16). The purified minicells were suspended into the minicell label medium to a cell density resulting in an optical density at 500 nm of 1.0. After 10 min of incubation at 37°C, 25 μCi of a ^{14}C-labeled amino acid mixture was added. The cells were labeled for 30 min and washed with saline, and the proteins were analyzed by SDS-PAGE and Western blotting. The blots were then subjected to autoradiography. Alternatively, the polyacrylamide gels were soaked in Amplify (Amersham Corp., Arlington Heights, Ill.), dried, and autoradiographed.

The induction of protein synthesis in strain K-12.H1.trp carrying various expression plasmids was analyzed as follows. Cells were grown in N-broth at 28°C to a cell density of 3×10^7 cells per ml. The cells were collected and suspended in their original volume of minimal-label medium supplemented with 0.2% glucose, 0.05% yeast extract, 0.05% Casamino Acids (Difco Laboratories, Detroit, Mich.), and 50 μg of L-tryptophan per ml. After 60 min of incubation at 28°C, the temperature was shifted to 42°C. At various times after the temperature shift, the cells were labeled with ^{14}C-amino acid mixture (2.5 μCi/ml final concentration). The cells were washed with saline and boiled for 5 min in 2% SDS–2% dithiothreitol–10% glycerol–0.01% bromophenol blue–50 mM Tris hydrochloride (pH 8.3). The proteins were separated by SDS-PAGE (12). After fixing, staining with Coomassie brilliant blue, and drying, they were autoradiographed.

RESULTS

Mapping of the genes encoding the 42- and 34-kd treponemal antigens. Plasmid pRIT4694 was previously described as the smallest derivative of a parental recombinant cosmid that expressed two protein antigens having molecular weights of 44 and 35 kd (27). These molecular weights differ slightly from our recent, more accurate determinations, which were estimated to be 42 and 34 kd, respectively. A physical map of the 5.5-kb treponemal part of pRIT4694 was constructed (Fig. 1). To map more precisely the position of the treponemal genes involved in the expression of the two antigens, plasmid pRIT4694 was deleted in various regions (Fig. 1) by cleaving the DNA with various restriction endonucleases (Table 1), followed by self ligation at a low DNA concentration and reintroduction of the plasmid *E. coli*. Expression of the treponemal antigens by these mutants was tested by the reaction of anti-*T. pallidum* antibodies with SDS-PAGE-fractionated proteins from lysates of cells carrying the various mutant plasmids. A number of representative examples of these Western blots are shown in Fig. 2, and all data are summarized in Fig. 1.

Deletion of treponemal DNA from the positions at 3.4 and 3.5 kb to the right on the map in Fig. 1 resulted in plasmids pRIT4632 and pRIT4695, respectively. These plasmids expressed the 42-kd protein normally, but the 34-kd antigen was expressed as truncated protein having sizes of 30 and 29 kd, respectively. This indicated that the 42-kd-protein gene is on the left of the 34-kd gene and that transcription of the latter gene is from left to right. The small *Sst*I deletion in plasmid pRIT4627, extending from 3.95 to 4.20 kb, did not affect expression of the 34-kd-protein gene. Therefore, its structural gene must have been located on the left of the *Sst*I site at 3.95 kb.

With the exception of pRIT4623, all mutants depicted in Fig. 1 that were deleted leftwards from the *Xho*I site at 3.6 kb or rightwards from the *Xho*I site at 2.4 kb failed to

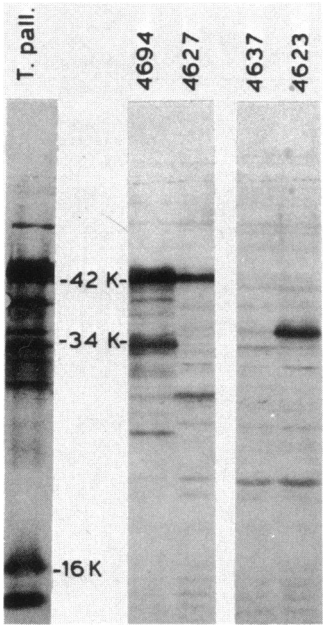

FIG. 2. Expression of treponemal antigens by pRIT4694 and some of its deletion mutants. Lysates of cells carrying the various plasmids were analyzed by Western blotting, using rabbit anti-*T. pallidum* antiserum as a probe. Numbers above the lanes indicate the numbers of pRIT-type plasmids present in strain SC181. T. pall. denotes the lane loaded with a lysate of *T. pallidum*.

produce detectable amounts of treponemal antigen. The deletion mutant pRIT4623, however, expressed a 36-kd protein that reacted with anti-*T. pallidum* antibodies. This 36-kd protein appeared to carry antigenic determinants of both the 42- and the 34-kd antigens. This was shown by incubation of the Western blot with sera that were made specific for either of the two treponemal antigens by absorption of anti-*T. pallidum* serum with ultrasonicated cells carrying either pRIT4637 or pRIT4632, respectively. The 36-kd protein encoded by pRIT4623 reacted with both of the specific antisera (Fig. 3). Therefore, this 36-kd protein was a fusion protein composed of the amino-terminal part of the 42-kd antigen and the carboxy-terminal part of the 34-kd antigen. From the sizes of the deletion in pRIT4623 and of the fusion protein, the position of the 42-kd protein gene relative to that of the 34-kd protein gene could be deduced accurately. It was concluded that both genes were positioned very close to each other with less than 100 base pairs of DNA in between. Furthermore, it followed that the direction of transcription of the 42-kd protein gene was the same as that of the 34-kd gene, i.e., from left to right. Absorption with ultrasonicated cells carrying pRIT4623 did not result in complete removal of antibodies reacting with the 42- and the 34-kd antigens (Fig. 3), indicating that epitopes on these antigens are also present on the nonoverlapping regions of the 36-kd protein with those of the 42- and the 34-kd proteins. Previously, we reported that monoclonal antibodies of hybridoma line 1-14M1 did react with the 42-kd protein (27). These antibodies don't react with the 36-kd fusion protein, indicating that they recognize an epitope on the carboxy-terminal part of the 42-kd protein (Fig. 3).

The position of the two treponemal antigen genes on the map of pRIT4694 is depicted in Fig. 1. As described below, both gene products appeared to be incorporated in the membrane fraction. Therefore, we designated the genes encoding the 42- and the 34-kd treponemal membrane proteins *tmpA* and *tmpB*, respectively.

Expression of *tmpA* and *tmpB* in minicells. The de novo protein synthesis mediated by treponemal DNA was studied in minicells carrying the plasmids pRIT4694 and pRIT4695, respectively. After labeling minicells carrying pRIT4694 with ^{14}C-labeled amino acids, four nonvector-encoded radioactive polypeptides were detected in autoradiographs of polyacrylamide gels (Fig. 4, lanes a and b). The molecular

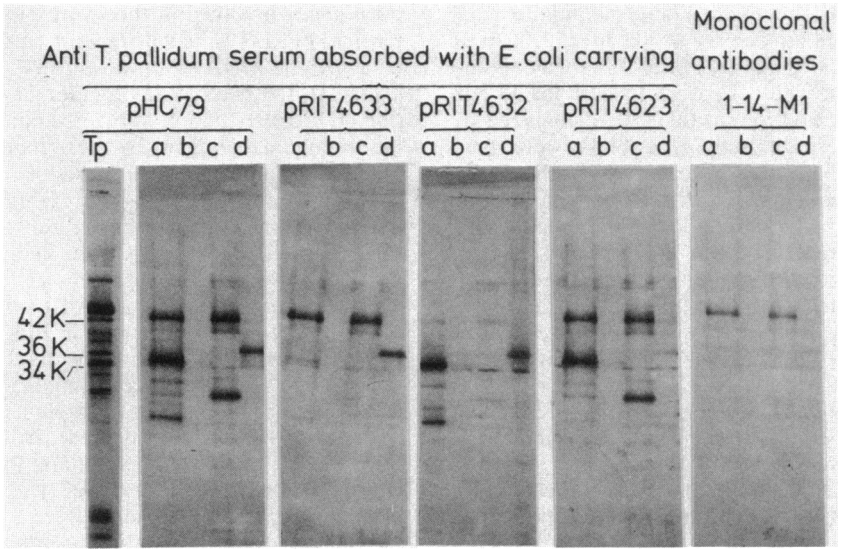

FIG. 3. Reaction of the pRIT4623-encoded, 36-kd fusion protein with anti-*T. pallidum* rabbit antiserum, which was made specific by absorption with ultrasonicated cells carrying various plasmid derivatives. Lanes a, Lysate from SC181 carrying pRIT4694; lanes b, pHC79; lanes c, pRIT4632; lanes d, pRIT4623. Tp, *T. pallidum* protein, which served as a control. Note that absorption of serum with lysates of cells carrying pRIT4633 and pRIT4632 resulted in complete removal of detectable antibodies reacting with the 34- and 42-kd proteins, respectively. The reaction with monoclonal antibodies of hybridoma cell line 1-14M1 is shown on the right.

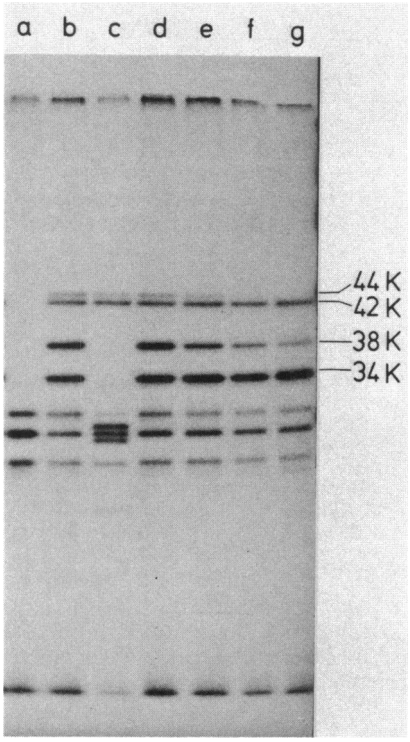

FIG. 4. Electrophoresis of ^{14}C-labeled proteins extracted from minicells, carrying pHC79 (lane a), pRIT4694 (lanes b and d–g), and pRIT4695 (lane c). Lanes a through c were loaded with extracts of minicells labeled for 30 min. Lanes d, e, f, and g were loaded with extracts pulse-labeled for 30 min and chased in nonradioactive medium for 0, 30, 60, and 90 min, respectively.

weights of these polypeptides were 44, 42, 38, and 34 kd. To precisely identify those ^{14}C-labeled proteins that were able to bind anti-treponemal antibodies, lysates of labeled minicells were first analyzed by Western blotting, and after visualization of the antibody binding proteins, the blots were subjected to autoradiography (data not shown). These data unambiguously identified the 42- and the 34-kd polypeptides as the main antibody binding components on Western blots.

Plasmid pRIT4695, which is deleted in the *tmpB* gene, expressed only the 44- and the 42-kd proteins in minicells (Fig. 4, lane c), so it seemed that *tmpA* coded for the 44-kd and the 42-kd proteins, and *tmpB* coded for the 38 and 34-kd proteins. In addition to the 42- and 34-kd polypeptides, the 44- and 38-kd components reacted weakly with anti-treponemal antibodies. This is in contrast to results obtained by Western blots of normal cells, in which the 44- and the 38-kd bands are hardly detected if at all (see Fig. 2). This suggested that the latter polypeptide components were more stable in minicells than in normal cells. Pulse-chase experiments of minicells carrying pRIT4694 showed that more than 90% of the radioactivity in the 44- and the 38-kd proteins disappeared within 1.5 h of incubation in cold medium and that this decrease was accompanied by an increase of labeled 42- and 34-kd proteins, respectively (see Fig. 4, lanes d through f). Thus, it seemed that the 42- and 34-kd proteins were the products of proteolytic cleavage of precursor proteins having sizes of 44 and 38 kd, respectively.

Because membrane proteins are known to be cleaved proteolytically during transport through the cytoplasmic membrane (8), we investigated the cellular localization of the 42- and the 34-kd antigens. The major portions of both antigens were found in the cell envelope fraction (data not shown), whereas breakdown products of the 34-kd protein, which are often observed on Western blots, were present exclusively in the cytoplasmic fraction (data not shown).

Construction of antigen-overproducing strains and mapping of transcription signals of *tmpA* and *tmpB*. Although the gene products of *tmpA* and *tmpB* were detected by Western blotting or after labeling in minicells, the production of these treponemal antigens was poor, because the proteins could not be detected among the *E. coli* proteins present when analyzed in Coomassie blue-stained polyacrylamide gels. For the construction of overproducing strains, we used the

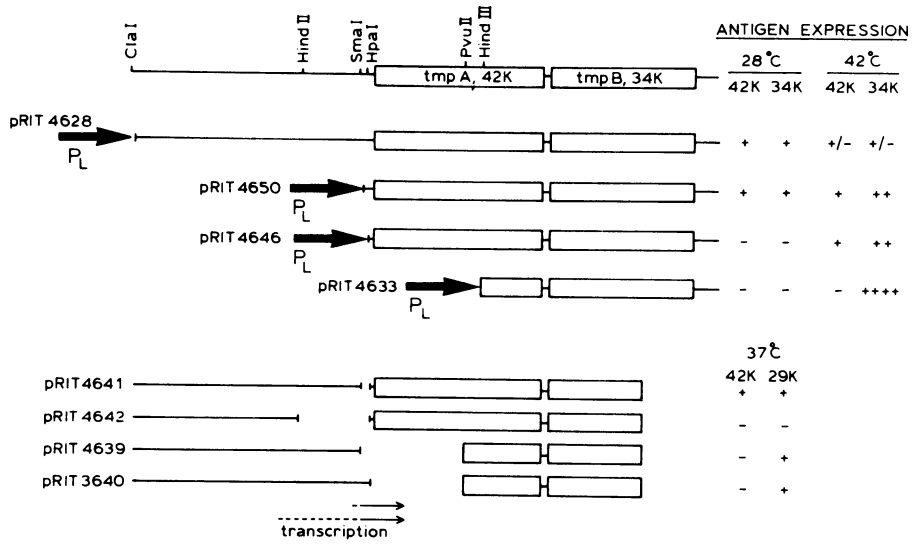

FIG. 5. Expression of *tmpA* and *tmpB* by plasmids carrying the lambda promoter at various distances upstream of these genes and mapping of the transcriptional signals of *tmpA* and *tmpB*. Expression of all plasmid derivatives was measured by Western blotting. Arrows represent both lambda DNA containing the left promoter p_L and the direction of transcription. Plasmids pRIT4639 through pRIT4642 are derivatives of pRIT4695, which is a spontaneous deletion mutant of plasmid pRIT4694.

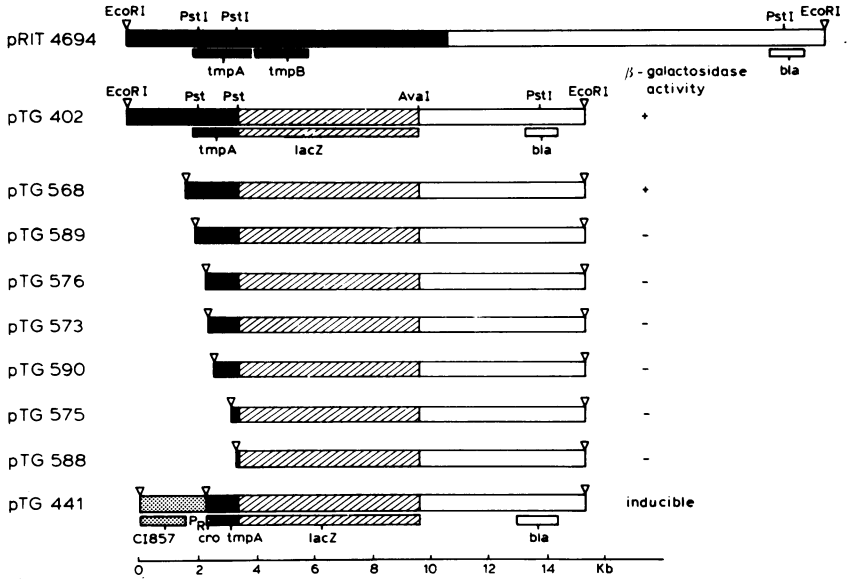

FIG. 6. Genetic and physical maps of plasmids carrying the entire *tmpA* operon of *T. pallidum* (pRIT4694) or segments of the *tmpA* gene fused to the *lacZ* gene of *E. coli*. Black bars (■), *T. pallidum* DNA; hatched bars (▨), *lacZ* DNA; dotted bars (▩), λ DNA; open bars (▭), vector DNA. The vector DNA in pRIT4694 is from pHC79. The other plasmids contain the *Ava*I-*Eco*RI fragment of pBR322 (coordinates 1424 to 4360 [25]). Relevant gene products are indicated below the bars representing plasmid DNA. For further discussion, see the text.

expression plasmid pPLc236 which carries the left promoter, p_L, of bacteriophage lambda. Promoter activity can be fully repressed at 28°C when the thermolabile repressor product of the lambda *c*I857 gene is functional, whereas strong transcription proceeds after heat induction at 42°C (19).

Four different plasmids carrying the p_L were constructed; the position of p_L relative to *tmpA* and *tmpB* is depicted in Fig. 5. In pRIT4628 p_L is located 1.6 kb upstream of *tmpA*, and in contrast to our expectation, production of the 42- and 34-kd proteins was decreased after heat induction of cells carrying this plasmid. By deletion of either the *Cla*I-*Sma*I or *Cla*I-*Hpa*I fragment of pRIT4628, we obtained the plasmids pRIT4650 or pRIT4646, respectively. As will be shown later, the distances between the lambda DNA and *tmpA* in these plasmids amounted to approximately 70 and 10 base pairs, respectively. Both plasmids overproduced the 34-kd protein at 42°C; however, no detectable increase in the production of the 42-kd protein was observed. By placing p_L about 200 base pairs upstream of *tmpB*, a still better producing strain for the 34-kd protein was obtained. The 34-kd *tmpB* protein was present 3 h after induction as the major protein in *E. coli*, carrying pRIT4633. A 10-min pulse-label of protein after 3 h of induction showed that virtually all protein synthesis was that of the 34-kd protein and the 38-kd precursor (data not shown).

In contrast to plasmid pRIT4650, no expression at 28°C of *tmpA* or *tmpB* was found from plasmids pRIT4646 and pRIT4633, indicating that a promoter is located between the *Sma*I and the *Hpa*I sites and is within approximately 70 base pairs of the front of *tmpA*. In addition, this lack of expression showed that expression of both *tmpA* and *tmpB* was dependent on the same transcriptional signal(s). To study in more detail the transcriptional control of *tmpA* and *tmpB*, we constructed the deletion mutants pRIT4639 through pRIT4642 (Fig. 5). The mere deletion of the 60-base-pair *Sma*I-*Hpa*I fragment upstream of *tmpA* or *tmpB* did not abolish expression of these genes (see Fig. 5). Therefore, a second transcriptional signal was located leftwards from the *Sma*I site. Deletion of the 400-base-pair *Hin*dII fragment leftwards from the *Hpa*I site resulted in a total loss of expression of *tmpA* and *tmpB*. Thus, at least two promoters appeared to be involved in the expression of the treponemal genes in *E. coli*. One was located less than 70 base pairs upstream of *tmpA*, and a second one was situated between 70 and 400 base pairs left of *tmpA*.

Determination of the *tmpA* sequence. The DNA sequence was determined in both directions from the unique *Hin*dIII site in pRIT4694 (about 0.5 kilobases in each direction [see Fig. 8]). This allowed identification of the *tmpA* reading frame and subsequent construction of a fusion between this reading frame and that of the *lacZ* gene of pEH253. The *tmpA* gene was fused to the *lacZ* gene to facilitate the analysis of gene expression. Construction of plasmid pTG402, carrying the fused genes, is described above (see Fig. 6). Plasmid pTG402 expressed a 145-kd protein, carrying *T. pallidum* epitopes, as was shown by Western blotting with anti-*T. pallidum* rabbit antiserum. The molecular size of the fragment of the *lacZ* protein in the fusion protein was 116 kd; therefore, the fusion protein carried a *T. pallidum*-derived fragment of about 29 kd. Staining of SDS gels with Coomassie brilliant blue indicated that this protein constituted only a minor fraction of the total protein content in the recombinant strain. The poor expression of the hybrid protein was no surprise, as transcription was initiated from a *T. pallidum* promoter and translation was initiated from a *T. pallidum* ribosome binding site.

The leftward extension of the *tmpA* gene was analyzed by the construction of deletion derivatives of pTG402. Deletions extending unidirectionally from the *Eco*RI site in pTG402 were constructed by DNase I cleavage (6). The deletion derivatives used in this study are shown in Fig. 6. pTG568 was the smallest plasmid (the largest deletion) still expressing β-galactosidase activity. Western blot analysis showed that pTG568 coded for the same hybrid protein as pTG402. The hybrid protein was produced in the same amount from both plasmids, suggesting that in these plas-

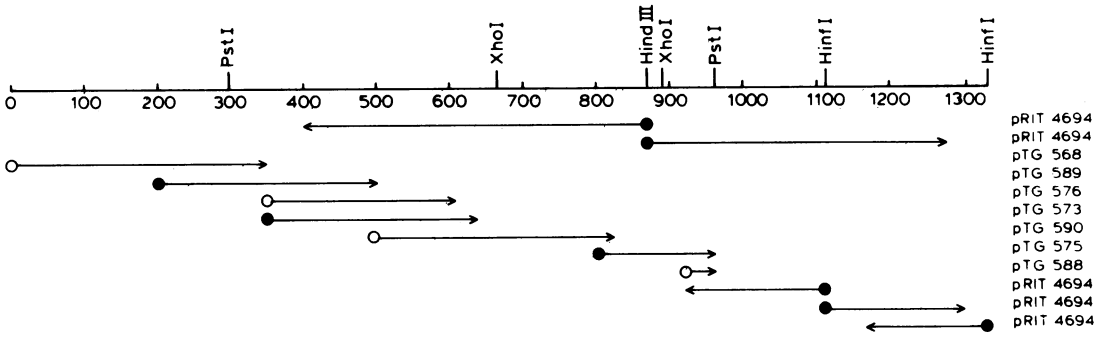

FIG. 7. Sequencing strategy for the establishment of the nucleotide sequence of the *tmpA* gene (see Fig. 8). The location of restriction enzyme cleavage sites relevant for the sequencing strategy is given along the base pair scale. *Hin*fI sites to the left of the *Hin*dIII site are not shown. The site of labeling and the amount of sequence obtained from each fragment are indicated by arrows. ●, 3′ end labeling with Klenow polymerase; ○, 5′ end labeling with polynucleotide kinase. The plasmid used as the source of DNA for each reaction is given to the right of each arrow. The pTG plasmids were labeled in their unique *Eco*RI site (cf. Fig. 1), as described in the text.

mids the same promoter was used to express *tmpA*. This was consistent with the earlier finding that a promoter was present at a distance less than 70 base pairs upstream of *tmpA*.

The deletion derivatives of pTG402 contained unique *Eco*RI sites at the endpoints of the deletions, that is, at various positions in the *tmpA* gene. Determination of the DNA sequence from the *Eco*RI sites of these deletion derivatives gave a set of overlapping sequences which at the same time allowed the sequence of the *tmpA* gene up to position 955 to be established as well as the determination of the precise endpoint of each deletion. The complete sequence strategy is shown in Fig. 7, and the nucleotide sequence of the *tmpA* gene is given in Fig. 8. The sequence presented starts from the first base of the *T. pallidum* DNA in pTG568 since this plasmid was the smallest deletion derivative obtained that was still able to express the fusion protein. Accordingly, the sequence presented contains a *tmpA* promoter as well as the start of the structural gene.

High level expression of the β-galactosidase-linked antigen. To explore the potential use of the fusion protein as an enzyme-linked antigen for serodiagnostic purposes, it was necessary to increase production of the fusion protein since expression of the hybrid protein from plasmid pTG402 was too low for large-scale production. The λ p_R promoter was used to introduce a controllable promoter at various positions in front of or within the *tmpA* gene. To facilitate this analysis, we constructed a plasmid, pTG428, containing the λ p_R promoter as well as the λ *c*I857 repressor gene on a 1,160-base-pair *Eco*RI fragment. Insertion of the λ p_R-containing fragment into the *Eco*RI site of pTG568 did not increase the yield of the hybrid protein. This observation is in accordance with the negative results obtained in the attempt to overproduce native *tmpA* protein by placing the left promoter of phage λ in front of *tmpA*.

Therefore, we constructed a plasmid in which the transcription and translation signals of the λ *cro* gene were used to express the fusion protein. This was accomplished by insertion of the *Eco*RI fragment of pTG428, containing p_R and the N terminus of the *cro* gene, into the *Eco*RI site of pTG576. In the resulting plasmid pTG441, the *cro* gene was fused in frame with the *tmpA-lacZ* hybrid gene. So pTG441 coded for a sandwich protein, starting with 21 amino acid residues of the λ *cro* protein, 6 residues coded for by two tandem *Eco*RI linkers, 204 residues of the *tmpA* protein, 6 residues coded for by the linker region of pEH254, and finally 1,020 residues of β-galactosidase. Induction of a culture of MC1000, carrying pTG441, resulted in a high level of β-galactosidase activity during the first 90 min at 42°C. After 1.5 h at 42°C, the sandwich protein amounted to about 20% of the total protein in pTG441 containing *E. coli* cells. The protein could be purified to about 90% purity by fractional ammonium sulfate precipitation of a crude extract of soluble proteins.

DISCUSSION

Expression of the tmpA-tmpB operon. Of the major protein immunogens during an infection with *T. pallidum*, one has previously been identified as a 42-kd protein (26; R. V. W. van Eijk, unpublished data). In this study we have shown that the structural gene of this protein, *tmpA*, is part of an operon that also contains the structural gene, *tmpB*, for another immunogenic protein of 34 kd. This operon is expressed in *E. coli* from one or more *T. pallidum* transcription signals less than 400 base pairs upstream of *tmpA*. While no data are available on the coordination of expression of *tmpA* and *tmpB* in *T. pallidum*, there seems to be no special reason to believe that these two genes would not be coordinately expressed in *T. pallidum* itself. Although both *tmpA* and *tmpB* are expressed from the same transcription signals in *E. coli*, the proteins are produced in different quantities. This was observed in minicells and also in normal cells when both genes were under the control of the lambda promoter. The *tmpB* product is produced in larger quantities than the *tmpA* product, and, because *tmpB* is located downstream from *tmpA*, this difference in expression is unlikely to be the result of different levels of mRNA. The reason for this conjectured difference in posttranscriptionally controlled expression between *tmpA* and *tmpB* has not been elucidated in this study. However, it seems probable that our inability to overproduce the 42-kd protein by heat induction of p_L- or p_R-containing plasmids is related to this difference in posttranscriptional control of expression between *tmpA* and *tmpB*.

Dependence of the expression of *tmpA* and *tmpB* on the same promoters might suggest that the 42- and 34-kd gene products have some functional relationship. The observation that both proteins are processed posttranslationally and found in the cell envelope of *E. coli* is consistent with this indication. Until now we have had no direct evidence that the 42- and 34-kd proteins in *T. pallidum* are also posttranslational cleavage products and are membrane associated. However, the observation that the apparent molecular

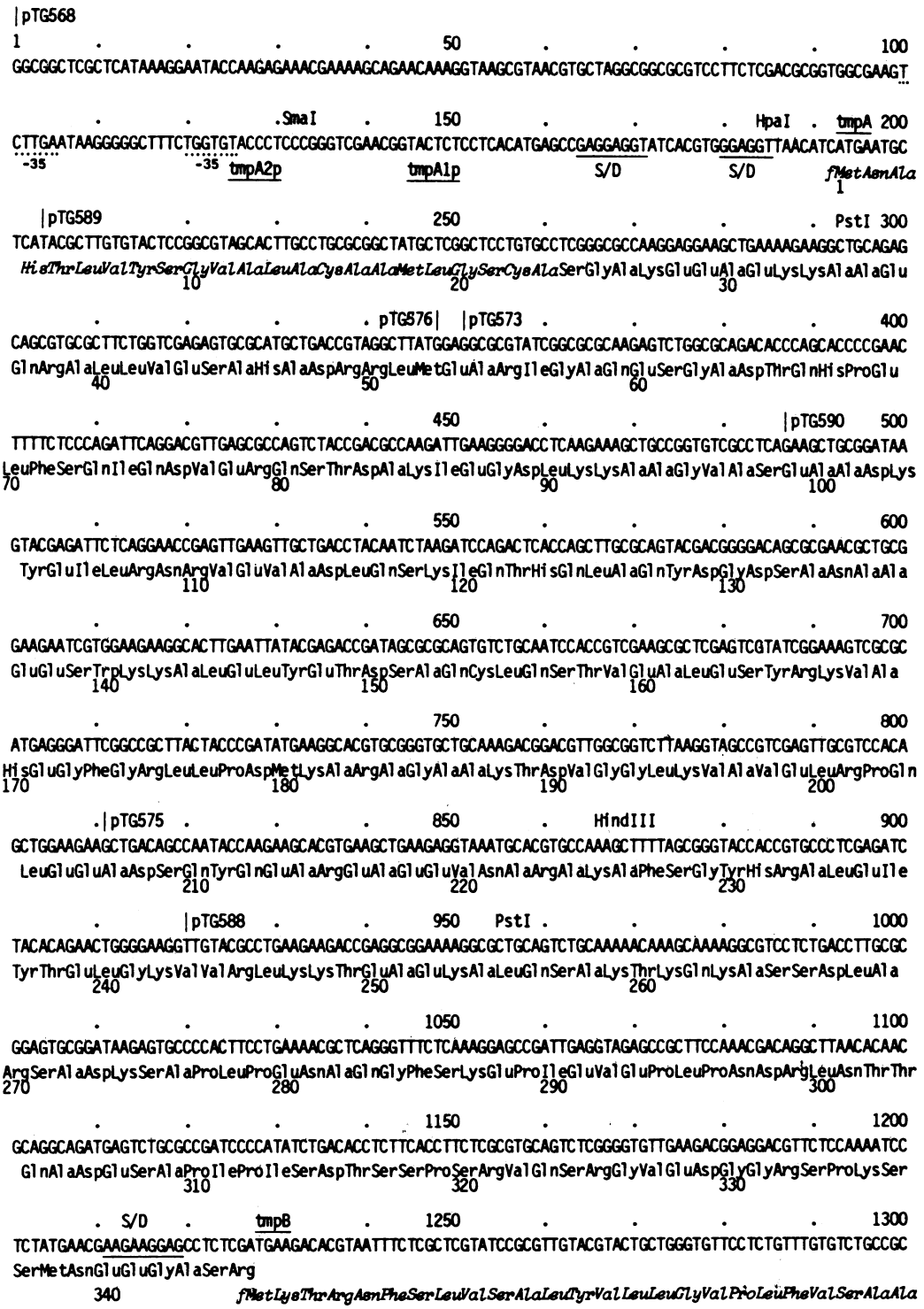

FIG. 8. Nucleotide sequence of the *tmpA* gene of *T. pallidum*. The amino acid sequence of the TmpA protein and the reading frame supposed to be the N terminus of the TmpB protein are given along the nucleotide sequence. Amino acid sequences printed in italics are potential signal peptides. End points of deletion derivatives of pTG402 (cf. Fig. 6) are given by plasmid designation and a vertical line pointing to the first base of *T. pallidum* DNA in this plasmid. Restriction enzyme cleavage sites discussed in the text are given above the sequences.

weights of these proteins in *E. coli* are exactly the same as those in *T. pallidum* strongly suggests that a similar cleavage occurs in this gram-negative organism. Preliminary experiments with specific anti-*tmpB* protein antisera showed binding of antibodies to whole *T. pallidum* cells, by an immunofluorescence test (unpublished data). This suggests that at least the *tmpB* protein is directly exposed to antibodies on the surface of *T. pallidum*.

Location of *tmpA* within the *T. pallidum* DNA sequence. The nucleotide sequence of the region known to contain the

tmpA gene has been determined. Analysis of this nucleotide sequence for open reading frames in both directions revealed only one open reading frame of sufficient length to code for a protein of about 40 kd. Expression of this reading frame in *E. coli* was verified by the construction of a hybrid gene in which this frame was fused to the reading frame of the *lacZ* gene of *E. coli*. The hybrid gene was found to express a protein having TmpA epitopes as well as β-galactosidase activity. By SDS-gel electrophoresis, the protein was found to be 240 to 290 amino acid residues longer than wild-type β-galactosidase. Based on the difference in size between the TmpA-β-galactosidase hybrid protein and wild-type β-galactosidase, the N terminus of the *tmpA* gene is calculated to be located between bases 150 and 300. In this region, four possible start codons for the *tmpA* reading frame are found, at positions 157, 193, 211, and 244. Only one of these potential start codons is preceded by a Shine-Dalgarno (23)-like sequence. This codon is the ATG codon at position 193, and two Shine-Dalgarno sequences are located in front of this start codon (Fig. 8).

tmpA is the first *T. pallidum* gene sequenced; therefore, no comparable sequences of *T. pallidum* ribosome binding sites are known. Because the treponemal translational signals of *tmpA* are recognized by the *E. coli* machinery for protein synthesis, the existence of a Shine-Dalgarno-like sequence in front of the ATG codon at position 193 strongly favors the assignment of this codon as the actual start codon for the *tmpA* gene. Further evidence comes from analysis of the deduced amino acid sequence (discussed below), which reveals the existence of a typical signal peptide only if translation is initiated from the ATG codon at position 193.

Codon usage. The codon usage of the *tmpA* gene (Fig. 9) indicates that *T. pallidum*, like other organisms, has a preferential use of synonymous codons (7). The pattern of codon usage in the *tmpA* gene does not deviate grossly from the pattern seen in *E. coli*. However, some minor differences are worth noticing. The preferred Leu codon in the *tmpA* gene is CTT, whereas the preferred codon in *E. coli* is CTG. None of the four Pro codons seems to be preferred in the *tmpA* gene; in *E. coli* CCG is preferred. Of the two Lys codons, *T. pallidum* seems to prefer AAG, whereas *E. coli* prefers AAA. Finally, the preferred codons for Arg and Gly appear to be the same as for *E. coli*, although the nonpreferred codons are used more frequently than in *E. coli*.

TTT PHE	1	TCT SER	13	TAT TYR	1	TGT CYS	2
TTC PHE	3	TCC SER	6	TAC TYR	7	TGC CYS	1
TTA Leu	2	TCA Ser	3	TAA ***	0	TGA ***	1
TTG Leu	1	TCG Ser	3	TAG ***	0	TGG TRP	1
CTT Leu	12	CCT Pro	2	CAT HIS	3	CGT ARG	9
CTC Leu	5	CCC Pro	3	CAC His	3	CGC ARG	4
CTA leu	2	CCA Pro	4	CAA Gln	6	CGA arg	2
CTG LEU	6	CCG PRO	3	CAG GLN	14	CGG arg	3
ATT ILE	4	ACT THR	1	ATT Asn	2	AGT Ser	3
ATC ILE	4	ACC THR	6	AAC ASN	6	AGC Ser	4
ATA ile	1	ACA thr	3	AAA LYS	5	AGA arg	0
ATG MET	5	ACG Thr	3	AAG Lys	20	AGG arg	3
GTT VAL	6	GCT ALA	14	GAT Asp	5	GGT GLY	5
GTC Val	5	GCC ALA	11	GAC Asp	14	GGC GLY	7
GTA VAL	5	GCA ALA	10	GAA GLU	21	GGA gly	4
GTG VAL	2	GCG ALA	19	GAG Glu	18	GGG gly	4

FIG. 9. Codon usage of the *tmpA* gene of *T. pallidum*. Frequency of codon usage by *E. coli* is indicated as follows (with examples): all capital letters, very frequently used (LEU); initial capital letter, frequently used (Leu); all lower-cased letters, seldom used (leu). Numbers indicate the codon usage in *tmpA*. Asterisks indicate stop codons.

The general pattern is that codons used frequently in *E. coli* are also used frequently in *T. pallidum* and also that the infrequently used codons are the same ones in both bacteria, although *T. pallidum* appears somewhat less discriminatory. The firm establishment of the preferred codon usage of *T. pallidum*, however, would require the sequences of more than one gene.

***tmpA* protein sequence.** The amino acid composition of the protein initiated at position 193 (the *tmpA* preprotein [see below]) deviates from the composition of average proteins (5) only for alanine and glutamic acid. Other amino acids occur approximately in the same ratio as in a standard protein. The *tmpA* protein is, from the amino acid composition, seen to be a weakly acidic protein with a calculated isoelectric point of 4.98 for the preprotein and for the mature protein as well. The hydrophilicity profile and the predicted secondary structure of the TmpA protein, as calculated by the methods of Hopp and Woods (9) and Chou and Fasman (3), respectively, reveal that the TmpA protein is usually a hydrophilic protein. The hydrophilicity profile reveals that the hydrophilic residues are rather evenly distributed along the protein. The only larger hydrophobic domain is the N-terminal part of about 30 amino acids carrying the conjectured signal peptide (see below).

The TmpA and TmpB proteins were found to be processed rapidly in minicells, and the mature products were found to be 2 and 4 kd smaller compared with the precursor proteins of *tmpA* and *tmpB*, respectively. Although we have no direct evidence about the nature of this apparent posttranslational modification, it seems highly probable that the *tmpA* and *tmpB* proteins carry a signal sequence in the amino-terminal part of the protein. Most membrane proteins have been shown to carry such a hydrophobic signal sequence, which is cleaved off during transport over the cytoplasmic membrane (8). Signal sequences from distantly related organisms are known to be recognized and cleaved by *E. coli*; therefore, it would not be surprising if treponemal signal peptides are normally cleaved by *E. coli*.

Signal peptides of eucaryotes and procaryotes have a number of common features. A charged residue (usually basic) within the first five amino acids, is followed by a core of at least nine hydrophobic residues in the α helical conformation (sufficient to span the membrane); a helix-breaking residue is usually found four to eight residues before the cleavage site, and the cleavage site is usually behind an alanine, glycine, serine, threonine, or cystine residue (28, 31). The N terminus of the *tmpA* protein fulfills all of these requirements, if the protein is initiated from the ATG start codon at position 193. Cleavage probability calculated by the method of von Heijne (28) gives the highest probability for cleavage after residue 23 (Ala) and the second highest value for cleavage after residue 26 (Ala). However, only 54% of the known procaryotic cleavage sites are cleaved at the sites giving the highest or second highest probabilities (3); therefore, the method of von Heijne cannot be used reliably to discriminate among the four possible cleavage sites.

We were able to show that the N terminus of the mature protein is probably close to the site predicted from the sequence analysis. This was done by comparing the apparent molecular weight of the mature *tmpA* protein with a hybrid protein in which the N terminus of the λ *cro* gene is fused to residue 53 of the *tmpA* protein. This hybrid was

constructed from pTG441 by reintroducing the original C terminus of the *tmpA* gene (data not shown). This hybrid protein was found to have a mobility identical to that of the mature *tmpA* protein. Since 27 amino acid residues of the *cro* protein were added in front of residue 53 of the *tmpA* protein, this identical motility shows that the N terminus of the mature *tmpA* protein is very close to residue 26 in the sequence shown.

Presently, nothing is known about the function of the *tmpA* protein in *T. pallidum*, and because no genetic techniques are available to make site-specific mutations in the genome of this noncultivable microorganism, we compared the sequence of the *tmpA* protein with those published for procaryotic and eucaryotic proteins, whose functions usually are known. All 14 segments of 25 amino acids of the *tmpA* protein were compared, in segments of 25 residues, to all 2,893 protein sequences in the database of the National Biomedical Research Foundation by using the program SEARCH. This program calculates the mutational distance between protein segments with a scoring matrix called the Mutation Data Matrix (5). The *tmpA* protein showed no significant homology to any single protein in this database. It is worth noticing, however, that several signal peptides were found among the protein segments that gave the highest scores with the N terminus of the *tmpA* sequence.

Location of the *tmpB* gene. As discussed previously, the *tmpB* structural gene was located within 100 base pairs after the *tmpA* gene. The start codon for the *tmpB* gene is therefore expected to be within the sequence presented here. Three start codons are located in reading frames which are not terminated within the sequence presented here. Of these three start codons, only the ATG codon at position 1227 is preceded by a Shine-Dalgarno sequence at the proper distance. The Shine-Dalgarno sequence found at position 1211 (AAGAAGGAG) can be aligned with the 3′ end of 16S rRNA (of *E. coli*) in two different ways to give either six consecutive or seven nonconsecutive base pairs. The amino acid sequence coded for by this open reading frame resembles a typical signal peptide. Since the *tmpB* protein is produced in *E. coli* cells as a preprotein, the presence of a Shine-Dalgarno sequence in front of an open reading frame able to code for a signal peptide makes it tempting to suggest that this reading frame is the N terminus of the *tmpB* preprotein. If this is actually the case, the *tmpA* and *tmpB* genes are overlapping genes, the overlap being one nucleotide (or 4 nucleotides if the TGA termination codon is also regarded as belonging to *tmpA* gene).

***T. pallidum* expression signals.** The experiments presented revealed two promoters, *tmpA1P*, located between the *Sma*I site at position 131 (Fig. 9) and the structural *tmpA* gene, and *tmpA2P*, located less than 300 base pairs to the left of this *Sma*I site. The DNA sequence between positions 131 and 193, which should contain *tmpA1P*, reveals one sequence resembling a Pribnow box (18). The −35 sequence corresponding to this Pribnow box shows no extensive homology to the consensus sequence (21); furthermore, this −35 sequence is located to the left of the *Sma*I site. The existence of *tmpA1P* is based on the expression at low temperature of the *tmpA* antigen from plasmid pRIT4650. This plasmid contains a small segment of pBR322-derived DNA between the p_L promoter and the *Sma*I site at position 131. This segment of pBR322 DNA is a 24-base-pair *Eco*RI-*Cla*I fragment containing the −35 sequence from the *tet* promoter (25). The sequence across the *Cla*I-*Sal*I junction in pTG4650 will then be as follows:

pBR322 < >*T. pallidum*
TTCTCATGTTTGACAGCTTATCAT GGGTCGAACGGTACTCTCC
−35 Pribnow box
tetP' *'tmpA1P'*

This sequence is almost identical to the consensus sequence for *E. coli* promoters.

The *tmpA1P* promoter identified in pRIT4650 might therefore be an artifact created at the junction between pBR322 and *T. pallidum* DNA. At least the *tmpA1P* of pRIT4650 does not contain the same −35 sequence as the native *tmpA1P*; therefore, the strength might be altered substantially. A potential treponemal promoter sequence, *tmpA2*, is located immediately to the left of the *Sma*I site (Fig. 9). The −35 sequence for this potential promoter agrees well with the consensus sequence for *E. coli* promoters, whereas the Pribnow box is TACCCT (consensus TAtaaT).

The sequence has been analyzed for promoters of the *E. coli* type since the genes are transcribed in *E. coli*. As the sequence is known to contain at least one promoter, it is reasonable to suggest that one (or both) of the promoterlike sequences identified above is actually a promoter (in *E. coli*). It is presently not known whether the genes are transcribed from the same promoter(s) in *T. pallidum* as in *E. coli*. However, the ability to express a variety of different *T. pallidum* antigens in *E. coli* (17, 24, 27, 30) indicates that the RNA polymerase of *E. coli* is able to recognize *T. pallidum* promoters.

In contrast to our expectation, deletion of the *Sma*I-*Hpa*I fragment (plasmid pRIT4641 [Fig. 6]) did not result in a complete loss of expression of *tmpA*, although this region carries the presumed translation signals for *tmpA*. The reason for this result remains obscure. An explanation might be that in the region between positions 86 and 99 various sequences are found that have a 4- or 5-base-pair homology with the 3′ end of *E. coli* 16S RNA. These sequences might act as ribosome binding sites of an mRNA that is initiated upstream.

Use of *T. pallidum* proteins for serodiagnostic purposes. Unfortunately, we have been unable to construct plasmids capable of hyperexpression of *tmpA*. However, the *tmpA-lacZ* fusions constructed in this study gave rise to strains overproducing hybrid proteins carrying *tmpA* epitopes and concurrently β-galactosidase enzyme activity. Preliminary experiments have shown that the enzyme-linked antigen can be used for the detection of antibodies against the *tmpA* protein in hyperimmune rabbit serum by a simple double-sandwich, solid-phase immunoassay (11; unpublished data). Studies to evaluate the potential value of this fusion protein and the highly expressible *tmpB* protein in the serodiagnosis of syphilis are in progress.

ACKNOWLEDGMENT

We are grateful to E. Remaut for providing p_L-containing expression plasmids, G. Aelbers for propagation and preparation of *T. pallidum* bacteria, Nils Axelsen for stimulating discussions, Winona C. Barker for the comparison of the TmpA protein sequence to the sequence in the database of the *Atlas of Protein Sequence and Structure* (M. O. Dayhoff [ed.], National Biomedical Research Foundation, Washington, D.C.) and to Randi Brundstedt and Tineke Olij for perfect secretarial assistance.

LITERATURE CITED

1. **Burnette, W. N.** 1981. "Western blotting": electrophoretic transfer of proteins from sodium dodecyl sulfate-polyacrylamide gels to unmodified nitrocellulose and radiographic detection with antibody and radioiodinated protein A. Anal. Biochem. **112**:195–203.
2. **Casadaban, M. J., and S. N. Cohen.** 1980. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. J. Mol. Biol. **138**:179–207.
3. **Chou, P. Y., and G. D. Fasman.** 1974. Prediction of protein conformation. Biochemistry **13**:222–245.
4. **Collins, J., and H. J. Bruning.** 1978. Plasmids usable as gene-cloning vectors in an in vitro packaging by coliphage: "cosmids". Gene **4**:85–107.
5. **Dayhoff, M. O., W. C. Barker, and L. T. Hunt.** 1983. Establishing homologies in protein sequences. Methods Enzymol. **91**:524–545.
6. **Frischauf, A. M., H. Garoff, and H. Lehrach.** 1980. A subcloning strategy for DNA sequence analysis. Nucleic Acids Res. **8**:5541–5549.
7. **Grantham, R., C. Gautier, and M. Gouy.** 1980. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. Nucleic Acids Res. **8**:1893–1912.
8. **Hall, M. N., and T. J. Silhavy.** 1981. Genetic analysis of the major outer membrane protein of *Escherichia coli*. Annu. Rev. Genet. **15**:91–142.
9. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. U.S.A. **78**:3824–3828.
10. **Kalnins, A., K. Otto, U. Rüther, and B. Müller-Hill.** 1983. Sequence of the *lacZ* gene of *Escherichia coli*. EMBO J. **2**:593–597.
11. **Koenen, M., U. Rüther, and B. Müller-Hill.** 1982. Immunoenzymatic detection of expressed gene fragments cloned in the *lacZ* gene of *E. coli*. EMBO J. **1**:509–512.
12. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
13. **Maniatis, T., E. F. Frisch, and J. Sambrook.** 1982. Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
14. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing endlabeled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
15. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Mooi, F. R., N. Harms, D. Bakker, and F. K. de Graaf.** 1981. Organization and expression of genes involved in the production of the K88ab antigen. Infect. Immun. **32**:1155–1163.
17. **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.
18. **Pribnow, D.** 1975. Nucleotide sequence of an RNA polymerase binding site at an early T7 promoter. Proc. Natl. Acad. Sci. U.S.A. **72**:784–788.
19. **Remaut, E., P. Stanssens, and W. Fiers.** 1981. Plasmid vectors for high-efficiency expression controlled by the P_L promotor of coliphage lambda. Gene **15**:81–93.
20. **Roozen, K. J., R. G. Fenwick, Jr., and R. Curtiss III.** 1971. Synthesis of ribonucleic acid and protein in plasmid-containing minicells of *Escherichia coli* K-12. J. Bacteriol. **107**:21–33.
21. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
22. **Sanger, F., A. R. Coulson, G. F. Hong, D. F. Hill, and G. B. Peterson.** 1982. Nucleotide sequence of bacteriophage λ DNA. J. Mol. Biol. **162**:729–772.
23. **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. U.S.A. **71**:1342–1346.
24. **Stamm, L. V., J. D. Folds, and P. J. Bassford, Jr.** 1982. Expression of *Treponema pallidum* antigens in *Escherichia coli* K-12. Infect. Immun. **36**:1238–1241.
25. **Sutcliffe, J. G.** 1979. Complete nucleotide sequence of the *E. coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43**:77–90.
26. **Van Eijk, R. V. W., and J. D. A. van Embden.** 1982. Molecular characterization of *Treponema pallidum* proteins responsible for the human immune response to syphilis. Antonie van Leeuwenhoek J. Microbiol. **48**:486–487.
27. **van Embden, J. D., H. J. van der Donk, R. V. van Eijk, H. G. van der Heide, J. A. de Jong, M. F. van Olderen, A. D. Osterhaus, and L. M. Schouls.** 1983. Molecular cloning and expression of *Treponema pallidum* DNA in *Escherichia coli* K-12. Infect. Immun. **42**:187–196.
28. **Von Heijne, G.** 1983. Patterns of amino acids near signal-sequence cleavage sites. Eur. J. Biochem. **133**:17–21.
29. **Von Meyenburg, K., B. B. Jørgensen, J. Nielsen, and F. G. Hansen.** 1982. Promoters of the *atp* operon coding for the membrane bound ATP synthase of *Escherichia coli* mapped by Tn*10* insertion mutations. Mol. Gen. Genet. **188**:240–248.
30. **Walfield, A. M., P. A. Hanff, and M. A. Lovett.** 1982. Expression of *Treponema pallidum* antigens in *Escherichia coli*. Science **216**:522–523.
31. **Watson, M. E. E.** 1984. Compilation of published signal sequences. Nucleic Acids Res. **12**:5145–5163.