JOURNAL OF BACTERIOLOGY, Oct. 1984, p. 395–401
0021-9193/84/100395-07$02.00/0
Copyright © 1984, American Society for Microbiology
Vol. 160, No. 1

DNA Sequence of the F *traALE* Region That Includes the Gene for F Pilin

LAURA S. FROST,[1] W. PARANCHYCH,[1]* AND NEIL S. WILLETTS[2]

Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada,[1] and Department of Molecular Biology, University of Edinburgh, Edinburgh EH9 3JR, Scotland[2]

Received 30 April 1984/Accepted 10 July 1984

The complete sequence of a 1.4-kilobase *Pst*I fragment containing the F transfer genes *traA*, *-L*, and *-E* is presented. The *traA* reading frame has been located both genetically and by comparing the primary structure of F pilin (the *traA* product) predicted by the DNA sequence to the amino acid composition and sequence of N- and C-terminal peptides isolated from purified F pilin. Taken together, these data show that there is a leader peptide of 51 amino acids and that F pilin contains 70 amino acids, giving molecular weights of 13,200 for F propilin and 7,200 for mature F pilin. Secondary structure predictions for F pilin revealed a reverse turn that precedes the sequence Ala-Met-Ala_{51}, a classic signal peptidase cleavage site. The N-terminal alanine residue is blocked by an acetyl group as determined by ^{1}H-nuclear magnetic resonance spectroscopy. The *traL* and *traE* genes encode proteins of molecular weights 10,350 and 21,200, respectively. According to DNA sequence predictions, these proteins do not contain signal peptide leader sequences. Secondary structure predictions for these proteins are in accord with *traLp* and *traEp* being membrane proteins in which hydrophobic regions capable of spanning the membrane are linked by sequences that form turns and carry positively charged residues capable of interacting with the membrane surface.

F pili are filamentous surface appendages of *Escherichia coli* strains harboring the F plasmid. They are required for establishing cell-to-cell contact during bacterial conjugation, and they also act as the site of attachment for a number of F-specific bacteriophages. F pili are hollow cylinders which are 8 nm in diameter, with an inner diameter of 2 nm. They are composed of a single, repeating subunit, F pilin, arranged in a helical manner with 3.6 U per turn and a pitch of 1.28 nm (9). F pili have been purified to homogeneity and characterized with respect to their amino acid composition and to the molecular weight of the pilin subunit which has previously been estimated at 11,500; they were found to contain no arginine, histidine, proline, or cysteine (4, 6, 7).

Although the biological properties of F pili have been studied in some detail, little is known about the structure-function relationships of these organelles. Although the F pilus is an apparently simple structure (9), the complexity of its synthesis and assembly is demonstrated by the fact that 14 or more genes in the F transfer (*tra*) operon are required for F pilus formation (33). So far, only two of these genes *traA* and *traQ*, have been assigned a role. The F pilin subunit has been shown to be encoded by the *traA* gene (19), which is located at the promoter proximal end of the 33-kilobase (kb) *traY*→*Z* mRNA transcript. An F pilin pool has been located in the inner membrane (20), and it has been postulated that the *traA* product of 13,000 daltons is processed to mature F pilin, of molecular weight 7,000, and that this cleavage may be mediated by *traQ*, a gene located ca. 14 kb downstream from *traA* in the *traYZ* operon (21).

To understand the F pilus structure and to relate this to its function, it is essential to know the sequence of the F pilin protein. F pili are present in low numbers (one to two per cell) in a bacterial culture, and this factor, together with the high hydrophobicity of F pilin protein and its resistance to dissociation by reagents such as urea or guanidine hydrochloride (3), has made purification and protein sequencing difficult. Both F and the closely related ColB2 pilin (8) contain highly insoluble, intransigent regions which made protein sequencing difficult, and furthermore their N termini are blocked (4). Also, no information on the putative leader sequence of F pilin could be gained from the protein sequence of the mature F pilin subunit. Therefore, the *traA* gene was isolated on a 1.4-kb DNA fragment and sequenced. Information on the protein sequence at the N and C termini was used to confirm the DNA sequence, the presence of a leader peptide, and the size of the F pilin subunit.

MATERIALS AND METHODS

Bacterial plasmids. Plasmid pED891 is a derivative of pBR328 (26) containing the 1.4-kb *Pst*I fragment found in the *Eco*RI fragment f6 of the F plasmid JCFL0 (33). Complementation of F *lac* mutants showed that pED891 carried the three genes *traA*, *traL*, and *traE*. pED851 is a chimera of pBR322 containing the entire F transfer region (12). JCFL1 is F *lac traA1* with an amber mutation in *traA* (2).

DNA sequence determination. The 1.4-kb *Pst*I fragment (100 μg) was electroeluted from a 0.7% agarose gel after digestion of pED891 (500 μg) with *Pst*I and electrophoretic separation of the DNA fragments. The purified fragment was digested with *Sau*3A and cloned into M13mp7 (17) and sequenced by the dideoxynucleotide method of Sanger et al. (24). The sequence near the *Pst*I site was determined by using pBR322 *Pst*I primer P1 (32). Overlapping sequences were determined by using single end-labeled DNA fragments isolated and sequenced by the method of Maxam and Gilbert (16).

F pilin purification. F pili were isolated from *E. coli* K-12 JE2571 (Fim^- Fla^- Sm^r Nal^r Rif^r) which has been freshly transformed with pED851. They were purified by the method described previously (3). F pilin was prepared by treating purified pili with 1% sodium dodecyl sulfate and precipitating the pilin protein with acetone to remove excess detergent (11).

Proteolytic Digestion of F pilin. F pilin was digested with carboxypeptidase A in 0.1 M NH_4HCO_3 (pH 8.0)–0.1%

* Corresponding author.

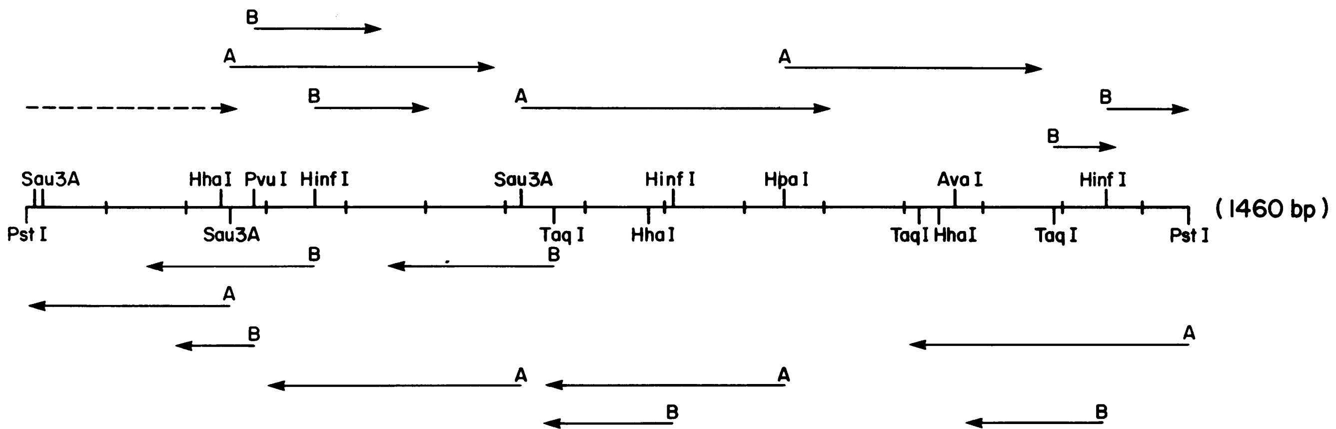

FIG. 1. Sequencing strategy for *traALE*. The 1.4-kb *Pst*I fragment containing *traA*, *traL*, and *traE* was sequenced by the Sanger dideoxynucleotide method (A), Maxam and Gilbert reactions (B), and the *Pst*I primer, P1 (- - -). Details are given in the text.

sodium dodecyl sulfate for 30 to 180 min at 37°C, with an enzyme-substrate ratio of 1:20. Samples equivalent to 20 to 50 nmol of protein were removed, acidified with 1 M acetic acid, lyophilized, and analyzed for amino acids as previously described (22). F pilin was digested with pronase as previously described (11).

Materials and media. Restriction enzymes, T4 DNA ligase, the large fragment of DNA polymerase I (Klenow), T4 polynucleotide kinase, and M13 26-base-pair (bp) primer were purchased from Bethesda Research Laboratories, Gaithersburg, Md., and used as suggested by the supplier. pBR322 *Pst*I primer P1 (12 bases) was supplied by Pharmacia-PL Biochemicals, Uppsala, Sweden. [α-^{32}P]dATP (3,000 Ci/mmol) and [γ-^{32}P]ATP (1,000 to 3,000 Ci/mmol) were obtained from New England Nuclear Corp., Boston, Mass., as were reagents for Maxam and Gilbert sequencing reactions. Nucleotide mixes used in dideoxynucleotide sequencing were supplied in a kit from New England Biolabs, Beverly, Mass.

Carboxypeptidase A was supplied by Worthington Diagnostics, Freehold, N.J., and pronase was supplied by Calbiochem-Behring, Los Angeles, Calif.

Synthetic peptides of the N termini of F pilin (N-Ac-Ala-Gly-Ser-Ser-Gly-Gln-Asp-Leu) and of ColB2 pilin (N-Ac-Ala-Gln-Gly-Gln-Asp-Leu) were provided by R. S. Hodges, Department of Biochemistry, University of Alberta, Edmonton, Canada.

RESULTS

Cloning and sequencing strategy. The f6 (*Eco*RI) fragment of the F transfer operon (33) contains the first seven genes of the operon, including the gene for F pilin, *traA*. A 1.4-kb *Pst*I fragment containing the C-terminal region of *traY* plus the coding regions for *traA*, *traL*, and *traE* was cloned into pBR328 to give the chimeric plasmid pED891. This plasmid was the source of the 1.4-kb *Pst*I fragment used for sequencing. The sequencing strategy is given in Fig. 1, and the sequence of the *Pst*I fragment is given in Fig. 2, where restriction sites for commercially available endonucleases are also indicated. Three nonoverlapping open reading frames were identified, and confirmation is presented below that these specify the *traA*, *traL*, and *traE* proteins.

These nonoverlapping open reading frames were preceded by potential Shine-Dalgarno sequences, followed by an initiating methionine codon in each case. No other substantial open reading frames were present in the three reading frames on each strand.

Location of the *traA* gene. The 1.4-kb *Pst*I fragment was partially digested with *Sau*3A, and the resultant fragments were subcloned into the *Bam*HI site of pBR322. Clones were screened for the presence of the *traA* gene by assaying for the ability to complement JCFL1, which carries a point mutation in *traA* (2).

Two chimeras, pLF31 and pLF32, were found to restore transfer ability to JCFL1 (0% transfer ability) at low levels (10^{-6} and 10^{-4}%, respectively) as compared with the transfer ability of the wild-type plasmid JCFL0 (100% transfer ability).

The chimera pLF31 contained the 369-bp *Sau*3A fragment (positions 245 to 614, whereas pLF32 contained a larger region (positions 26 to 614) including the 218-bp *Sau*3A fragment (positions 26 to 244) upstream and adjacent to the 369-bp *Sau*3A fragment. Inspection of the sequence argued against the 369-bp *Sau*3A fragment including the complete *traA* gene since there was no Shine-Dalgarno sequence (25) upstream from the methionines at positions 263, 266, and 287 and since the open reading frame continued upstream into the 218-bp *Sau*3A fragment to position 134. The presence of a potential Shine-Dalgarno sequence beginning at position 129 and a methionine codon at position 139 suggested that this is the true N terminus of the *traA* gene product. This would give a protein of 121 amino acids with a molecular weight of 13,200, which agrees with the findings of Kennedy et al. (13), who estimated that the *traA* gene product was 13,700 daltons by using minicells or in vitro translation techniques.

The low level of complementation given by the 369-bp *Sau*3A fragment can be explained once it is realized that cloning this fragment into the *Bam*HI site in the tetracycline gene gives rise to a fusion peptide consisting of the first 89 amino acids at the N terminus of the A179 protein (28) and the last 86 amino acids of F propilin. The A179 protein is a highly hydrophobic membrane protein which can presumably mimic the leader sequence of F propilin, resulting in processing of the leader peptide and insertion of the pilin into the inner membrane. The 369-bp *Sau*3A fragment, when inserted into a vector such as pUC8 (30) in which such a fusion protein would not result, failed to complement at all.

Sizing the F pilin protein. To locate the precise point at which the *traA* product propilin was cleaved to give F pilin,

it was necessary to obtain data directly from the protein. The purification of F pilin in large enough amounts for conventional protein chemistry and sequencing is now possible by using the chimera pED851 which both contains the complete F transfer region cloned into pBR322 and synthesizes three- to fivefold more pilin than does F itself (12). Purified F pilin was used for characterizations of N and C termini.

Examination of the predicted amino acid sequence of the *traA* protein reveals a putative signal peptidase cleavage site between alanine$_{51}$ and alanine$_{52}$ (-Ala$_{49}$-Met$_{50}$-Ala$_{51}$-Ala$_{52}$-), suggesting that residues 52 to 121 may represent the mature pilus protein. Summing up the amino acids in F pilin from alanine$_{52}$ to leucine$_{121}$ gives a protein with an amino acid composition which agrees well with that for F pili reported by Date et al. (7), Brinton (6), and this laboratory (3) based on five lysine residues per monomer (Table 1).

F pilin was found to contain a blocked N terminus on the basis of dansylation and automated Edman sequencing studies. To characterize the N terminus, 300 to 500 nmol of F pilin was treated with pronase, and any neutral or acidic peptides were isolated by cation-exchange chromatography by the procedure previously described (11). The amino acid composition of this peptide mixture is given in Table 2. This peptide preparation was insensitive to cadmium-ninhydrin reagent (11) after electrophoresis at pH 2.1 on Polygram Cel 400 cellulose thin-layer sheets (Brinkmann Instruments Inc., Westbury, N.Y.), suggesting that the peptides contained therein had blocked N termini.

The observed composition of the blocked N-terminal pilus peptide is consistent with the predicted N-terminal sequence. However, it is evident that the preparation contains a population of peptides of various lengths, all having the

FIG. 2. DNA sequence of *traA*, *-L*, and *-E*. The sequence of a 1.4-kb *Pst*I fragment encoding the C-terminal portion of *traY* and the complete *traA*, *-L*, and *-E* genes is presented. Shine-Dalgarno sequences are underlined, and the nucleotides are marked every 25 bases with a dot.

TABLE 1. Amino acid composition of F pilin

Amino acid	Residues per F pilin monomer[a]			
	A	B	C	D
Asp	4.7	3.9	4.0	4
Thr	4.0	3.9	4.0	4
Ser	5.8	6.1	5.5	6
Glu	2.3	2.2	2.0	2
Gly	8.8	7.2	7.0	7
Ala	8.8	7.8	7.5	9
Val	13.3	10.6	10.5	14
Met	4.7	4.4	4.0	5
Ile	3.2	2.2	2.0	3
Leu	5.2	4.4	4.5	5
Tyr	1.3	1.1	1.0	1
Phe	3.8	3.3	3.5	4
Trp	1.0	1.1	1.0	1
Lys	4.9	5.0	5.0	5
Arg	0	0	0	0
His	0	0	0	0
Pro	0	0	0	0
Cys	0	0	0	0

[a] Amino acid analyses were taken from (A) Armstrong et al. (3), (B) Date et al. (7), (C) Brinton (6), and (D) the DNA sequence presented here. All previous analyses were adjusted to give five lysines per monomer. The amino acid composition determined from the DNA sequence used residues Ala_{52}-Leu_{121}.

same N terminus. Moreover, the presence of small amounts of Thr, Val, and Ile suggests that the preparation also contains small quantities of one or more contaminating acidic or neutral peptides. These peptides were not detected with ninhydrin on the thin-layer electropherogram, suggesting that they were probably heterogeneous and in low concentration.

^{1}H-nuclear magnetic resonance spectroscopy of the foregoing peptide sample (data not shown) showed clearly that the blocking group is an acetyl moiety with methyl protons showing a resonance peak at 2.05 ppm. The acetyl methyl protons were in a 1:1 relationship with the protons in the methyl group of alanine. This observation is analogous to that obtained with the N terminus of another conjugative pilus protein, EDP208, in which the N-terminus is N-Ac-Thr (11) and similar to that observed with the acetylated N terminus of ColB2 pilin, which begins with N-Ac-Ala (8).

To provide further proof that the N terminus of mature F pilin is N-Ac-Ala-Gly-Ser-Ser-Gly-Gln-Asp-Leu, the acetyl peptide was prepared synthetically by a similar protocol to that previously described (35) and then tested for its ability to compete with pure intact pili for antipilus antibodies by a competition enzyme-linked immunosorbent assay procedure. The results (to be reported in detail elsewhere) showed that the synthetic peptide competes very strongly for polyclonal antipilus antibodies. In contrast, the synthetically prepared N-terminal region from ColB2 pili (8) competed only weakly with anti-F pilus antibodies. These observations provide strong support for the proposed structure of F pilin.

Finally, purified F pilin (40 nmol per sample) was treated with carboxypeptidase A by the method of Narita (22). The digestion was carried out for 3 h at 37°C. The amino acid analyses of the released amino acids are given in Table 3. The release of leucine in nearly quantitative amounts agreed with the C terminus for F pilin predicted from the DNA sequence. Glycine is a poor substrate for carboxypeptidase A, which could explain the slow release of the penultimate glycine residue. The two valine residues before glycine were presumably also being released in lower amounts. No other amino acids appeared in the digestion in significant amounts.

Structural predictions for F propilin (*traA*), *traL*, and *traE*. By using values for each amino acid residue derived by Levitt (14) and by averaging these values over six amino acids at each position, predictions for the secondary structure were made for the three proteins encoded by *traA*, *traL*, and *traE* (Fig. 3); these are discussed below.

DISCUSSION

The sequence of a 1.4-kb *Pst*I fragment derived from the F transfer region is presented here. Three genes, *traA*, *-L*, and *-E* are present on this fragment as is the C-terminal portion of *traY*. These three genes are encoded by the only plausible open reading frames on either strand of the DNA. All three genes have an initiating methionine codon preceded by a Shine-Dalgarno ribosome binding site. The three genes encode proteins of molecular weights 13,200, 10,350, and 21,200. The *traA* gene (13,200 daltons) encodes F propilin, the precursor of the F pilus subunit, pilin. The first 280 bp have recently been published by Fowler et al. (10) and are identical to the sequence presented here.

The findings of Kennedy et al. (13) and Moore et al. (21) showed that the immediate gene product from *traA* is 13,000 to 13,700, which agrees well with the size of the protein encoded by the *traA* open reading frame downstream from the Shine-Dalgarno sequence at position 134. Moore et al. (21) have reported that two protein bands of molecular weights 13,000 and 7,000 react with anti-F pilin rabbit serum and that the *traQ* gene product is required for processing *traAp*, the immediate *traA* gene product, to mature F pilin of molecular weight 7,000. We have also found that F pilin migrates at 7,200 daltons, provided that 0.75 mm or thinner Laemmli polyacrylamide gels are used. The thicker gels used previously (6, 7) were apparently unable to resolve small-molecular-weight proteins (6,000 to 12,000) accurately. Studies on the N and C termini of F pilin confirm that the immediate gene product, propilin, is cleaved after $alanine_{51}$, giving pilin 70 amino acid residues with a leader sequence of 51 amino acid residues. The amino acid sequence of pilin

TABLE 2. Composition of peptides released from F pilin by pronase digestion[a] and fractionated by cation-exchange chromatography[b]

Amino acid	Composition[c] (nmol)	No. of residues (based on Glx = 1.0)	DNA sequence prediction for nine-residue N terminus[d]
Asx	14.2	0.9	1
Thr	7.0	0.4	0
Ser	23.0	1.5	2
Glx	15.7	1.0	1
Gly	25.0	1.6	2
Ala	18.2	1.2	1
Val	6.7	0.4	0
Met	3.2	0.2	1
Ile	4.1	0.3	0
Leu	6.9	0.4	1

[a] The conditions for the proteolytic digestion are given in the text.
[b] AG 50W X8 resin (50 to 100 mesh; Bio-Rad Laboratories, Richmond, Calif.) was used in a column (0.8 by 7 cm).
[c] The effluent (acidic peptides in deionized water) was passed through a cation-exchange column a second time before subjecting it to amino acid compositional analysis.
[d] Ala-Gly-Ser-Ser-Gly-Gln-Asp-Leu-Met.

predicted from the DNA sequence corresponds very well with the amino acid composition of F pilin protein determined previously (3, 6, 7). Similar studies of tryptic peptides derived from ColB2 pili, closely related to F pili, show that ColB2 propilin is cleaved at an analogous position (8).

The amino acid sequence preceding alanine$_{52}$ (Ala-Met-Ala) forms a classic signal peptidase cleavage site. Thus the leader sequence for pilin is 51 amino acids in length, which is unusually long for a procaryotic leader sequence (18). Since *traQ* is required for processing propilin and since the signal peptidase cleavage site at alanine$_{51}$ resembles that of the signal peptidase of the host cell, it is possible that processing propilin to pilin is a two (or more)-step process. These various steps do not seem to include transporting the pilin to the outer membrane since pools of mature pilin are found in the inner membrane (20). Other posttranslational modifications may include glycosylation and phosphorylation (4, 6, 7). The N terminus of F pilin is acetylated, and this is analogous to the acetylated N termini of ColB2 and EDP208 pilin (8, 11). Preliminary studies have shown that a synthetic peptide representing nine residues at the N terminus of F pilin reacts strongly with anti-F pilus antibodies. Work is currently in progress to delineate the antigenic region more precisely and to determine which residues are exposed on the surface of the pilus.

A prediction for the secondary structure of F propilin (Figure 3) assists in imagining the insertion of propilin into the inner membrane, followed by its cleavage by a signal peptidase in preparation for its assembly into a mature F pilus. The propensity for a particular region of the protein to assume a β-structure or an α-helix or to be involved in a turn indicating a change of direction in the peptide chain is given in Fig. 3. The hydrophobicity and polarity index along the peptide chain is also shown. The first eight amino acids are hydrophobic and capable of forming either a β-structure or an α-helix. Because a β-structure is less compact than an α-helix, it can span a typical membrane with ca. 9 amino acids instead of the 12 or more required for an α-helix. If the first eight to nine amino acids in F propilin form a β-structure, then the β-transorption model of Steiner et al. (27) can be invoked. This model suggests that the hydrophobic strand entering the membrane can form a β-sheet structure with one or more extended polypeptide chains of existing transmembrane protein (a receptor) and is thus drawn through the membrane. This suggests a possible function for one of the many transfer gene products that remain uncharacterized. The next five amino acids Gly-Ala-Ser-Ala-Pro encompass a strong probability for a reverse turn since the amino acids glycine, serine, and proline are disrupters of both β- and α-helical structures (23). von Heijne and Blomberg (31) have suggested that transmembrane proteins typically carry positively charged residues on the cytoplasmic side of the membrane and that these residues are usually found in conjunction with a more disordered secondary structure to facilitate binding to the membrane.

After the reverse turn centered at residue 11, there is a sequence of 19 amino acids which is characterized by seven positively charged amino acids, three contiguous lysines at the beginning of the sequence, and then a positive residue every additional three or five residues. The predicted structure for this region favors an α-helix with the possibility of some β-structure at the amino terminus. If this region did form an α-helix whose positive charges interacted with the cytoplasmic side of the membrane, the positive charges might be expected to orient along one side of the helix. Analyzing residues 15 to 33 on a helical wheel with 3.6 amino acids per turn gave Lys$_{15}$, Lys$_{22}$, and Arg$_{33}$ clustered on one side, with the other four positively charged residues roughly equally spaced around the wheel.

TABLE 3. Composition of amino acids released from F pilin by carboxypeptidase A digestin[a]

Amino acid released	Amt detected (nmol)
Gly	11.8
Val	14.6
Leu	32.8

[a] The condition of the proteolytic digestion are given in the text. Pilin (40 nmol) were subjected to carboxypeptide A digestion. The predicted C-terminal sequence is -Val-Val-Gly-Leu-COOH.

Residues 34 to 45 could form an α-helix that would span the membrane, followed by the postcore sequence pro-gln-leu-ala-met-ala which favors a reverse turn and contains the classic recognition sequence for a signal peptidase. The active site of this putative enzyme may be positioned on the periplasmic side of the inner membrane, whereas the first part of the pilin polypeptide chain could be pulled into the membrane. This would serve to anchor the pilin protein in the inner membrane where pools of processed pilin protein have been observed (20).

The F pilin sequence itself has reverse turns centered at alanine$_{61}$ and threonine$_{70}$, which would allow the protein to fold back on itself and span the membrane at least twice. The structure of the bulk of the protein (residue 70 to 121) could be either an α-helix or a β-sheet. Previous studies in which circular dichroism was used to determine the conformation of F pilin in its native state estimated that there was ca. 70% α-helix in F pilin after its incorporation into a pilus (3, 7). This does not preclude the possibility of a conformational transition in F pilin as it leaves the inner membrane and a lipid environment and aggregates in the helical array characteristic of F pili.

Recently, Lory et al. (15) proposed, for exotoxin A, a model of protein excretion via Bayer junctions in gram-negative bacteria. They suggested that the inner and outer membranes are continuous at the junction and that the exotoxin molecule is inserted into the inner membrane via its leader sequence. The leader sequence is cleaved away as the protein passes through the junction and into the outer membrane from which it is released into the medium. This model is helpful in explaining how pilin protein could pass from the inner to outer membrane. No basal structure for pili has been identified so far, and it is presumed that the pilus itself rises from the outer membrane, whereas pools of pilin protein are located in the inner membrane. Bayer (5) suggested that F pili were extruded at or near these pores and that conjugation might involve localized fusion of the membranes of the donor and recipient cell by apposition of such pores between mating cells (33). Thus a process whereby pili are assembled at Bayer junctions and disaggregated during retraction, concomitantly bringing the mating cells together, is attractive. This model predicts that the F pilus tip may recognize some feature of Bayer junction architecture of the recipient cell.

One approach to the question of structure-function relationships in conjugative pili is to identify changes in the sequence of the pilin protein that affect the function of the pilus with respect to its biological properties. To this end, we are currently sequencing point mutants reducing sensitivity to F-specific phages, but not preventing conjugation (34), and characterizing the pilin genes of other plasmids carrying F-like transfer regions (33).

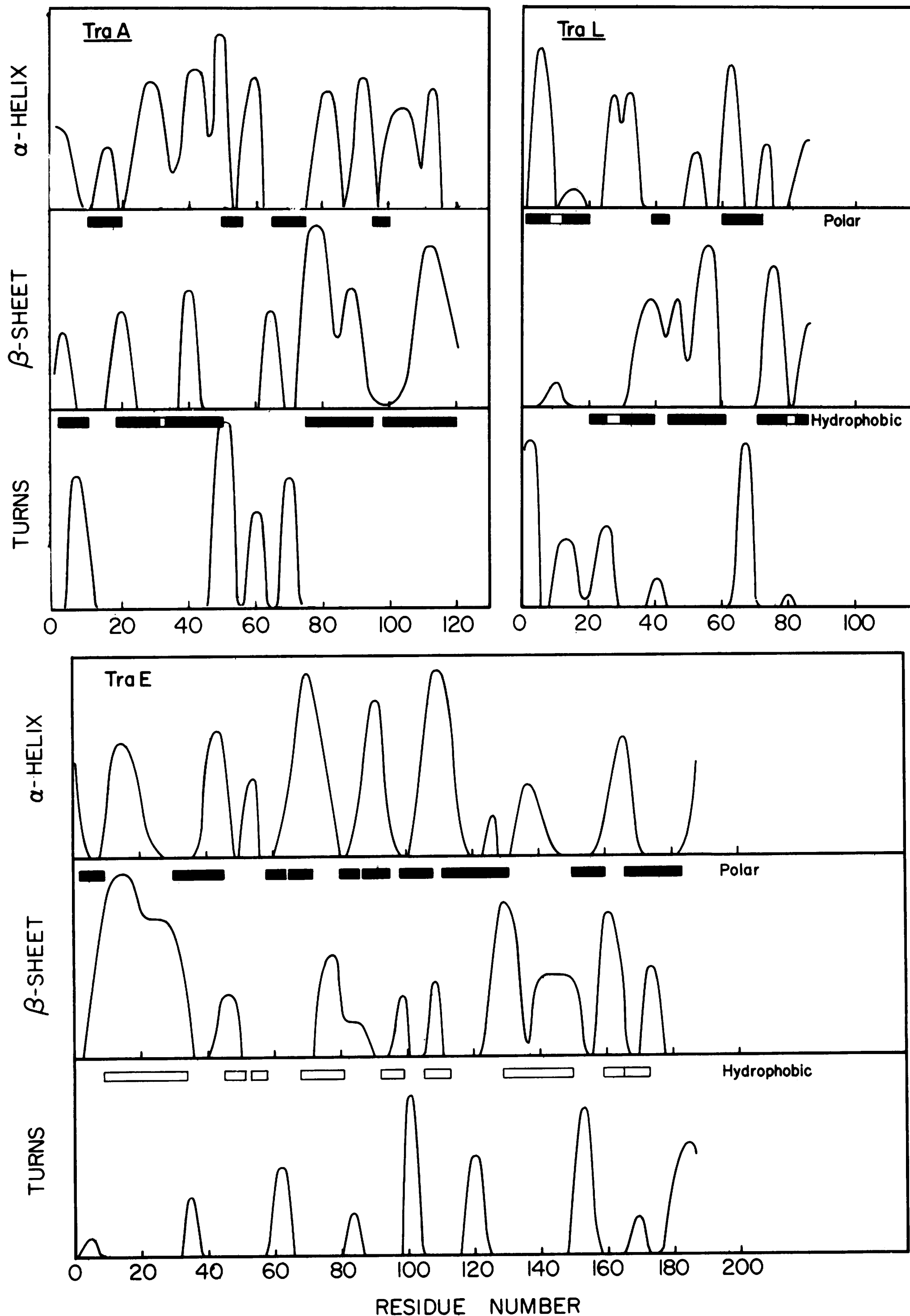

FIG. 3. Predicted secondary structure of *traAp*, *traLp*, and *traEp*. The sequence of F-pilin (*traAp*), plus its leader sequence, was analyzed for α-helix, β-sheet, and reverse turns according to the values for each amino acid residue suggested by Levitt (14). Similar predictions for *traLp* and *traEp* are also calculated. Each point represents an average of the values for six consecutive amino acids.

The *traL* and *traE* genes encode proteins required for F pilus assembly. *traL* protein has been identified as an outer membrane protein of molecular weight 11,000 (13). *traE* protein is found in the inner membrane and has a molecular weight of 19,000 as determined by minicell experiments and by in vitro translation (13). The molecular weights of these proteins, as determined from the DNA sequence, are 10,350 (*traLp*) and 21,200 (*traEp*). Although Achtman et al. (1) reported that the *traE* protein is apparently not cleaved during insertion into the membrane, the size descrepancy between the *traE* protein from the DNA sequence (21,000) and that determined by sodium dodecyl sulfate-polyacryl-

amide gel electrophoresis (19,000) could be due to an as yet undetected leader sequence. It is interesting to note that the proteins encoded by *traM* and *traJ* also do not appear to have leader sequences (10, 29). Inspection of the amino acid sequence and the secondary structure predictions for *traLp* and *traEp* shows that there is no obvious signal peptidase site after a reverse turn as found in the *traA* protein sequence. The predicted secondary structures of *traLp* and *traEp* follow a general pattern which is typical of membrane proteins in which blocks of hydrophobic amino acid residues, of sufficient length to form an α-helix or β-structure capable of spanning a membrane, are connected by regions containing charged amino acids (positively charged residues predominate) which specify a turn in the protein. Thus the general structure is of a peptide chain passing back and forth through the membrane with charged groups interacting with the membrane surface. The predicted structure of *traLp* is that of two α-helices bracketed by reverse turns, followed by two regions of β-structure connected by a segment that strongly predicts a reverse turn. The most prominent feature of the predicted structure of *traEp* is the presence of three to four α-helices connected by reverse turns from residues 50 to 120. The amino and carboxyl regions of the protein favor β-structure. The number of charged amino acids increases toward the C terminus with seven of eight amino acids carrying a charge at the C terminus itself. It is not yet know what the functions of the *traE* and *traL* proteins are; it is hoped that a knowledge of the sequence may suggest a method of probing for these proteins, thereby elucidating their contribution to pilus assembly and function.

ACKNOWLEDGMENTS

We thank Kathy Volpel for pili preparations, Mike Nattriss for amino acid analyses, and John Maule and Wendy Smith for technical support.

This work was supported by Medical Research Council Program grant G978 88 (N.S.W.) and by the Medical Research Council of Canada (W.P.).

LITERATURE CITED

1. **Achtman, M., P. A. Manning, C. Edelbluth, and P. Herrlich.** 1979. Export without proteolytic processing of inner and outer membrane proteins encoded by F sex factor *tra* cistrons in *Escherichia coli* minicells. Proc. Natl. Acad. Sci. U.S.A. **76**:4837–4841.
2. **Achtman, M., N. Willetts, and A. J. Clark.** 1972. Conjugational complementation analysis of transfer-deficient mutants of F*lac* in *Escherichia coli*. J. Bacteriol. **110**:831–842.
3. **Armstrong, G. D., L. S. Frost, P. A. Sastry, and W. Paranchych.** 1980. Comparative biochemical studies on F and EDP208 conjugative pili. J. Bacteriol. **141**:333–341.
4. **Armstrong, G. D., L. S. Frost, H. J. Vogel, and W. Paranchych.** 1981. Nature of the carbohydrate and phosphate associated with ColB2 and EDP208 pilin. J. Bacteriol. **145**:1167–1176.
5. **Bayer, M. E.** 1976. Role of adhesion zones in bacterial cell-surface function and biogenesis, p. 393–427. *In* A. Tzagoloff (ed.), Membrane biogenesis. Plenum Publishing Corp., New York.
6. **Brinton, C. C., Jr.** 1971. The properties of sex pili, the viral nature of "conjugal" genetic transfer systems, and some possible approaches to the control of bacterial drug resistance. Crit. Rev. Microbiol. **1**:105–160.
7. **Date, T., M. Inuzuka, and M. Tomoeda.** 1977. Purification and characterization of F pilin from *Escherichia coli*. Biochemistry **16**:5579–5585.
8. **Finlay, B. B., L. S. Frost, and W. Paranchych.** 1984. Localization, cloning, and sequence determination of the conjugative plasmid ColB2 pilin gene. J. Bacteriol. **160**:402–407.
9. **Folkhard, W., K. R. Leonard, S. Malmsey, D. A. Marvin, J. Dubochet, A. Engel, M. Achtman, and R. Helmuth.** 1979. X-ray diffraction and electron microscope studies on the structure of bacterial F pili. J. Mol. Biol. **130**:145–160.
10. **Fowler, T., L. Taylor, and R. Thompson.** 1983. The control region of the F plasmid transfer operon: DNA sequence of the *traJ* and *traY* genes and characterization of the *traY→Z* promoter. Gene **26**:79–89.
11. **Frost, L. S., G. D. Armstrong, B. B. Finlay, B. F. P. Edwards, and W. Paranchych.** 1983. N-terminal amino acid sequencing of EDP208 conjugative pili. J. Bacteriol. **153**:950–954.
12. **Johnson, D. A., and N. S. Willetts.** 1980. Construction and characterization of multicopy plasmids containing the entire F transfer region plasmid. Plasmid **4**:292–304.
13. **Kennedy, N., L. Beutin, M. Achtman, R. Skurray, U. Rahmsdorg, and P. Herrlich.** 1977. Conjugation proteins encoded by the F sex factor. Nature (London) **270**:580–585.
14. **Levitt, M.** 1978. Conformational preferences of amino acids in globular proteins. Biochemistry **17**:4277–4285.
15. **Lory, S., P. C. Tai, and B. D. David.** 1983. Mechanism of protein excretion by gram-negative bacteria: *Pseudomonas aeruginosa* exotoxin A. J. Bacteriol. **156**:695–702.
16. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base specific chemical cleavage. Methods Enzymol. **65**:499–580.
17. **Messing, J., R. Crea, and P. H. Seeburg.** 1981. A system for shotgun DNA sequencing. Nucleic Acids Res. **9**:309–322.
18. **Michaelis, S., and J. Beckwith.** 1982. Mechanism of incorporation of cell envelope proteins in *Escherichia coli*. Annu. Rev. Microbiol. **36**:435–466.
19. **Minkley, E. G., Jr., S. Polen, C. C. Brinton, Jr., and K. Ippen-Ihler.** 1976. Identification of the structural gene for F-pilin. J. Mol. Biol. **108**:111–121.
20. **Moore, D., B. A. Sowa, and K. Ippen-Ihler.** 1981. Location of an F-pilin pool in the inner membrane. J. Bacteriol. **146**:251–259.
21. **Moore, D., B. A. Sowa, and K. Ippen-Ihler.** 1982. A new activity in the F*tra* operon which is required for F-pilin synthesis. Mol. Gen. Genet. **188**:459–464.
22. **Narita, K., H. Matsuo, and T. Nakajima.** 1975. End group determination. Mol. Biol. Biochem. Biophys. **8**:30–103.
23. **Perlman, D., and H. O. Halvorson.** 1983. A putative signal peptidase recognition site and sequence in eukaryotic and prokaryotic signal peptides. J. Mol. Biol. **167**:391–409.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.
25. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature (London) **254**:34–38.
26. **Soberon, X., L. Covarrubias, and F. Bolivar.** 1980. Construction and characterization of new cloning vehicles. IV. Deletion derivatives of pBR322 and pBR325. Gene **9**:287–305.
27. **Steiner, D. F., P. S. Quinn, S. J. Chan, J. Marsh, and H. S. Tager.** 1980. Processing mechanisms in the biosynthesis of proteins. Ann. N.Y. Acad. Sci. **343**:1–16.
28. **Sutcliffe, J. G.** 1978. Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbor Symp. Quant. Biol. **43**:77–90.
29. **Thompson, R., and L. Taylor.** 1982. Promoter mapping and DNA sequencing of the F plasmid transfer genes *traM* and *traJ*. Mol. Gen. Genet. **188**:513–518.
30. **Viera, J., and J. Messing.** 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
31. **von Heijne, G., and C. Blomberg.** 1979. Transmembrane translocation of protein. Eur. J. Biochem. **97**:175–181.
32. **Wallace, R. B., M. J. Johnson, S. V. Suggs, K. Miyoshi, R. Bhatt, and K. Itakura.** 1981. A set of synthetic oligodeoxyribonucleotide primer for DNA sequencing in the plasmid vector pBR322. Gene **16**:21–26.
33. **Willetts, N., and R. Skurray.** 1980. The conjugation system of F-like plasmids. Annu. Rev. Genet. **14**:41–76.
34. **Willetts, N. S., P. M. Moore, and W. Paranchych.** 1980. Variant pili produced by mutants in the F*lac* plasmid. J. Gen. Microbiol. **117**:455–464.
35. **Worobec, E. A., A. K. Taneja, R. S. Hodges, and W. Paranchych.** 1983. Localization of the major antienic determinant of EDP208 pili at the N-terminus of the pilus protein. J. Bacteriol. **153**:955–961.