JOURNAL OF BACTERIOLOGY, Jan. 1989, p. 213–221
0021-9193/88/010213-09$02.00/0

Vol. 171, No. 1

Nucleotide Sequence of *traQ* and Adjacent Loci in the *Escherichia coli* K-12 F-Plasmid Transfer Operon

JUNE H. WU† AND KARIN IPPEN-IHLER*

Department of Medical Microbiology and Immunology, Texas A&M University, College Station, Texas 77843

Received 1 July 1988/Accepted 10 October 1988

The F *tra* operon region that includes genes *trbA*, *traQ*, and *trbB* was analyzed. Determination of the DNA sequence showed that on the *tra* operon strand, the *trbA* gene begins 19 nucleotides (nt) distal to *traF* and encodes a 115-amino-acid, M_r-12,946 protein. The *traQ* gene begins 399 nt distal to *trbA* and encodes a 94-amino-acid, M_r-10,867 protein. The *trbB* gene, which encodes a 179-amino-acid, M_r-19,507 protein, was found to overlap slightly with *traQ*; its start codon begins 11 nt before the *traQ* stop codon. Protein analysis and subcellular fractionation of the products expressed by these genes indicated that the *trbB* product was processed and that the mature form of this protein accumulated in the periplasm. In contrast, the protein products of *trbA* and *traQ* appeared to be unprocessed, membrane-associated proteins. The DNA sequence also revealed the presence of a previously unsuspected locus, *artA*, in the region between *trbA* and *traQ*. The *artA* open reading frame was found to lie on the DNA strand complementary to that of the F *tra* operon and could encode a 104-amino-acid, 12,132-dalton polypeptide. Since this sequence would not be expressed as part of the *tra* operon, the activity of a potential *artA* promoter region was assessed in a *galK* fusion vector system. In vivo utilization of the *artA* promoter and translational start sites was also examined by testing expression of an *artA*–β-galactosidase fusion protein. These results indicated that the *artA* gene is expressed from its own promoter.

The F pilus, a cylindrical surface structure composed of F-pilin subunit helixes, is essential to formation of the recipient-cell contact that leads to conjugal DNA transfer from F^+ donor cells (for reviews, see references 9, 31, and 32). F-pilus outgrowth is a complex procedure that requires expression of a large number of genes in the transfer region of F, the fertility factor of *Escherichia coli* K-12. The precise function of most of these gene products is not known. However, the direct product of the structural gene for F pilin, *traA*, is known to be a 121-amino-acid polypeptide that must undergo processing and modification to form the mature pilin subunit (6, 12, 21). Expression of the F *traQ* gene is required for efficient expression of the 70-amino-acid pilin polypeptide (10, 12, 23).

Although pilin subunit synthesis is clearly *traQ* dependent, the nature of the *traQ* product and the mechanism through which it acts have been unclear. The studies of Moore et al. (23) showed that TraQ activity stemmed from the F *tra* operon region between *traF* and *traH*. Recent cloning and polypeptide product analyses placed three genes in this region in the order *traF*, *trbA*, *traQ*, *trbB*, *traH* (35). This work also demonstrated that TraQ activity was associated with the expression of a protein of apparent molecular weight (M_a) 12,500, whereas *trbA* and *trbB* expressed products of M_a 14,200 and 18,400, respectively. As a first step toward characterizing these products, we determined the DNA sequence of the region of the F transfer operon that includes these genes. Additional analyses were also performed to test two predictions that derive from the sequence we obtained. These data suggest that the *trbB* gene product is transported to the periplasm and undergoes signal sequence processing. They also indicate that the DNA just prior to *traQ* includes a functional promoter and translational start site for expression of the anti-*tra* operon strand.

* Corresponding author.
† Present address: Cell Biology Department, Baylor College of Medicine, Texas Medical Center, Houston, TX 77030.

MATERIALS AND METHODS

Bacterial strains and plasmids. The construction and origin of plasmids containing F *tra* operon segments are described in Table 1. The relevant region of *tra* DNA carried by these plasmids is depicted in Fig. 1. Strain JM103 [Δ(*lac-pro*) *thi strA supE endA sbcB15 hsdR4* (F′ *traD36 lacI*^{q} *lacZ*Δ*M15 proAB*) (18)], plasmid pMC1403 (1), and phage vectors M13mp8 and M13mp9 (19) were obtained from R. Young, Texas A&M University, College Station. Plasmids pK01 (17) and pK06 (26) and strains C600 *galK* (F^- *galK thi-1 thr-1 leuB6 lacY1 tonA21 supE44*) and N100 (F^- *galK recA pro*) were obtained from E. G. Minkley, Jr., Carnegie Mellon Institute, Pittsburgh, Pa. Plasmids pDR720 and pDR540 (28) were purchased from Pharmacia LKB Biotechnology Inc. SE5000, an F^- *araD139 lacU169 rpsL relA thi recA56* strain, was used for maxicell labeling experiments (24).

DNA cloning and sequencing. The cloning of F DNA fragments followed standard procedures (15). For sequencing, the two F *tra* *Pst*I-*Sma*I fragments were first cloned from pKI181 to M13mp8 and M13mp9. Subsequently, *Sau*3A, *Taq*I, or *Alu*I restriction enzyme-digested DNA fragments derived from these two *tra* DNA segments were inserted into M13mp8 and M13mp9 *Bam*HI, *Acc*I, or *Hin*cII sites, respectively. These clones, together with appropriate primers, were used to determine the complete sequence of both DNA strands. DNA-sequencing methods were as described previously (34). The University of Wisconsin Genetics Computer Group (UWGCG) and IntelliGenetics, Inc. (BIONET), programs were used for sequence analyses.

Galactokinase assays. The level of galactokinase expressed by a strain C600 *galK* derivative that carries plasmid pKI351 (see Fig. 5) was determined as described by Newman et al. (26). The [^{14}C]galactose was purchased from Amersham Corp., Arlington Heights, Ill. DE81 2.5-cm-diameter filter disks (Whatman, Inc., Clifton, N.J.) and Scinti-Verse counting fluid (Fisher Scientific Co., Pittsburgh, Pa.) were used. All galactokinase levels reported are the average of several assays.

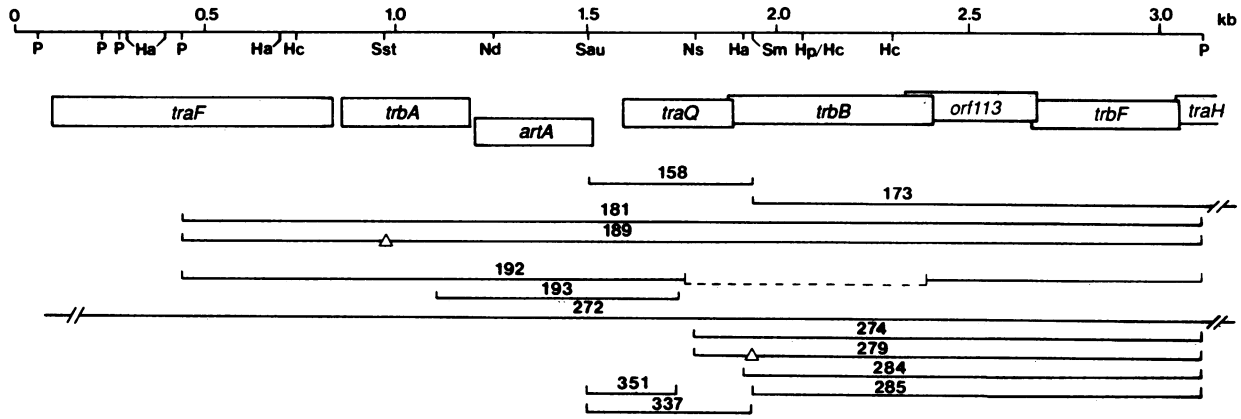


FIG. 1. Map of the F-plasmid *traQ* region. Boxes indicate the positions and lengths of open reading frames. Numbered lines indicate the segments carried by corresponding pKI plasmids; breaks indicate small (△) or large (- - -) deletions, and - // - indicates that the segment extends beyond the *tra* region shown. Numbers at the top indicate nucleotide length in kilobases, starting with the sequence of Wu et al. (34); restriction sites shown on these lines are: *Hae*II (Ha), *Hin*cII (Hc), *Hpa*I (Hp), *Nde*I (Nd), *Nsi*I (Ns), *Pst*I (P), *Sau*3A (Sau), *Sma*I (Sm), and *Sst*I (Sst).

Strains carrying other plasmids which express *galK* from known promoters were also assayed as controls. Numbers obtained were comparable with values reported elsewhere (28; Pharmacia LKB Biotechnology Products catalogue). The number of GalK units obtained for expression of the pK01 control (no promoter) was 22. For pDR720 (*trp* promoter-operator), GalK unit values were 670 (uninduced) and 1,230 (induced); for pDR540 (*tac* promoter, *lac* operator), GalK unit values of 41 (uninduced) and 949 (induced) were obtained. Induced values of pDR720 and pDR540 were obtained by adding 10 μl of either 3-β-indoleacrylic acid (5 μg/ml) or isopropyl-β-D-thiogalactoside (IPTG; 1 M) to 10-ml cultures 1 h before cell lysis. Strain C600 *galK* was the host for all assays except those of pDR540, in which the *lacI*^{q} host JM103 was used. During these experiments, we noticed that galactokinase levels appeared to fluctuate somewhat according to the growth phase of the culture tested; harvesting slowly growing cultures at a low optical density (650 nm) appeared to result in unusually high values, and this was avoided.

β-Galactosidase assays. Expression of β-galactosidase from plasmid pKI337 (see Fig. 5) was assayed in strain SE5000 according to the procedure described by Miller (20) except that a mixture of amino acids (final concentration, 400 μg/ml) was added to the minimal growth medium.

Detection and localization of plasmid protein products. All plasmids were introduced into strain SE5000, and plasmid products were labeled with [^{35}S]methionine or ^{14}C-amino acid mix according to the maxicell procedure described previously (24). When indicated, ethanol (final concentration, 10%) was added to the UV-irradiated cells before addition of [^{35}S]methionine label as described by Wu et al. (34). For protein localization experiments, [^{35}S]methionine-labeled maxicells were fractionated into samples enriched for periplasmic, membrane, or cytoplasmic components according to the protocol reported by Wu et al. (34). All samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis and autoradiography as described previously (10, 22, 24). Marker proteins used to determine M_a were β-galactosidase (116,000), phosphorylase *b* (97,400), catalase (60,000), ovalbumin (45,000), glyceraldehyde-3-phosphate dehydrogenase (36,000), carbonic anhydrase (29,200), α-chymotrypsinogen (25,700), β-lactoglobin (18,400), myoglobin (17,000), lysozyme (14,300), and bovine trypsin inhibitor (6,500).

RESULTS

We sequenced the 2.676-kilobase (kb) *Pst*I F *tra* region DNA fragment carried by plasmid pKI181. This fragment encompasses an F region extending from a *Pst*I site in *traF* through genes *trbA*, *traQ*, and *trbB* to the *traH* amino-terminal region (Fig. 1; 35). A relevant 1,900-nucleotide (nt) portion of the *Pst*I fragment sequence is shown in Fig. 2. The sequence of the first 364 nt of the *Pst*I fragment (not shown) was reported previously, together with the upstream sequence that completes the *traF* gene; numbering in Fig. 2 begins with nt 801 of this published sequence (34). Similarly, Fig. 2 does not include the last 412 nt of the *Pst*I fragment sequence; for the region distal to nt 2602, our data overlap and agree with an independently obtained sequence that is to be reported elsewhere (L. M. Ham, N. Firth, and R. Skurray, Gene, in press).

Figure 2 also shows the translation products of the major open reading frames contained in this segment of F DNA. The genetic identity of each reading frame is discussed below. It is noteworthy that although most of the reading frames occur in the forward direction appropriate to expression from the *tra* operon promoter, one reading frame is in the opposite orientation.

***trbA*.** Our previous work showed that small deletions, introduced at the *Sst*I site at position 971, affected the expression of the TrbA protein (35). The corresponding open reading frame indicates that *trbA* translation initiates 16 nt distal to the *traF* stop codon. The 115-amino-acid, M_r-12,946 *trbA* product is quite hydropathic and includes several long stretches of hydrophobic residues (Met-8 to Tyr-23, Phe-34 to Leu-79, and Ile-95 to Gly-115).

***traQ*.** We have previously demonstrated that pKI158 and other clones carrying the *Sau*3A-*Sma*I fragment that extends from nt 1509 to 1941 are able to express a functional *traQ* product (Fig. 1; 35). Thus, *traQ* was identified as the only major open reading frame contained within this segment (the second forward frame in Fig. 2). This sequence specifies a 94-amino-acid, M_r-10,867 product which contains two major hydrophobic regions (Ile-14 to Leu-32 and Ala-40 to Leu-60),

TABLE 1. Bacterial plasmids

Plasmid	*tra* fragment size (kb) and ends[a]	Vector and vector junction sites[b]	Comment or reference
pKI158	0.432 [*Hin*cII][c] *Sau*3A-*Sma*I	pACYC177 *Hin*cII	35
pKI173	4.1 *Ava*I	pACYC177 *Xma*I	35
pKI181	2.670 *Pst*I	pACYC177 *Pst*I	35
pKI189	2.670 *Pst*I	pACYC177 *Pst*I	*Sst*I-BAL 31 deletion of pKI181 (35)
pKI192	2.0 *Pst*I	pACYC177 *Pst*I	*Hpa*I-BAL 31 deletion of pKI181 (35)
pKI193	0.633 *Hae*III-*Taq*I [*Hin*dIII][c,d]	pACYC177 *Sma*I-*Hin*dIII	Anti-*tra* orientation to *kan* (this paper)
pKI272	3.8 *Eco*RV	pBR322 *Eco*RV	14
pKI274	1.322 *Nsi*I-*Pst*I	pBR322 *Pst*I	This paper
pKI279	1.318 *Nsi*I-*Pst*I	pBR322 *Pst*I	4-bp deletion at *Sma*I (pKI274 + *Xma*I + S1; this paper)
pKI284	1.193 *Hae*II-*Pst*I	pBR322 *Sca*I-*Pst*I	TrbB$^-$ (this paper)
pKI285	1.170 *Sma*I-*Pst*I	pBR322 *Sca*I-*Pst*I	TrbB$^-$ (this paper)
pKI351	0.229 [*Eco*RI][c] *Alu*I-*Sau*3A	pDR540 *Eco*RI-*Bam*HI	Anti-*tra* orientation to *galK* (see Fig. 5)
pKI337	0.432 [*Eco*RI][c] *Sma*I-*Sau*3A [*Hin*cII][c]	pMC1403 *Eco*RI-*Sma*I	Anti-*tra* orientation to *lacZ* (see Fig. 5)

[a] Except as noted otherwise, *tra* DNA inserts in pKI plasmids are oriented such that transcription from the promoter of the vector gene that contains the insert will proceed in the *tra* operon direction (left to right in Fig. 1 and 2).

[b] Unique sites used in pACYC177 are in the *bla* (*Pst*I, *Hin*cII) or *kan* (*Xma*I/*Sma*I, *Hin*dIII) genes; for pBR322, they are in *tcy* (*Eco*RV, *Bam*HI) or *amp* ((*Sca*I, *Pst*I).

[c] Restriction sites in brackets are not in *tra* DNA. The *tra* fragment was first cloned in M13mp8 or M13mp9; sites in brackets were then used to facilitate moving the *tra* fragment to the plasmid vector.

[d] The F DNA in this plasmid extends from nt 1748 (*Hae*III) to nt 1116 (*Taq*I).

followed by a hydrophilic, 29-amino-acid C-terminal segment in which six positively and six negatively charged amino acids occur.

***trbB*.** The third forward open reading frame indicated in Fig. 2 extends from nt 1875 to 2412 and crosses the *Sma*I and *Hpa*I sites expected to be in *trbB*; plasmids pKI272 and pKI181 have been shown to express the M_a-18,400 TrbB polypeptide, but pKI173 and pKI181 derivatives that carry deletions introduced at the *Hpa*I site do not (e.g., pKI192, Fig. 1; 35). Since the *trbB* reading frame appeared to overlap with *traQ*, we tested whether *trbB* might actually initiate at a more distal start site by constructing and examining the products of several additional clones. Plasmid pKI274, which carries the sequences distal to the *Nsi*I site at position 1789, expressed TrbB (Fig. 3). However, TrbB was not expressed by pKI284 (*Hae*II-*Pst*I), pKI285 (*Sma*I-*Pst*I), or pKI279, a derivative of pKI274 containing a 4-base-pair (bp) frameshift deletion introduced at the *Sma*I site by *Xma*I and S1 digestion (data not shown). We concluded that initiation of *trbB* translation occurs within the *traQ* sequence, at nt 1875.

Processing and localization of the *trbB* product. The *trbB* sequence encodes a 179-amino-acid protein of molecular weight 19,508. Since this protein is larger than the M_a-18,400 TrbB polypeptide we had identified and contains a hydrophobic amino-terminal segment, we investigated whether processing of the direct *trbB* product occurred. Figure 3 shows the effect of ethanol on protein labeling in maxicells strains carrying plasmids that express *trbB*. The presence of ethanol reduced the amount of 18.4-kilodalton polypeptide detectable in samples of cells carrying the TrbB$^+$ pKI plasmids 181, 274, and 272. At the same time, labeling of an M_a-21,000 protein expressed by each of these plasmids was enhanced, which suggested that the protein is the precursor to the M_a-18,400 product. In contrast, ethanol did not prevent the appearance of the *trbA* and *traQ* product bands observable in the pKI181 and pKI272 samples, although the intensity of the labeled *trbA* band was somewhat diminished. No potential precursor band could be detected when plasmids expressing only *trbA* (pKI192) or *traQ* (pKI158) were examined.

We also examined the appearance of these proteins in maxicell-labeled samples that were separated into fractions enriched for cytoplasmic, periplasmic, and membrane components. The 18.4-kilodalton TrbB band, although detected in all fractions, appeared to be enriched in the periplasmic protein sample (Fig. 4). In contrast, the *traQ* and *trbA* products were enriched in the membrane preparation. The sequences of the *traQ*, *trbA*, and *trbB* products (Fig. 2) are consistent with these findings. Except in the amino-terminal signal peptide segment, stretches of amino acids with strong hydrophobic character were much less marked in the *trbB* product.

***artA*.** In the *tra* operon direction, there is no significant open reading frame in the 399-bp sequence between *trbA* and *traQ*. Translation of a polypeptide greater than 30 amino acids in length is precluded by 34 stop codons, distributed as 11, 7, and 16 stop codons in the three possible forward reading frames. However, there is a 104-codon open reading frame on the opposite strand (nt 1523 to 1212) that could be translated from RNA transcribed in the anti-*tra* operon direction. Just before the start site of the reading frame there are sequences (TGTAATA and TTTACA) that correspond well with the TATAATG and TTGACA −10- and −35-region promoter consensus sequences (Fig. 2; 27). This suggested that a transcript might indeed be expressed from this strand. Since further experiments (see below) indicated that both transcription and translational initiation of the 104-amino-acid open reading frame does occur in vivo, we spelled *tra* backwards to denote the unusual orientation of this gene and named it *artA*. Its sequence specifies an M_r-12,132 protein that (distal to Leu-22) is highly hydropathic in character.

Since only a single methionine codon (the start codon) is present in the *artA* reading frame, we attempted to detect its product through examination of the ^{14}C-amino acid-labeled polypeptides expressed in maxicells carrying pKI181, pKI189, and pKI193. However, these experiments were inconclusive. Unambiguous resolution of ^{14}C-labeled proteins in the expected size range was difficult; codon preference analysis (7) further indicated that *artA* might be translated at a low level. Therefore, we used fusion protein

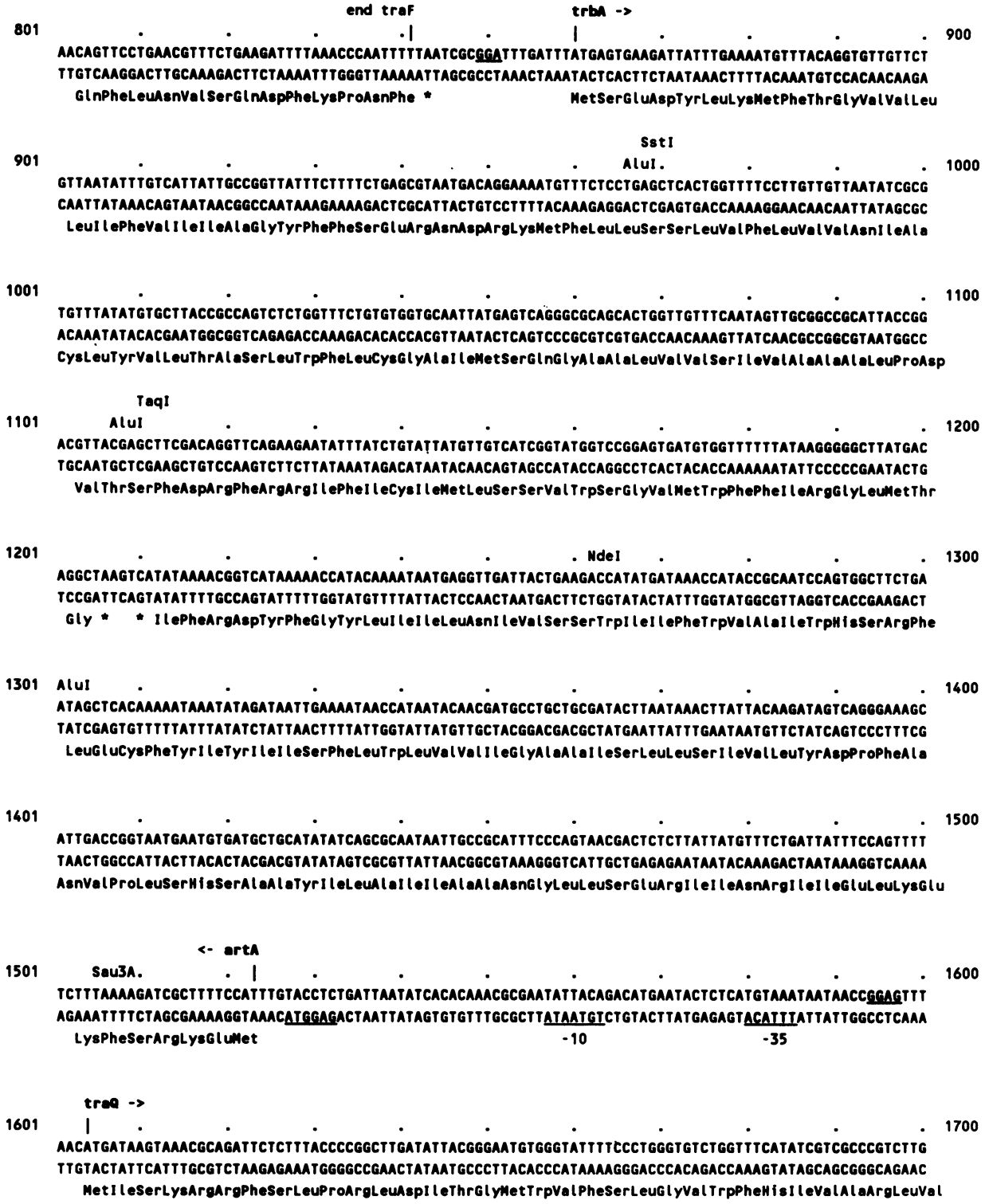
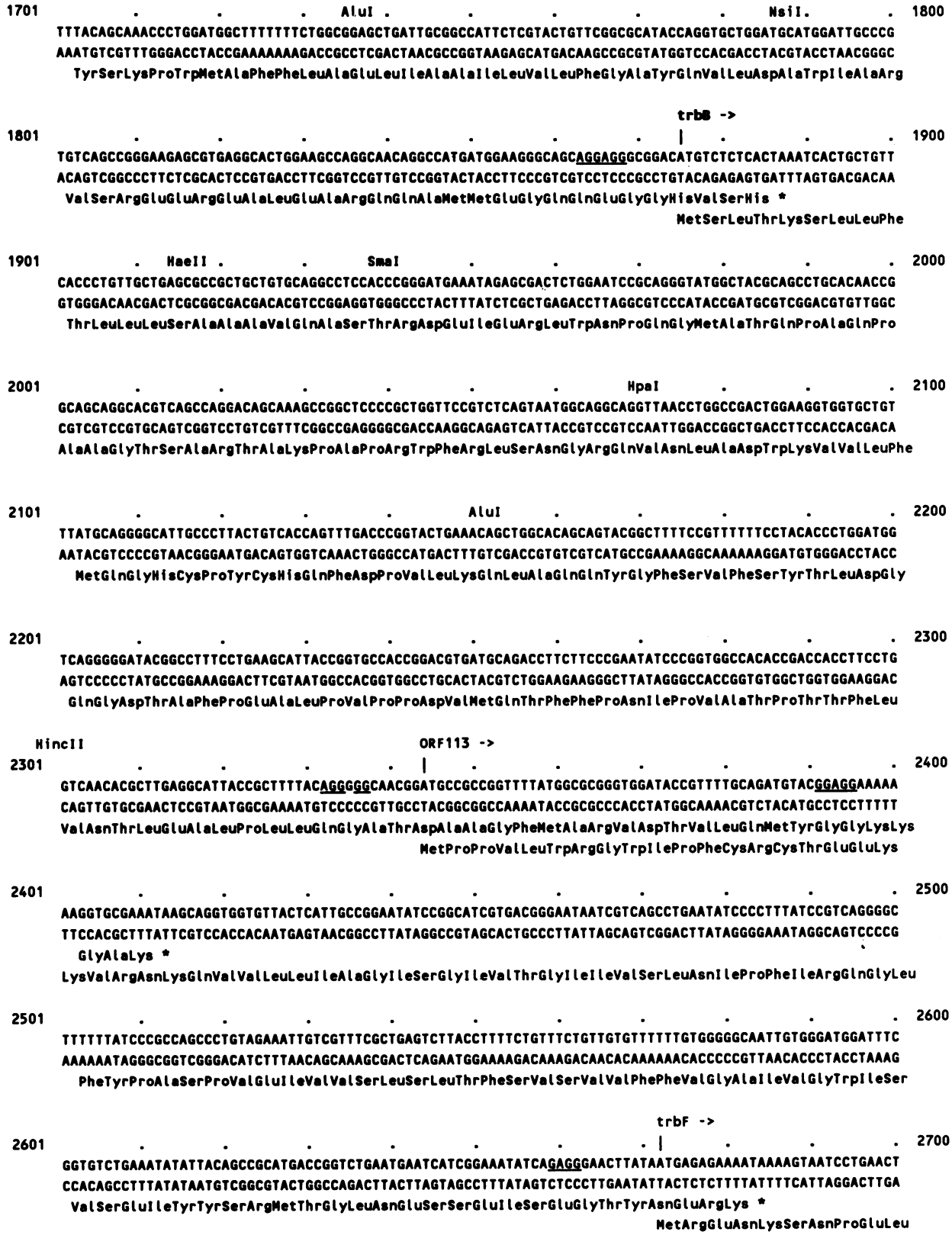


FIG. 2. DNA sequence of the *traQ* region. Translation products of the open reading frames for *trbA*, *artA*, *traQ*, *trbB*, and ORF113 are shown, as are the TraF C terminus (34) and TrbF N terminus (Ham et al., in press). The *artA* reading frame is on the strand complementary to that which encodes the *tra* operon. Numbering continues the sequence reported by Wu et al. (34). Potential ribosome-binding sequences and possible −35 and −10 sequences for the *artA* promoter are underlined; positions of pertinent restriction sites are also indicated.



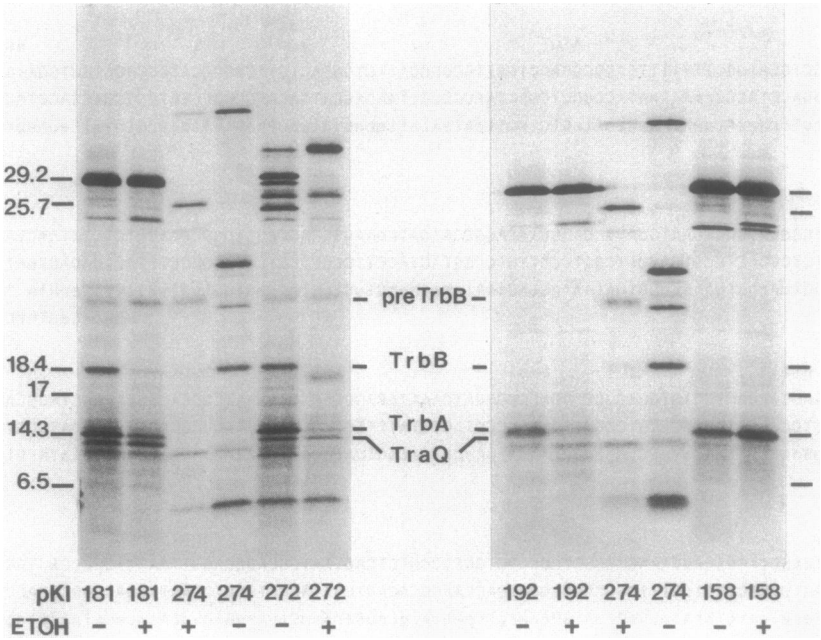


FIG. 3. Effect of ethanol on the ^{35}S-labeled polypeptides synthesized in maxicells carrying various plasmids. Labeling was performed in the presence (+) or absence (−) of ethanol (ETOH); the lower portion of the sodium dodecyl sulfate-polyacrylamide gel autoradiogram is shown. The plasmid present is indicated by the pKI number below each lane. Central lines indicate positions of the products TraQ, TrbA, TrbB, and the TrbB precursor (preTrbB). Lines at the left and right indicate the positions and sizes (in kilodaltons) of marker proteins detected on the stained gel.

vectors to test the expression potential of the promoter and translational start site for *artA*.

Expression of *galK*. To test for *artA* promoter activity, the 229-bp *Alu*I-*Sau*3A fragment (extending from nt 1738 to 1509) was inserted into a *galK* vector. The vector DNA used was a purified *Eco*RI-*Bam*HI pDR540 fragment that does not include the pDR540 *tac* promoter or *lacO* site (Fig. 5). The *Alu*I-*Sau*3A *tra* fragment was cloned from a *Sau*3A digest of a purified 0.43-kb *Eco*RI-(*Alu*I at nt 1738 to *Alu*I at nt 1304)-*Hin*dIII fragment obtained from an M13mp9 clone. This insured an appropriate orientation for the *tra* fragment insert.

Among 130 Amp^r transformants obtained from the ligation mixture, 122 were Gal$^+$ on galactose MacConkey plates. Plasmid DNA from five of these Gal$^+$ colonies was examined; restriction enzyme digests confirmed that each *galK* plasmid carried the F DNA insertion expected. When cultures made from one of these colonies (carrying plasmid pKI351) were assayed for galactokinase, the average level of activity was 438 U. No dramatic change in galactokinase levels was observed when pKI351 expression was assayed in cells carrying F *lac*. We concluded that the sequence tested includes a competent promoter for in vivo transcription of *artA*.

To test for promoter activity in the *tra* operon direction, we also cloned an M13mp9 *Hin*dIII-(*Alu*I at nt 1304 to *Alu*I at nt 1738)-*Eco*RI fragment into *Hin*dIII-*Eco*RI-digested pKO6 DNA. This resulted in *galK* plasmids that carried the *Alu*I fragment in the *tra* orientation. When these pKO6 derivatives were introduced into strain N100, all transformants were Gal$^-$, which suggested that no independent promoter in this region contributes to transcription of *traQ*. Transformation of these strains with the TraJ$^+$ plasmid pRS27 (29) did not convert the strains to Gal$^+$.

Expression of a β-galactosidase fusion protein. We also used the vector pMC1403 (1) to test the capacity of the *artA* promoter and translational start site to express β-galactosidase. In this case, a 0.43-kb purified *Eco*RI-(*Sma*I at nt 1941 to *Sau*3A at nt 1509)-*Hin*cII fragment from an M13mp8 clone was coprecipitated and ligated with the purified pMC1403 *Eco*RI-*Sma*I vector fragment. This construction placed the *artA* promoter upstream of the pMC1403 *lacZ* gene (Fig. 5). Initiation of translation at the *artA* start codon would result in expression of β-galactosidase activity, since the *artA-lacZ* fusion is in frame. After selection of transformants with ampicillin, 13 of 17 colonies were found to be Lac$^+$. The restriction profile of plasmid DNA from all five Lac$^+$ colonies tested confirmed that the 0.43-kb fragment had inserted into pMC1403 in the orientation expected. One of these plasmids was named pKI337.

Assays of cultures of the five Lac$^+$ isolates showed that these strains expressed an average of 880 U of β-galactosidase. As expected, addition of IPTG had no effect on the β-galactosidase expression level. A culture of the negative control strain carrying the promoterless vector, pMC1403, developed no yellow color; β-galactosidase expression (approximately 2,000 U) of an F *lac* strain used as a positive control was dependent on induction with IPTG.

Additional open reading frames. A 113-amino-acid open reading frame (*orf113*; Fig. 2) with the potential to encode a 12,588-dalton protein extends from nt 2344 to 2682. Alternatively, a 93-amino-acid (M_r 10,067) translation product could be derived if translation began at the valine codon (nt 2404) near the end of *trbB*. We presently have no evidence concerning in vivo expression of *orf113*. Another open-reading frame overlaps with the end of *orf113*. The work of others (Ham et al., in press) suggests that this locus (*trbF*) is translated.

DISCUSSION

The segment of F *tra* DNA that we sequenced includes the genes which encode the TrbA, TraQ, and TrbB products that

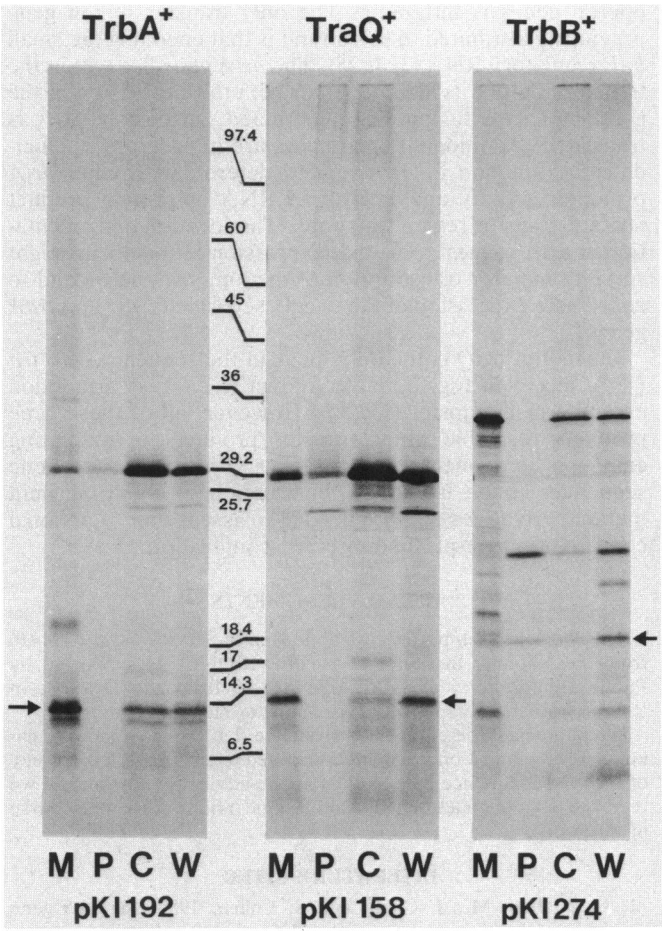


FIG. 4. Subcellular fractions of ^{35}S-labeled proteins expressed in maxicells carrying the TrbA$^+$ plasmid pKI192, the TraQ$^+$ plasmid pKI158, or the TrbB$^+$ plasmid pKI274. Arrows indicate positions of the Trb and Tra products expressed by these plasmids. Whole-cell (W) samples were separated into fractions enriched for membrane (M), periplasmic (P), and cytoplasmic (C) proteins. Numbered lines indicate the sizes and positions of molecular weight marker proteins. The autoradiograms of relevant lanes from two sodium dodecyl sulfate-polyacrylamide gels are shown.

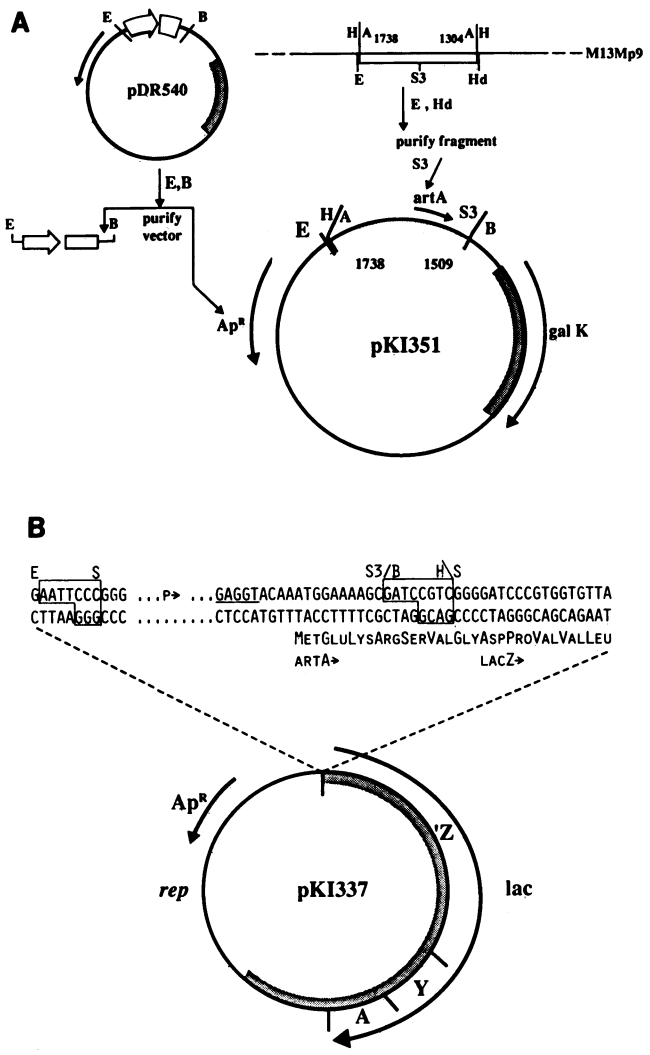


FIG. 5. Structure of plasmids pKI351 (A) and pKI337 (B). (A) Plasmid pDR540 was digested with *Eco*RI (E) and *Bam*HI (B) to remove its *tac* promoter-*lacO* segment, and the large vector fragment (pKO1) was purified. An M13mp9 clone containing an *Alu*I (nt 1738 to 1304) *tra* DNA fragment insert in the phage *Hin*cII site (H) was digested with *Eco*RI (E) and *Hin*dIII (Hd). After purification, the small *Eco*RI-*Hin*dIII fragment was digested with *Sau*3A (S3) and ligated to the vector DNA to give a 0.23-kb *tra* insert in the orientation shown. Expression of *galK* reflects transcription initiated from a promoter (*artA*) oriented in the anti-*tra* operon direction. (B) An M13mp9 clone carrying a 0.43-kb *Sma*I (S)-*Sau*3A (S3) *tra* fragment (nt 1941 to 1509) inserted into the M13mp9 *Sma*I and *Bam*HI sites was used as a source of F DNA. The fragment was excised with *Eco*RI and *Hin*cII, purified, and ligated with *Eco*RI-*Sma*I-digested pMC1403 vector DNA. The figure shows the nucleotide sequence at each end of the insertion; bases within boxed regions are the linker sequences that derive from M13mp9. The insertion creates an in-frame fusion of the amino-terminal segment of *artA* and the vector *lacZ* sequence that can be translated as shown; for expression of *lacZYA*, transcription must be initiated within the F-DNA insert (in the anti-*tra* operon direction), and the translational start of *artA* must be utilized.

we identified previously. It also includes additional open reading frames that could express previously unsuspected *tra* region products. The most interesting of these, *artA*, occurs on the DNA strand opposite that encoding *tra* operon products and is preceded by a promoter sequence. Analysis of the expression of *galK* and *lacZ* fusion vectors indicate that this promoter is active in vivo and that the *artA* sequence would be translated.

The *trbA* gene was originally identified through detection of an M_a-14,200 *tra* product that was affected by small deletions introduced at the unique *Sst*I site in pKI181. The open reading frame that crosses this site encodes a product of M_r 12,946, a value in very reasonable agreement with the product size estimated in our protein experiments; our M_as determined by sodium dodecyl sulfate-polyacrylamide gel electrophoresis appear to slightly overestimate *tra* proteins of this size. In our gel system, TrbA migrated just fractionally slower than did the *traA* product (M_a 14,000; a 121-amino-acid M_r-13,200 protein; 6, 35). That the *trbA* product contains seven methionines and a preponderance of hydrophobic amino acids is also consistent with the appearance of TrbA protein as a prominent [^{35}S]methionine-labeled polypeptide associated with maxicell membrane fractions.

It is interesting that the next gene on the *tra* operon strand, *traQ*, is separated from *trbB* by a 399-base interval that is not

expected to be translated in the *tra* operon direction. For expression of *traQ* to occur from the *tra* operon promoter (or from vector promoters located upstream to *trbA*, e.g., pKI181 and pKI272), transcription must proceed without terminating during synthesis of this untranslated RNA as well as in opposition to any transcription originating at the *artA* promoter. Our *galK* fusion provided no evidence that a promoter located immediately before *traQ* could be responsible for the TraQ expression we have observed.

The *traQ* gene was also originally located through analysis of plasmid clones and deletion mutants. The TraQ activity of pKI158 and other plasmids can be assayed by detection of *traA* product processing; the 7-kilodalton pilin polypeptide is expressed when $TraA^+$ and $TraQ^+$ plasmids are coresident in the cell (35). However, it is still unclear whether TraQ directly catalyzes *traA* product processing or facilitates leader peptide removal in some other manner. The characteristics of the M_r-10,867, 94-amino-acid protein encoded by the *traQ* sequence correlate well with our identification of TraQ as an M_a-12,500, membrane-associated polypeptide. Following an N-terminal cluster of charged residues (Lys-4, Arg-5, and Arg-6), there are hydrophobic regions in the molecule that should allow membrane insertion. Like leader-peptidase, TraQ has been detected in association with both inner and outer membrane fractions (2, 33, 35, 36). However, participation in TraA processing suggests that TraQ should be anchored in the inner membrane. The C-terminal segment of TraQ seems likely to be a region of the molecule that interacts with the pilin precursor; whether this region extends into the periplasm or remains in the cytoplasm after membrane insertion remains to be tested. In the latter case, the TraQ structure would be somewhat analogous to that proposed for the 14-kilodalton N-terminal segment of the signal recognition particle receptor or docking protein (8, 13).

An interesting feature of *trbB* is that its sequence starts within *traQ*. Whereas the codons of previously characterized contiguous *tra* genes do not overlap (5, 6, 11, 30), sequencing studies in progress in this laboratory and elsewhere (S. Maneewannakul, unpublished data; Ham et al., in press) indicate that the start codons for a number of other *tra* operon products are situated prior to (and out of frame with) the stop codon of the upstream gene. It is also of interest that *trbB* expresses a protein with a characteristic membrane signal sequence. The size of this product (M_r 19,508) is consistent with that of the M_a-21,000 precursor we observed to be expressed from $TrbB^+$ clones in the presence of ethanol. Removal of a 22-amino-acid signal peptide from the *trbB* amino terminus would result in an M_r-17,188 protein, comparable in size to the M_a-18,400 TrbB product we have identified. Our cellular fractionation experiments indicate that, unlike TrbA and TraQ, TrbB is periplasmic. This finding is supported by previous analysis of sucrose density-fractionated preparations, which indicated that TrbB is found in fractions of lighter density than those typical of inner membranes (35). Parallel observations were recently reported for TraF, which is also a periplasmic protein (34).

The position of the *artA* promoter we detected appears to correspond to a strong RNA polymerase binding site observed by Manning et al. (16). Interestingly, the amount of galactokinase expressed by the *artA* promoter in plasmid pKI351 appeared to be higher than levels associated with promoters for the *traM*, *traS*, *traT*, and *traI* genes and approximated levels reported for the *traJ* promoter (11, 25). The presence of the *artA* reading frame and promoter sequence on the strand opposite that which encodes *tra* operon genes is intriguing. The only transfer region gene previously attributed to this strand is that encoding the small *finP* regulatory RNA (3, 4, 25). The *artA* gene lies within the transfer operon; its size, together with the expression of the β-galactosidase fusion we constructed, suggests that it is translated. Additional experiments are in progress to elucidate the function of *artA* and to determine whether *artA* promoter activity and/or an *artA* RNA or protein product affect F transfer region functions. The location and orientation of *artA* suggest to us that expression of this locus might serve to monitor transcription of the long *tra* operon and to coordinate expression of the widely separated *traQ* and *traA* genes.

Currently, *traQ* is the only locus in the sequence of F *tra* DNA shown in Fig. 2 that is known to express a function essential to F transfer (23, 35). However, all of these gene products resemble other transfer products in exhibiting envelope component characteristics. We expect the gene sequences to be useful in construction of the F-plasmid mutant derivatives that are needed to assess their functional contribution to F-plasmid-directed conjugation.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant AI14426 from the National Institutes of Health. BIONET is supported by Public Health Service award RR01865-05 from the Division of Research Resources, National Institutes of Health.

We appreciate the effort and assistance that Sumit Maneewannakul and Raymond VonderHaar contributed to ensuring the accuracy of the DNA sequence data as well as the advice and assistance we received from Michael Benedik and Ryland Young during the course of this work.

LITERATURE CITED

1. **Casadaban, M., J. Chou, and S. Cohen.** 1980. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins; *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. J. Bacteriol. **143**:971–980.
2. **Dalbey, R. E., A. Kuhn, and W. Wickner.** 1987. The internal signal sequence of *Escherichia coli* leader peptidase is necessary but not sufficient for its rapid membrane assembly. J. Biol. Chem. **262**:13241–13245.
3. **Dempsey, W. B.** 1987. Transcript analysis of the plasmid R100 *traJ* and *finP* genes. Mol. Gen. Genet. **209**:533–544.
4. **Finlay, B. B., L. S. Frost, W. Paranchych, and N. Willetts.** 1986. Nucleotide sequences of five IncF plasmid *finP* alleles. J. Bacteriol. **167**:754–757.
5. **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**:76–89.
6. **Frost, L., W. Paranchych, and N. Willetts.** 1984. DNA sequence of the F *traALE* that includes the gene for F pilin. J. Bacteriol. **160**:395–401.
7. **Gribskov, M., J. Devereux, and R. R. Burgess.** 1984. The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Res. **12**:539–549.
8. **Hortsch, M., D. Avossa, and D. I. Meyer.** 1985. A structural and functional analysis of the docking protein. Characterization of active domains by proteolysis and specific antibodies. J. Biol. Chem. **260**:9137–9145.
9. **Ippen-Ihler, K. A., and E. G. Minkley, Jr.** 1986. The conjugation system of F, the fertility factor of *Escherichia coli*. Annu. Rev. Genet. **20**:593–624.
10. **Ippen-Ihler, K., D. Moore, S. Laine, D. A. Johnson, and N. Willetts.** 1984. Synthesis of F-pilin polypeptide in the absence of F *traJ* product. Plasmid **11**:116–129.
11. **Jalajakumari, M. B., A. Guidolin, H. J. Buhk, P. A. Manning,**

L. M. Ham, A. L. M. Hodgson, K. C. Cheah, and R. A. Skurray. 1987. Surface exclusion genes *traS* and *traT* of the F sex factor of *Escherichia coli* K-12. Determination of the nucleotide sequence and promoter and terminator activities. J. Mol. Biol. **198**:1–11.

12. **Laine, S., D. Moore, P. Kathir, and K. Ippen-Ihler.** 1985. Genes and gene products involved in the synthesis of F-pili, p. 535–553. *In* D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson, and A. Hollaender (ed.), Plasmids in bacteria. Plenum Publishing Corp., New York.
13. **Lauffer, L., P. D. Garcia, R. N. Harkins, L. Coussens, A. Ullrich, and P. Walter.** 1985. Topology of signal recognition particle receptor in endoplasmic reticulum membrane. Nature (London) **318**:334–338.
14. **Maneewannakul, S., P. Kathir, D. Moore, L.-A. Le, J. H. Wu, and K. Ippen-Ihler.** 1987. Location of F plasmid transfer operon genes *traC* and *traW* and identification of the *traW* product. J. Bacteriol. **169**:5119–5124.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Manning, P. A., G. Morelli, and C. Fisseau.** 1984. RNA-polymerase binding sites within the *tra* region of the F factor of *E. coli* K-12. Gene **27**:121–123.
17. **McKenny, K., H. Shimatake, D. Court, U. Schmeissner, C. Brady, and M. Rosenberg.** 1981. A system to study promoter and terminator signals recognized by *E. coli* RNA polymerase, p. 383–415. *In* J. Chirikjian and T. Papas (ed.), Gene amplification and analysis, vol 2. Structural analysis of nucleic acid. Elsevier/North-Holland Publishing Co., New York.
18. **Messing, J., R. Crea, and P. Seeburg.** 1981. A system of shotgun DNA sequencing. Nucleic Acids Res. **9**:309–321.
19. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double digest restriction fragments. Gene **19**:269–276.
20. **Miller, J. H.** 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **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.
22. **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.
23. **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.
24. **Moore, D., J.H. Wu, P. Kathir, C. M. Hamilton, and K. Ippen-Ihler.** 1987. Analysis of transfer genes and gene products within the *traB-traC* region of the *Escherichia coli* fertility factor, F. J. Bacteriol. **169**:3994–4002.
25. **Mullineaux, P., and N. Willetts.** 1985. Promoters in the transfer region of plasmid F, p. 605–614. *In* D. R. Helinski, S. N. Cohen, D. B. Clewell, D. A. Jackson and A. Hollaender (ed.). Plasmids in Bacteria. Plenum Publishing Corp., New York.
26. **Newman, A. M. J., K. Howe, I. Garner, and R. Hayward.** 1982. Evidence that rifampicin can stimulate readthrough of transcriptional terminators in *E. coli*, including the attenuator of the *rpoBC* operon. Nucleic Acids Res. **10**:7409–7424.
27. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
28. **Russell, D., and G. Bennett.** 1982. Construction and analysis of in vivo activity of *E. coli* promoter hybrids and promoter mutants that alter the −35 to −10 spacing. Gene **20**:231–243.
29. **Skurray, R. A., H. Nagaishi, and A. J. Clark.** 1976. Molecular cloning of DNA from F sex factor of *Escherichia coli* K-12. Proc. Natl. Acad. Sci. USA **73**:64–68.
30. **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.
31. **Willetts, N. S., and R. A. Skurray.** 1980. The conjugative system of F-like plasmids. Annu. Rev. Genet. **14**:41–76.
32. **Willetts, N. S., and R. A. Skurray.** 1987. Structure and function of the F factor and mechanism of conjugation, p. 1110–1133. *In* F. C. Neidhart, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
33. **Wolfe, P. B., W. Wickner, and J. M. Goodman.** 1983. Sequence of the leader peptidase gene of *Escherichia coli* and the orientation of leader peptidase in the bacterial envelope. J. Biol. Chem. **258**:12073–12080.
34. **Wu, J. H., P. Kathir, and K. Ippen-Ihler.** 1988. The product of the F plasmid transfer operon gene, *traF*, is a periplasmic protein. J. Bacteriol. **170**:3633–3639.
35. **Wu, J. H., D. Moore, T. Lee, and K. Ippen-Ihler.** 1987. Analysis of *E. coli* K12 F factor transfer genes: *traQ*, *trbA* and *trbB*. Plasmid **18**:54–69.
36. **Zwizinski, C., T. Date, and W. Wickner.** 1981. Leader peptidase is found in both the inner and outer membranes of *Escherichia coli*. J. Biol. Chem. **256**:3593–3597.