JOURNAL OF BACTERIOLOGY, Apr. 1994, p. 1924–1931
0021-9193/94/$04.00+0

Vol. 176, No. 7

The Nature of the *traK4* Mutation in the F Sex Factor of *Escherichia coli*

S. S. PENFOLD, K. USHER,† AND L. S. FROST*

Department of Microbiology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

Received 18 October 1993/Accepted 25 January 1994

The sequence of *traK* gene of the F sex factor of *Escherichia coli* is presented; the *traK* gene product is predicted to be a protein of 25,627 Da with a signal sequence of 21 amino acids to give a mature protein of 23,307 Da. The *traK4* mutation is an extremely polar mutation in the F plasmid that affects F pilus synthesis and plasmid transfer. *traK* genes carrying the *traK4* mutation and a nonpolar mutation *traK105* were cloned, sequenced, and identified as an amber nonsense and a frameshift mutation, respectively. The *traK4* mutation occurred within one predicted rho-dependent transcription termination element (TTE) and immediately upstream of another, while the *traK105* mutation occurred after the two potential TTEs within the *traK* gene. S1 nuclease protection analysis and Northern (RNA) blot analysis were used to confirm that the *traK4* mutation, but not the *traK105* mutation, caused premature termination of transcription. Computer analysis of the F transfer region suggested the presence of TTE motifs at regular intervals throughout the 33.4-kb sequence.

The F plasmid is a self-transmissible plasmid, approximately 100 kb in size, with the genes required for transfer located within a 33.4-kb region adjacent to the origin of transfer (for a review of conjugation, see references 8, 13, and 36). At least 15 of the 28 genes known to be required for transfer are involved in pilus synthesis and assembly (8, 13). These genes are encoded within the transfer (*tra*) operon, which has been estimated to have a maximum length of 31.5 kb (8, 33). The *tra* operon is regulated by the TraJ transcriptional activator, which positively controls the expression of *tra* genes from the P_Y promoter.

Early experiments suggested that the P_Y transcript extends from *traY* through *traI* (Fig. 1). Recent work has identified a number of promoters in distal *tra* operon genes, and it is presently unclear where the P_Y transcript ends (13). Nevertheless, this transcript is predicted to be long, with the next known promoter within the transfer region occurring before the *trbF* gene, nearly 16 kb downstream from the P_Y promoter (11).

F*lac traK4* and F*lac traK105* (JCFL4 and JCFL105, respectively) are two transfer-deficient mutants of the F plasmid; *tra-4* (later known as F*lac traK4*) was shown to contain an amber-suppressible mutation and was predicted to lie within *traK*, while F*lac traK105* was predicted to contain a frameshift mutation (2, 3, 34). Complementation tests with F *tra* amber mutants indicated that the *traK4* mutation was extremely polar and affected expression of genes in the interval from *traK* through *traG* (12) (Fig. 1). By using a series of Hfr strains carrying deletions extending into the transfer region, the position of this polar mutation was confirmed to be within the *traK* gene (35).

Recently, the salient features of rho-dependent termination in *Escherichia coli* have been characterized (19, 37). A sequence motif common to all rho-dependent terminators characterized thus far, consisting of a region of high cytosine-over-guanosine content, or C>G-rich bubble, prior to the 3′ end of a terminated transcript has been described (5). It was subsequently shown that any DNA fragment bearing this consensus motif could activate rho-mediated release of transcripts in the absence of translation, regardless of its physiological role (21). Richardson (19, 20) has proposed that intragenic rho-dependent terminators act to prevent further transcription of long inessential operons during times of stress. Thus, when a transcript containing a latent rho-dependent transcriptional terminator is no longer being translated, the transcriptional termination signal is recognized and transcription ceases.

In this report, we present the sequence of the F *traK* gene, identify the positions of the polar *traK4* and the nonpolar *traK105* mutations in the F*lac* plasmid JCFLO, and characterize their effect on transcription. Alifano et al. (5) have proposed a consensus sequence for intragenic rho-dependent terminators based on a computer algorithm for DNA sequence analysis. Using a similar program, we predicted the presence of possible transcription termination elements (TTEs) within the transfer operon and used the *traK* mutations to test the validity of these predictions. We have shown that the *traK4* mutation occurs before a predicted termination signal, while the *traK105* mutation is located after both predicted TTEs in the *traK* gene. In addition, we verified the positions of the 3′ ends of the transcripts by using S1 nuclease analysis and found that premature termination hastened the rate of degradation of *traK4* transcripts, possibly by subtle alterations in the secondary structure.

MATERIALS AND METHODS

Bacterial strains, plasmids, and media. Bacterial strains used in transformations were *E. coli* DH5α [*supE44* Δ*lacU169* (ϕ80 *lacZ*Δ*M15*) *hsdR17 recA1 endA1 gyrA95 thi-1 relA1*], MV1193 {Δ(*lac-proAB*) *rpsL thi endA sbcB15 hsdR4* Δ(*srl-recA*)*306*::Tn*10*(Tet^r) F′ [tra*D36* proAB$^+$ *lacI*^{q} *lacZ*Δ*M15*]}, BL21 (*hsdS gal* [λ*cIts857 ind1 S*am 7 *nin5 lacUV5*-T7 gene 1]), and HB101 (*supE44 hsdS20 recA13 ara-14 proA2 lacY1 galK2 rpsL20 xyl-5 mtl-1*). M176, used as the donor strain for mating assays, contains the F plasmid, JCFL0 (2), in *E. coli* JC3272 (F$^-$ *lac* ΔX74 *galK his trp lys str* λ^r T6^r), while *E. coli* ED24 (F$^-$ Lac$^-$ Spc^r) was the recipient. Plasmids F*lac traK4* (JCFL4) and

* Corresponding author. Phone: (403) 492-0458. Fax: (403) 492-2216.

† Present address: Department of Chemistry, University of Oregon, Eugene, OR 97403-1253.

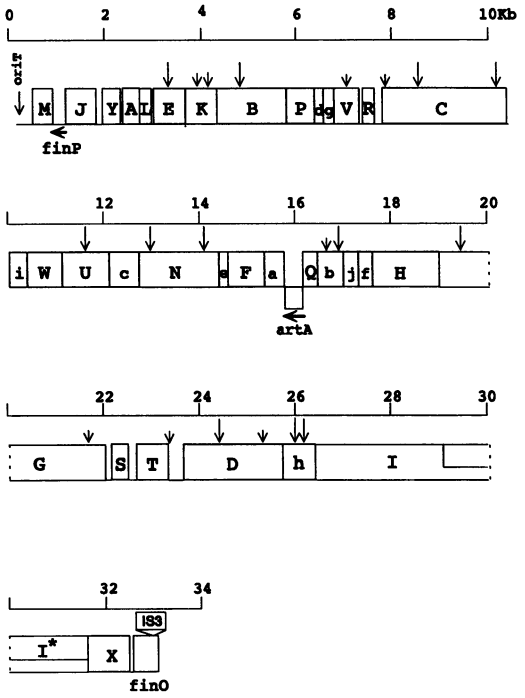

FIG. 1. Genetic map of the transfer region of the conjugative plasmid F, including the positions of C>G bubbles or potential TTEs. Arrows above the open reading frames indicate the location of TTEs; small arrows represent bubbles for which the percentage of C's is less than two times greater than the percentage of G's, and large arrows represent bubbles for which the percentage of C's is more than two times the percentage of G's. The scale represents the number of kilobases from the *Bgl*II site located 141 nt upstream of *oriT*. *tra* genes capitalized; *trb* genes are in lowercase. Two counter transcripts, *finP* and *artA*, are indicated. IS*3* represents the insertion element that interrupts the *finO* gene of the F plasmid.

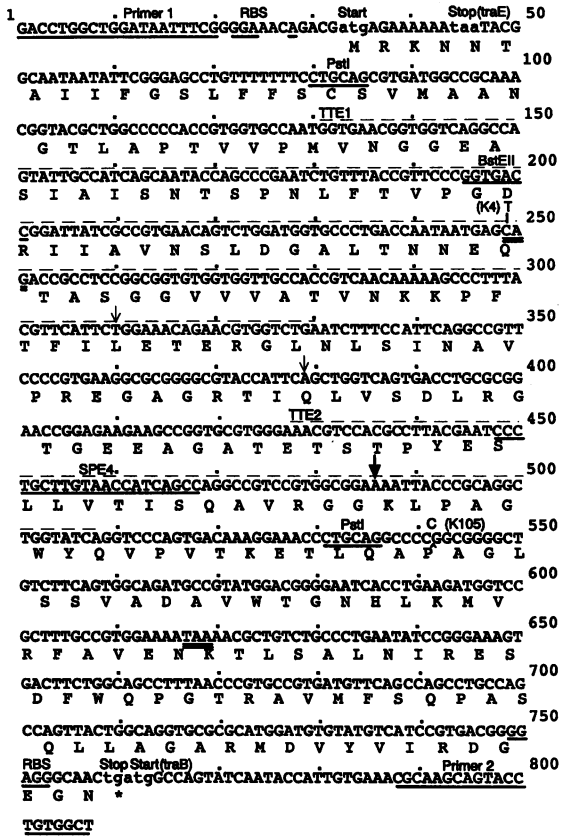

FIG. 2. Nucleotide sequence of the *traK* gene of the F plasmid. The *traK4* and *traK105* mutations are indicated at nt 249 and 541, respectively; the PCR primers 1 and 2 and oligonucleotide SPE4 are underlined. Lowercase letters indicate the overlapping start and stop codons of *traK* and *traE*, respectively, as well as the overlapping stop and start codons of *traK* and *traB*, respectively. RBS indicates ribosome binding sites for *traK* and *traB*. The stop codons found in *traK4* and *traK105* are double underlined. In addition, the two potential TTEs are indicated by overbars. The open arrows denote the 3′ ends of the truncated transcripts found by S1 nuclease analysis of the wild-type and K105 *traK* genes, while the filled arrow indicates the 3′ end of the *traK4* transcript.

F*lac traK105* (JCFL105) have been described elsewhere (2, 34) and were supplied by N. S. Willetts and K. Ippen-Ihler, respectively; pKO4 and pKL200 were obtained from N. Willetts and are related to plasmids developed by McKenney et al. (17). Plasmids pTTQ18 and pTTQ19 (26) were purchased from Amersham, and LB (Luria-Bertani) broth was as previously described (15).

Recombinant DNA techniques and reagents. Restriction and DNA modification enzymes were purchased from Boehringer Mannheim and, unless otherwise stated, were used according to the manufacturer's instructions. The wild-type *traK* gene was subcloned from pRS27 (25), using convenient restriction sites (8), into M13mp18 (38) by standard methods and sequenced by the chain termination method (24), using a Sequenase sequencing kit (United States Biochemical Corp.). Two primers of sequences 5′-GACCTGGCTGGATAATT TCG-3′ and 5′-AGCCACAGGTACTGCTTGCG-3′ (Fig. 2) were used in the PCR to amplify an 850-bp fragment containing the entire *traK* gene from purified F plasmid DNA isolated from M176 and JC3272 containing each of the two mutant F plasmids. The amplified fragments were then treated with Klenow enzyme and T4 polynucleotide kinase and were ligated into *Sma*I-digested pUC118 (30). These constructs were transformed into *E. coli* MV1193, and single-stranded DNA was prepared for sequencing (30). Plasmid pSPK1 contained the wild-type *traK* gene oriented away from the *lac* promoter, while pSPK4 and pSPK5 contained the *traK4* and *traK105* genes, respectively, oriented in the same direction as the *lac* promoter. The *Bam*HI-*Eco*RI fragments from pSPK1, pSPK4, and pSPK5 were cloned into pTTQ18 and pTTQ19 vectors (26) to create plasmids pQSK1, pQSK4, and pQSK5, respectively, such that the *traK* gene was aligned with the *tac* promoter.

Assay for transfer efficiency. Plasmid transfer ability was quantitated by the procedure of Frost et al. (9). Donor cells (*E. coli* JC3272 containing wild-type JCFL0, F*lac traK105*, or F*lac traK4* and either pSPK1, pSPK4, or pSPK5) and recipient cells (*E. coli* ED24) were grown to mid-log phase in LB broth at 37°C, and 0.1 ml of each was mixed together with 1.0 ml of fresh broth. The cells were incubated at 37°C for 1 h, diluted 100-fold in 1× SSC (0.15 M NaCl plus 0.015 M sodium citrate), vortexed to interrupt mating, and plated on selective media after further dilution.

***galK* assay.** The assay to measure *galK* activity was performed as described previously (17) except that the specific activity of the [^{14}C]galactose was 59.6 mCi/mmol and the background strain was *E. coli* HB101. The units were expressed as nanomoles of galactose phosphorylated per minute per unit of optical density of cells at 650 nm.

Computer-predicted termination sites. By using the parameters suggested by Alifano et al. (5), a computer program for the IBM PC was written in Turbo Pascal to assess the presence of possible termination sites within the entire F transfer region. TTEs were defined as regions of at least 78 nucleotides (nt) where the percentage of cytosine residues was greater than the percentage of guanosine residues. Details of the program and the program itself are available upon request.

RNA preparation. Cells (1.5 ml) containing pQSK1, pQSK4, or pQSK5 were induced for 3 min by the addition of IPTG (isopropylthiogalactoside) to a final concentration of 0.5 mM followed by the addition of rifampin to a final concentration of 200 μg ml^{-1} for 2 min. RNA was prepared by the modified hot phenol extraction method described previously (7).

S1 nuclease analysis. Nuclease protection experiments were carried out by using the 337-bp *Bst*EII-*Pst*I (Fig. 2) fragment as a probe. The fragment was labeled with [α-^{32}P]dATP, using Klenow enzyme to fill in the recessed 3′ end of the *Bst*EII site (15), and purified by passage through a Bio-Gel P-30 column. Hybridization of 0.1 pmol of probe to 50 μg of total RNA was allowed to proceed overnight at 50°C prior to treatment with 400 U of S1 nuclease. Maxam and Gilbert sequencing reactions (16) were carried out simultaneously on the labeled *Bst*EII-*Pst*I fragment, and the reactions were electrophoresed on a 6% denaturing polyacrylamide gel.

Northern (RNA) blot analysis. Oligonucleotide SPE4 (5′-GGCTGATGGTTACAAGCAGGG-3′) is complementary to the 3′ end of TTE2 (Fig. 2) and was end labeled with [γ-^{32}P]ATP, using T4 polynucleotide kinase (23). RNA extracted from *E. coli* containing pQSK1, pQSK4, and pQSK5 was subsequently probed with the labeled primer. RNA was separated by electrophoresis on a 5% polyacrylamide gel containing 8.5 M urea and transferred to a Hybond N nylon membrane (Amersham), using a Biorad Trans-Blot Cell. The membrane was prehybridized for a minimum of 2 h at 37°C in 2.5× SSC–5× Denhardt's solution (23)–1.5% sodium dodecyl sulfate (SDS)–100 μg of *E. coli* W tRNA type XX (Sigma) per ml–100 μg of denatured calf thymus DNA per ml. After hybridization at 56°C overnight, the membrane was washed twice for 10 mm each time with 6× SSC–0.1% SDS at 56°C and twice for 10 min each time at 61°C. Autoradiography was performed at −70°C with an intensifying screen, using Kodak X-AR5 film.

Secondary structure analysis. Sequence analysis and secondary structure predictions were generated by RNAFOLD (39), using the PC/Gene program (Intelligenetics, Mountain View, Calif.) and the LoopViewer RNA secondary structure viewing program (10).

GenBank accession number. The nucleotide sequence of *traK* can be accessed through GenBank accession number U01159.

RESULTS

Sequence analysis of the *traK* gene. As part of a project to complete the sequence of the entire F transfer region (8), the *traK* gene was cloned into M13mp18, using *Pst*I and *Sma*I restriction sites within the *traE*, *traK*, *traB*, and *traP* genes (8, 9). Two *Pst*I fragments of 461 and 2,408 bp extended the sequence presented earlier (9) to within the *traP* gene downstream from *traK* and *traB*. A single *Sma*I fragment of 1,498 bp contained the 3′ portion of the *traE* gene and extended into the *traB* gene following *traK*. These fragments were sequenced in both directions, using oligonucleotide primers to extend the sequence, and the portion representing the *traK* gene is shown in Fig. 2. *traK* is located 1.7 kb downstream from the P_Y promoter and overlaps with the genes upstream (*traE*) and downstream (*traB*) from it. The *traE* gene overlaps the *traK* sequence by three codons, while the stop codon for *traK* is part of the initiation codon for *traB*. This pattern of translational coupling is a common feature of the F transfer operon (8).

TABLE 1. Comparison of *galK* activities observed for wild-type, K4, and K105 *traK* genes assayed with a termination assessment vector

Construct	*GalK* activity[a]	% *galK* activity[b] compared with wild type
pKL200 (vector)	14	25
Wild type	57.0	100
K4	5	9
K105	9	16

[a] Expressed as nanomoles of galactose phosphorylated per minute per unit of optical density of cells at 650 nm.
[b] Average for two assays.

The *traK* gene product is predicted to be a protein of 25,627 Da, as determined by the PEPTIDESORT program of PC/Gene. With use of the PROSITE and SURFACEPLOT programs, TraK is predicted to be a periplasmic protein of 23,307 Da after cleavage of a signal sequence of 21 amino acids.

Identification of the *traK4* and *traK105* mutations. The phenotypes of the F plasmids F*lac traK4* and F*lac traK105* were confirmed by performing mating efficiency assays and comparing the results with those for the wild-type F plasmid, JCFL0. The *traK105* mutation could be complemented by wild-type *traK* supplied in *trans* from pSPK1, but the *traK4* mutation could not, which agrees with data presented previously (2, 34) (data not shown). By using PCR and primers described in Materials and Methods, the *traK* genes from JCFL0, F*lac traK4*, and F*lac traK105* were amplified and cloned into pUC118 to give pSPK1, pSPK4, and pSPK5. pSPK4 was found to contain a C-T transition at nucleotide 249 resulting in an amber stop codon at amino acid 73 (Fig. 2), while the *traK105* mutation resulted from the addition of an extra CG base pair after nt 541, giving a frameshift mutation and termination of translation after codon 195. Several isolates from each amplification and cloning reaction were sequenced in order to guard against mutations introduced by *Taq* polymerase during PCR.

Measurement of termination activity within the *traK* gene. The highly polar nature of the *traK4* mutation suggested the presence of a rho-dependent terminator which would affect expression of downstream genes in the *tra* operon. Termination activity was measured by using the vector pKL200 (17), which contains a multiple cloning site flanked by the *lac* promoter upstream and the *galK* gene downstream. Insertion of a fragment containing a transcription terminator results in a decrease in *galK* activity proportional to the efficiency of the terminator. The *traK* genes from pSPK1, pSPK4, and pSPK5 were inserted into pKL200 by using convenient restriction sites, and the level of galactokinase activity was measured in an *E. coli* HB101 background. The results of the *galK* assays are shown in Table 1. Measurements of *galK* activity in these constructs showed greatly reduced *galK* activity both for *traK4* and *traK105* compared with the wild-type gene. The low *galK* activity measured for *traK105* was unexpected, as this mutation has not been demonstrated to be polar in the F plasmid. It was therefore not predicted to cause termination of transcription. Interestingly, pKL200 containing the wild-type *traK* gene showed strong *galK* activity, approximately fourfold greater than that for the vector alone. To test for the presence of a

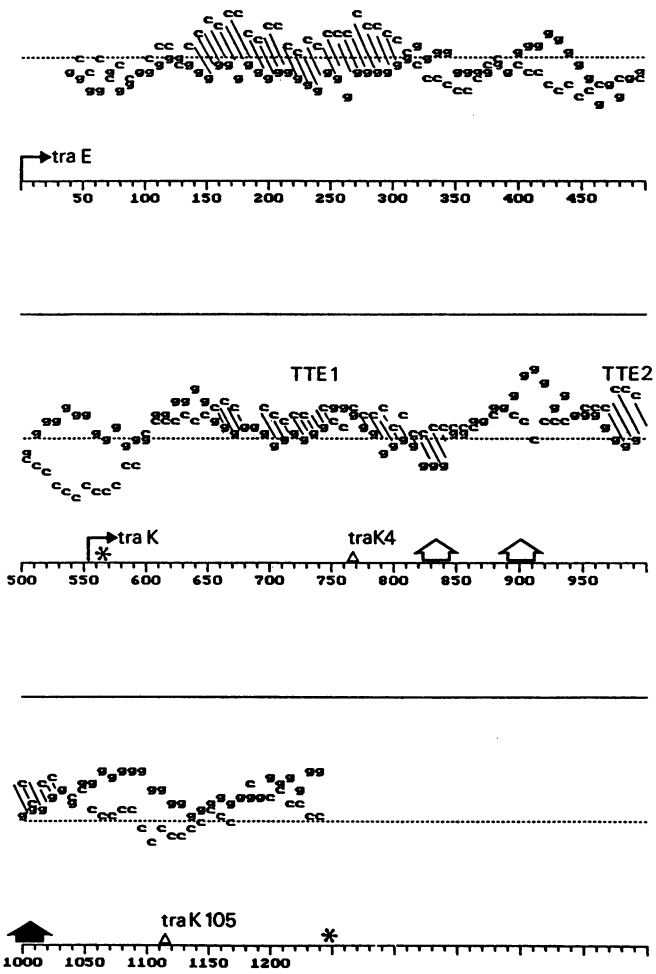

FIG. 3. Computer analysis of the *traEK* genes of the F plasmid for potential TTEs. The dotted line represents 25% cytosine or guanosine content for each 78-nt block searched. The *traK4* and *traK105* mutations (△) as well as the positions of the 3′ ends of protected fragments found by S1 analysis of wild-type and K105 *traK* (open arrows) and *traK4* (filled arrows) are indicated. Asterisks identify the translation stop points for the *traE* and *traK* genes. TTEs are shaded.

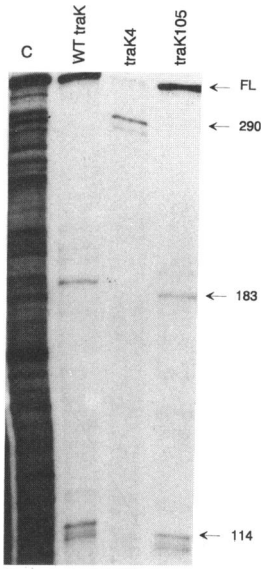

FIG. 4. S1 nuclease analysis of the wild-type (WT) *traK*, *traK4*, and *traK105* transcripts, using the 337-bp *Bst*EII-*Pst*I fragment (nt 196 to 533 in Fig. 2) as a probe. Sizes of protected fragments is shown in base pairs. C indicates the sequence of C residues obtained from Maxam and Gilbert sequencing of the same fragment. FL indicates the position corresponding to the full-length fragment. The autoradiogram is overexposed in order to visualize the protected fragments.

promoter near the end of *traK*, we inserted the gene in the promoter assessment vector, pKO4 (17), and measured the resulting *galK* activity. No difference in *galK* activity could be detected between this construct and the vector alone, suggesting the absence of a promoter in *traK*.

Computer-predicted TTEs in *traK*. Using the features of the computer program described by Alifano et al. (5), we searched the entire transfer region for possible C>G bubbles, also known as TTEs. TTEs were defined as regions of at least 78 nt which contained a higher percentage of cytosine residues than guanosine residues. It is noteworthy that it is the ratio of C residues to G residues, rather than the absolute number of C's, that is important. Sequential blocks of 78 nt were searched at 1-bp intervals throughout the transfer region to identify sequences fitting these criteria. The search for TTEs in *traK* identified two such regions at nt 131 to 330 and 429 to 510, the second of which lay between the *traK4* and *traK105* mutations (Fig. 2 and 3). In addition, a number of these motifs occurred throughout the *tra* operon and are depicted in Fig. 1. Strong TTE motifs (large arrows) were characterized by a percentage of C residues at least twice that of G residues, while weaker TTEs (small arrows) contained a greater percentage of C's than G's, but the percentage of C's was less than two times greater than the percentage of G's. The two TTEs in the *traK* gene were considered to be weak signals by these criteria.

Analysis of the effectiveness of the TTEs in *traK*. The second TTE in *traK*, positioned between the *traK4* and *traK105* mutations, would be expected to truncate mRNA that is not actively translated. If this TTE is recognized under these conditions, a truncated transcript would be predicted for *traK4* but not *traK105*, since translation continues downstream of the TTE in the latter mutant. Because of the extreme length of the *tra* operon transcript in the F plasmid itself, the fate of the RNA in the wild-type and mutant plasmids could not be assessed easily. Therefore, the *traK* genes from the pSPK constructs were recloned into pTTQ118 and pTTQ119 such that they were under the control of an IPTG-inducible *tac* promoter in the vector, generating pQSK1, pQSK4, and pQSK5. S1 nuclease protection experiments were performed as described in Materials and Methods in order to define the 3′ ends of the mRNA resulting from induction of the *tac* promoter in each construct. A protected fragment which terminated 290 bp from the *Bst*EII site was identified for pQSK4 (Fig. 4); a small amount of full-length probe corresponding to the 337-bp *Bst*EII-*Pst*I fragment was detected upon overexposure of the gel. Since equal numbers of counts were loaded in all lanes of Fig. 4, these results indicate that most of the *traK4* transcripts terminated near the 3′ end of TTE2. The protected fragments detected for mRNAs induced from pQSK1 and pQSK5 were identical; the strongest band represented the full-length probe, while weaker bands were found to terminate at 114 and 183 bp from the *Bst*EII site (Fig. 4). Assuming that the transcripts began within the vector at the *tac* promoter, the ends of shorter transcripts detected in pQSK1 and pQSK5

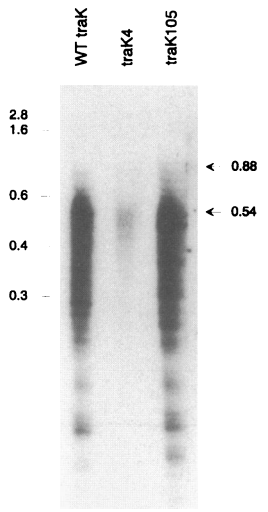

FIG. 5. Northern blot analysis of the wild-type (WT) *traK*, *traK4*, and *traK105* transcripts, using oligonucleotide SPE4 as a probe. Each sample contained 10 μg of RNA; size markers are in kilobases.

samples lie between TTE1 and TTE2, while the 3′ end of the truncated product in pQSK4 is located near the end of TTE2, suggesting efficient termination at this site.

The mRNA induced from the pQSK clones was subjected to Northern blot analysis in order to estimate its size and abundance (Fig. 5). The pattern of degradation and/or processing of the mRNA in pQSK5 contained only minor differences from the pattern in pQSK1. The largest transcripts detected were approximately 880 nt long, corresponding to the predicted length for full-length RNA. A very weak signal corresponding to a transcript approximately 540 nt long was detected for pQSK4 upon overexposure of the autoradiograph. The size of this fragment is within the expected size range for a transcript expressed from the *tac* promoter and terminating at nt 486 (Fig. 2) and thus was thought to correspond to the protected fragment detected by S1 analysis.

The overall increase in the rate of degradation of *traK4* RNA may be the result of changes in secondary structure. To investigate whether either the *traK4* or *traK105* mutation affected the secondary structure of the RNA, we made used of the computer algorithm RNAFOLD (39) to predict the secondary structures of these transcripts. Nucleotides 122 to 490 of the wild-type *traK* (and *traK105*, since the sequence is identical in this region) sequence and nt 122 to 486 of the *traK4* sequence were folded into their most stable predicted configurations (Fig. 6). The C-to-T transition at nt 249 in *traK4* was found to alter the predicted structure of the RNA between nt 241 and 381. Note that despite their different structures, the molecules have similar calculated free energies of approximately −193 kcal (ca. −807.5 kJ). The one-base-pair change that defines the *traK4* mutation may affect the pattern of degradation of the transcript by affecting the overall structure of the RNA. In contrast, the secondary structure predictions for nt 402 to 760 of the wild-type *traK* and *traK105* transcripts, which encompass the *traK105* mutation, were identical.

DISCUSSION

Unlike rho-independent terminators, which usually occur between genes, rho-dependent terminators have been identified both between genes, as in λtR_1 (31) and *trpt′* (18), and within an operon, as in the *his* operon of *Salmonella typhimurium* (4, 6) and the *ilvGMEDA* operon (32) and *lacZ* gene (22) of *E. coli*. In the latter two cases, the polarity of certain mutations has been attributed to rho-dependent termination under conditions in which translation has stopped. It has been proposed that the recognition of latent intragenic terminators under such conditions is a general mechanism to reduce unnecessary transcription (19, 20). Because of the large size of the *tra* operon, regulatory elements such as rho-dependent intragenic terminators would be expected to be present within this operon. The polar nature of the *traK4* mutation suggested that a rho-dependent terminator may have been unmasked and that this was the cause of the poor expression of genes downstream from *traK*. While *traB* is predicted to a have strong ribosome binding site, expression of this gene is greatly affected by the *traK4* mutation, suggesting that the primary effect of the mutation is at the transcriptional rather than the translational level.

With use of the PROSITE and SURFACEPLOT programs, TraK is predicted to have a molecular mass of 23,307 Da after processing from the original product of 25,627 Da. In previous analyses of *tra* products, only one polypeptide of approximately 24,000 Da has been identified, and it was weakly expressed in all cases (1, 14, 29); no precursor polypeptide was detected. Resolution of *tra* proteins in this size range is difficult because of the large number of *tra* proteins close to the 25,000-Da range; this may explain the inability to detect a precursor product. Attempts to overexpress the TraK protein from the *traK*-containing fragment of pQSK1 placed under the control of an inducible T7 RNA polymerase promoter were unsuccessful. This suggests that TraK expression may be influenced by the presence of one or more *tra* genes within the larger fragments used previously (1, 14, 29) to detect the *traK* gene product.

In galactokinase assays, reduced *galK* activity could be demonstrated with both the *traK4* and *traK105* genes. While reduced *galK* activity was expected for the *traK4* gene, the activity for *traK105* was unexpectedly low. This reduced activity appears to be the result of reduced translation rather than transcriptional termination, since S1 nuclease analysis and Northern blot analysis demonstrated the presence of full-length *traK* transcripts in this construct. The *traK105* mutation could affect ribosome loading of the *galK* gene by allowing an alternative secondary structure to form in the absence of translation at the 3′ end of this transcript. Also, the origin of the increased *galK* activity found for the wild-type *traK* gene is unknown, since no promoter activity could be detected by using the promoter assessment vector pKO4 (data not shown) and no fortuitous promoter was found at the cloning junction in pKL constructs.

The current model for the mechanism of rho-dependent termination suggests that these TTEs function by coupling transcription to translation to ensure that transcription does not continue when translation has been aborted (20). In the absence of translation, specific sequences, called *rut* sequences, which are recognized by Rho protein, are exposed in the nascent RNA. Rho binds to the *rut* sequences and causes termination at a point just downstream of *rut* by a mechanism which is still unclear. In 1991, Alifano et al. (5) described a motif which was common to all rho-dependent terminators and is proposed to constitute a *rut* site: a region of high cytosine-over-guanosine content called a C>G-rich bubble.

A search for TTE motifs in the *traK* gene revealed two of these TTEs, one of which lies between the *traK4* and *traK105* mutations. The location of this TTE in *traK* supports the theory that the polar nature of the *traK4* mutation is due to rho-

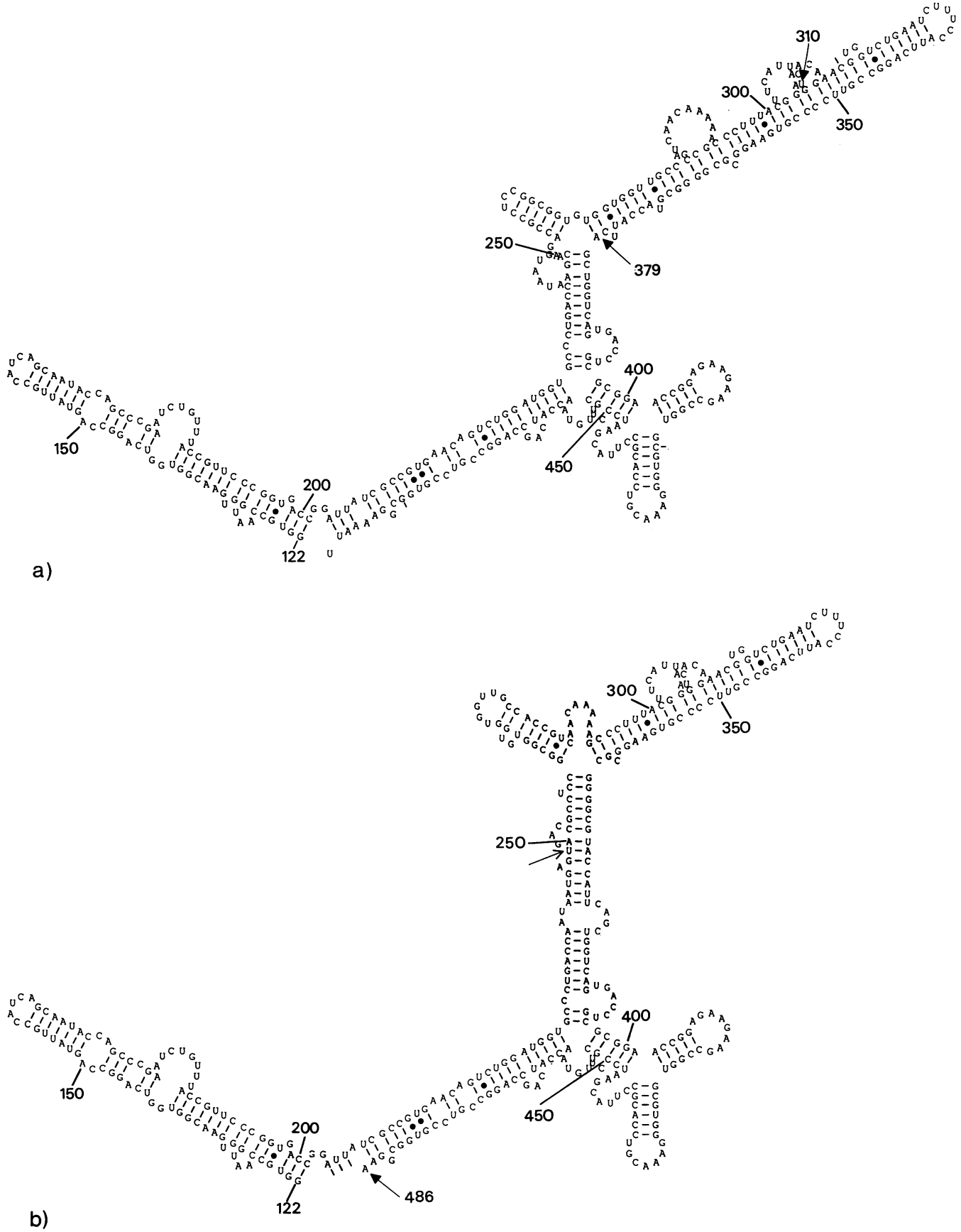

FIG. 6. Predicted secondary structures of wild-type *traK* and *traK105* (a) and *traK4* (b) transcripts from nt 122 to 490 and 122 to 486, respectively, from Fig. 2. The 3′ ends of the protected fragments detected by S1 nuclease analysis are indicated by filled arrows (nt 310 and 370 in panel a and nt 486 in panel b), and the site of the *traK4* mutation is indicated by the open arrow. The secondary structure was generated by RNAFOLD (39), using the PC/Gene program (Intelligenetics), and visualized by LoopViewer (10).

dependent termination. Abortion of translation at codon 73 in *traK4* would result in recognition by Rho of TTE2, and termination of transcription would occur. In *traK105*, none of the TTE2 were predicted downstream of the translation stop codon, suggesting that transcription would proceed into the next gene.

Nuclease protection experiments identified truncated transcripts in *traK4* which terminated at the 3′ end of TTE2. In contrast, the majority of transcripts identified in wild-type *traK* and *traK105* samples protected the probe from digestion by S1, suggesting that these transcripts were not terminated prematurely. The truncated transcripts detected for *traK4* in Northern blot analysis appeared unstable and were very rapidly degraded, possibly because of an altered secondary structure of the prematurely terminated transcripts which triggered ribonucleolytic degradation. In comparison, transcripts isolated from cells expressing cloned wild-type *traK* and *traK105* were considerably more stable, and fragments corresponding to full-length transcripts were detected. It is possible that in the F plasmid itself, alternate secondary structures and degradation patterns, differing from those described here, may occur when these genes are part of the large *tra* operon.

The RNA secondary structure predictions were identical for the wild-type *traK* and *traK105* mRNAs, but an alternate structure was predicted for *traK4* mRNA. It is interesting to note that a single base change in the *traK4* mutant significantly altered the predicted secondary structure. This altered structure could result in rapid degradation of the *traK4* transcripts by an alternate pathway to that used for wild-type or *traK105* transcripts. The longer of the two truncated transcripts observed in wild-type *traK* and the *traK105* mutant (nt 379 in Fig. 2) ends at the 3′ side of a predicted stem-loop structure, suggesting that this may be a polymerase pause or RNase processing site. In *traK4*, this stem-loop no longer occurs, explaining the absence of an RNA transcript of corresponding size. These results suggest that subtle changes in sequence (1 bp) can greatly affect RNA stability either by altering target sequences for ribonucleolytic cleavage or by decreasing chemical stability in the absence of translation.

The absence of any full-length transcript from the *traK4* gene supports the theory that the polarity exhibited by this mutant is due to the recognition of a rho-dependent terminator in *traK*, and computer analysis to detect possible terminators suggests that TTE2 could fulfill this role. Since the consensus sequence for *rut* sites can be characterized as having low stringency, this mechanism of termination could be widely used in a variety of genes. That a number of TTEs were found throughout the *tra* operon and were especially prominent in the longer genes such as *traE*, *-B*, *-C*, *-U*, *-N*, *-G*, and *-D* and *trbH* suggests that they play a role in terminating unnecessary transcription in this long operon. These results provide another example of how intragenic elements may function to control gene expression.

ACKNOWLEDGMENTS

We thank Tim van Biesen for assistance with the RNA secondary structure predictions.

This work was supported by the Medical Research Council of Canada.

REFERENCES

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. USA **76**:4837–4841.
2. **Achtman, M., N. S. Willetts, and A. J. Clark.** 1971. Beginning a genetic analysis of transfer determined by the F factor in *Escherichia coli* by isolation and characterization of transfer-deficient mutants. J. Bacteriol. **106**:529–538.
3. **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.
4. **Alifano, P., M. S. Ciampi, A. G. Nappo, C. B. Bruni, and M. S. Carlomagno.** 1988. In vivo analysis of the mechanisms responsible for strong transcriptional polarity in a "sense" mutant within an intercistronic region. Cell **55**:351–360.
5. **Alifano, P., F. Rivellini, D. Limauro, C. B. Bruni, and M. S. Carlomagno.** 1991. A consensus motif common to all Rho-dependent prokaryotic transcription terminators. Cell **64**:553–563.
6. **Ciampi, M. S., P. Alifano, A. G. Nappo, C. B. Bruni, and M. S. Carlomagno.** 1989. Features of the Rho-dependent transcription terminator polar element within the *hisG* cistron of *Salmonella typhimurium*. J. Bacteriol. **171**:4472–4478.
7. **Frost, L., S. Lee, N. Yanchar, and W. Paranchych.** 1989. *finP* and *fisO* mutations in FinP antisense-RNA suggest a model for FinOP action in the repression of bacterial conjugation by the F*lac* plasmid JCFLO. Mol. Gen. Genet. **218**:152–160.
8. **Frost, L. S., K. Ippen-Ihler, and R. A. Skurray.** An analysis of the sequence and gene products of the transfer region of the F sex factor. Submitted for publication.
9. **Frost, L. S., W. Paranchych, and N. S. Willetts.** 1984. DNA Sequence of the F *traALE* region that includes the gene for F pilin. J. Bacteriol. **160**:395–401.
10. **Gilbert, D. G.** 1990. *Loopviewer*, a Macintosh program for visualizing RNA secondary structure. Published electronically on the Internet, available via anonymous ftp to iubio.bio.indiana.edu.
11. **Ham, L. M., D. Cram, and R. Skurray.** 1989. Transcriptional analysis of the F plasmid surface exclusion region: mapping of *traS*, *traT*, and *traD* transcripts. Plasmid **21**:1–8.
12. **Ippen-Ihler, K., M. Achtman, and N. Willetts.** 1972. Deletion map of the *Escherichia coli* K12 sex factor F: the order of eleven transfer cistrons. J. Bacteriol. **110**:857–863.
13. **Ippen-Ihler, K., and S. Maneewannakul.** 1991. Conjugation among enteric bacteria: mating systems dependent on expression of pili, p. 35–69. *In* M. Dworkin (ed.), Microbial cell-cell interactions. American Society for Microbiology, Washington, D.C.
14. **Ippen-Ihler, K., D. Moore, S. Laine, D. A. Johnson, and N. S. Willetts.** 1984. Synthesis of F-pilin polypeptide in the absence of F traJ product. Plasmid **11**:116–129.
15. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, p. 113–114. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–580.
17. **McKenney, K., H. Shimatake, D. Court, U. Schmeissner, C. Brady, and M. Rosenberg.** 1981. A system to study promoter and terminator signals recognized by *Escherichia coli* RNA polymerase, p. 383–415. *In* J. Chirikjian, and T. Papas (ed.), Gene amplification and analysis, vol. II. Elsevier-North Holland, New York.
18. **Platt, T.** 1986. Transcription termination and the regulation of gene expression. Annu. Rev. Biochem. **55**:339–372.
19. **Richardson, J. P.** 1990. Rho-dependent transcription termination. Biochim. Biophys. Acta **1048**:127–138.
20. **Richardson, J. P.** 1991. Preventing the synthesis of unused transcripts by Rho factor. Cell **64**:1047–1049.
21. **Rivellini, F., P. Alifano, C. Piscitelli, V. Blasi, C. Bruni, and M. S. Carlomagno.** 1991. A cytosine- over guanosine-rich sequence in RNA activates rho-dependent transcription termination. Mol. Microbiol. **5**:3049–3054.
22. **Ruteshouser, E. C., and J. P. Richardson.** 1989. Identification and characterization of transcription termination sites in the *Escherichia coli lacZ* gene. J. Mol. Biol. **208**:23–43.
23. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 5.68–5.69. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

25. **Skurray, R. A., H. Nagaishi, and A. J. Clark.** 1978. Construction and *Bam*HI analysis of chimeric plasmids containing *Eco*RI DNA fragments of the F sex factor. Plasmid **1**:174–186.
26. **Stark, M. J. R.** 1987. Multicopy expression vectors carrying the *lac* repressor gene for regulated high level expression of genes in *Escherichia coli*. Gene **51**:255–267.
27. **Studier, F. W., and B. A. Moffat.** 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. J. Mol. Biol. **189**:113–130.
28. **Tabor, S., and C. C. Richardson.** 1985. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. Proc. Natl. Acad. Sci. USA **82**:1074–1078.
29. **Thompson, R., and M. Achtman.** 1979. The control region of the F sex factor DNA transfer cistrons: physical mapping by deletion analysis. Mol. Gen. Genet. **169**:49–57.
30. **Vieira, J., and J. Messing.** 1987. Production of single-stranded plasmid DNA. Methods Enzymol. **153**:3–11.
31. **von Hippel, P. H., D. G. Bear, W. D. Morgan, and J. A. McSwiggen.** 1984. Protein-nucleic acid interactions in transcription: a molecular analysis. Annu. Rev. Biochem. **53**:389–446.
32. **Wek, R. C., J. H. Sameshima, and G. W. Hatfield.** 1987. Rho-dependent transcriptional polarity in the *ilvGMEDA* operon of wild-type *Escherichia coli* K12. J. Biol. Chem. **262**:15256–15261.
33. **Willetts, N.** 1977. The transcriptional control of fertility in F-like plasmids. J. Mol. Biol. **112**:141–148.
34. **Willetts, N., and M. Achtman.** 1972. Genetic analysis of transfer by the *Escherichia coli* sex factor F, using P1 transductional complementation. J. Bacteriol. **110**:843–851.
35. **Willetts, N., J. Maule, and S. McIntire.** 1976. The genetic locations of *traO*, *finP* and *tra-4* on the *E. coli* sex factor F. Genet. Res. (Cambridge) **26**:255–263.
36. **Willetts, N., and R. 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. Schaecter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
37. **Yager, T. D., and P. H. von Hippel.** 1991. A thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. Biochemistry **30**:1097–1118.
38. **Yanisch-Perron, C., J. Vieira, and J. Messing.** 1985. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene **33**:103–109.
39. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxilliary information. Nucleic Acids Res. **9**:133–148.