

## Nucleotide Sequence of the *traI* (Helicase I) Gene from the Sex Factor F

HARVEY D. BRADSHAW, JR.,<sup>1</sup> BETH A. TRAXLER,<sup>2†</sup> EDWIN G. MINKLEY, JR.,<sup>3\*</sup> EUGENE W. NESTER,<sup>4</sup>  
AND MILTON P. GORDON<sup>1</sup>

*Departments of Biochemistry<sup>1</sup> and Microbiology<sup>4</sup> University of Washington, Seattle, Washington 98195, and Department of Biological Sciences,<sup>2</sup> and Biotechnology Center, Mellon Institute,<sup>3</sup> Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213*

Received 9 March 1990/Accepted 30 April 1990

**A 6.9-kilobase region of the *Escherichia coli* F plasmid containing the 3' half of the *traD* gene and the entire *traI* gene (encodes the TraI protein, DNA helicase I, and TraI\*, a polypeptide arising from an internal in-frame translational start in *traI*) has been sequenced. A previously unidentified open reading frame (tentatively *trbH*) lies between *traD* and *traI*.**

The conjugal transfer of the *Escherichia coli* F plasmid is mediated by the gene products of the *tra* regulon (reviewed in references 9 and 21). TraI protein is required for conjugation of F, is now known to be a DNA helicase (2), and has been purified and partially characterized (1, 10). Recent genetic and biochemical studies have shown that TraI protein also plays a role in nicking at the origin of transfer (19), presumably supplying the TraZ function required for nicking at F *oriT*. Mapping experiments (18) have suggested that TraI\* (which is probably the same as protein 2b [12], for which an independent map locus, *traZ*, was once postulated [5]), is the result of an internal in-frame translational start in *traI*. Consequently, in the revised genetic map of F there is a region between *traD* and *traI* sufficiently large to encode additional *tra* polypeptides (18). To facilitate a biochemical analysis of TraI and TraI\* function and to investigate the possibility that an unrecognized *tra* gene lies between *traD* and *traI*, we have sequenced the relevant region from the F plasmid. The TraI amino acid composition and N-terminal amino acid sequence deduced from the nucleotide sequence were verified by amino acid composition and N-terminal sequence analyses performed on purified TraI protein.

The DNA to be sequenced was taken from pPD1 (18), which spans the *traDI* locus. Two *EcoRV* (all enzymes except Sequenase [United States Biochemical Corp., Cleveland, Ohio] were from Bethesda Research Laboratories, Inc., Gaithersburg, Md.) fragments (6.4 kilobases and 511 base pairs) were sequenced, and their junction was confirmed by sequencing the overlapping 1-kilobase *SmaI* fragment. The gel-purified fragments were electroeluted and passed over NACS columns (Bethesda Research Laboratories), self-ligated, sonicated to randomly shear the DNA (4), end repaired with T4 DNA polymerase, and cloned into *SmaI*-cleaved M13mp18. Sequencing was by the dideoxy method (17) by using either the Klenow fragment of DNA polymerase I or Sequenase with [ $\alpha$ -<sup>35</sup>S]dATP (Dupont, NEN Research Products, Boston, Mass.) and buffer gradient gels (3). We routinely used dTTP in place of dGTP with Sequenase to eliminate band compression artifacts. Shotgun sequence data were assembled by using SEQMAN and SEQ-

MANED software (DNASTAR, Madison, Wis.). Subsequent analysis was performed with GENEPRO (Riverside Scientific, Seattle, Wash.) or FASTP (11). TraI protein was purified as previously described (19), and its amino acid composition was determined on a Durrum D-500 automated amino acid analyzer following hydrolysis in 6 N HCl at 110°C in vacuo for 24 h. Primary sequence determination was made on a 96-kilodalton amino-terminal fragment of TraI protein purified from an *E. coli* strain harboring the TraI chain-terminating derivative pEM10 (19) by using a gas-phase sequencer (model 470A; Applied Biosystems, Inc., Foster City, Calif.) and the standard Edman degradation sequencing program of Hunkapiller et al. (8).

The nucleotide sequence of a 6,881-base-pair region from the F plasmid containing the 3' half of *traD* (1 to 768), a previously unknown open reading frame (tentatively *trbH* [771 to 1487]), and *traI* (1487 to 6754), along with the predicted translation products, is shown in Fig. 1. The N-terminal amino acid sequence of purified TraI protein, MetMet???IleAlaGlnVal, confirms the initiation of translation derived from nucleotide sequence analysis. (The Ser in position 3 would not be expected to be recovered.) Further support for the validity of the DNA sequence comes from a comparison between the amino acid composition of TraI protein predicted by the nucleotide sequence and that determined experimentally (Table 1). The amino acid analysis is the result of a single 24-h hydrolysis, uncorrected for destruction (would increase Ser and Thr and, to a lesser extent, Tyr and Met) or incomplete hydrolysis (would increase Leu and Ile).

Inspection of the nucleotide sequence shows that the *traD* reading frame is terminated by consecutive opal codons, TGATGA, whose internal ATG (underlined) is the probable initiation codon for an open reading frame designated *trbH*. The *trbH* reading frame ends with two opal codons preceded by an A (ATGATGA), to form the start sequence (underlined) consistent with the two amino terminal residues of TraI protein (we have discounted the first methionine codon in the *traI* reading frame at 1466 to 1468 as a likely translation start because it does not agree with the experimentally determined N-terminal amino acid sequence and because it lacks a consensus ribosome-binding site [Fig. 1]). This arrangement of stop-start codons implies that *traD*, *trbH*, and *traI* are translationally coupled (7) and explains the polar

\* Corresponding author.

† Present address: Department of Microbiology and Molecular Genetics, Harvard Medical School, Boston, MA 02115.

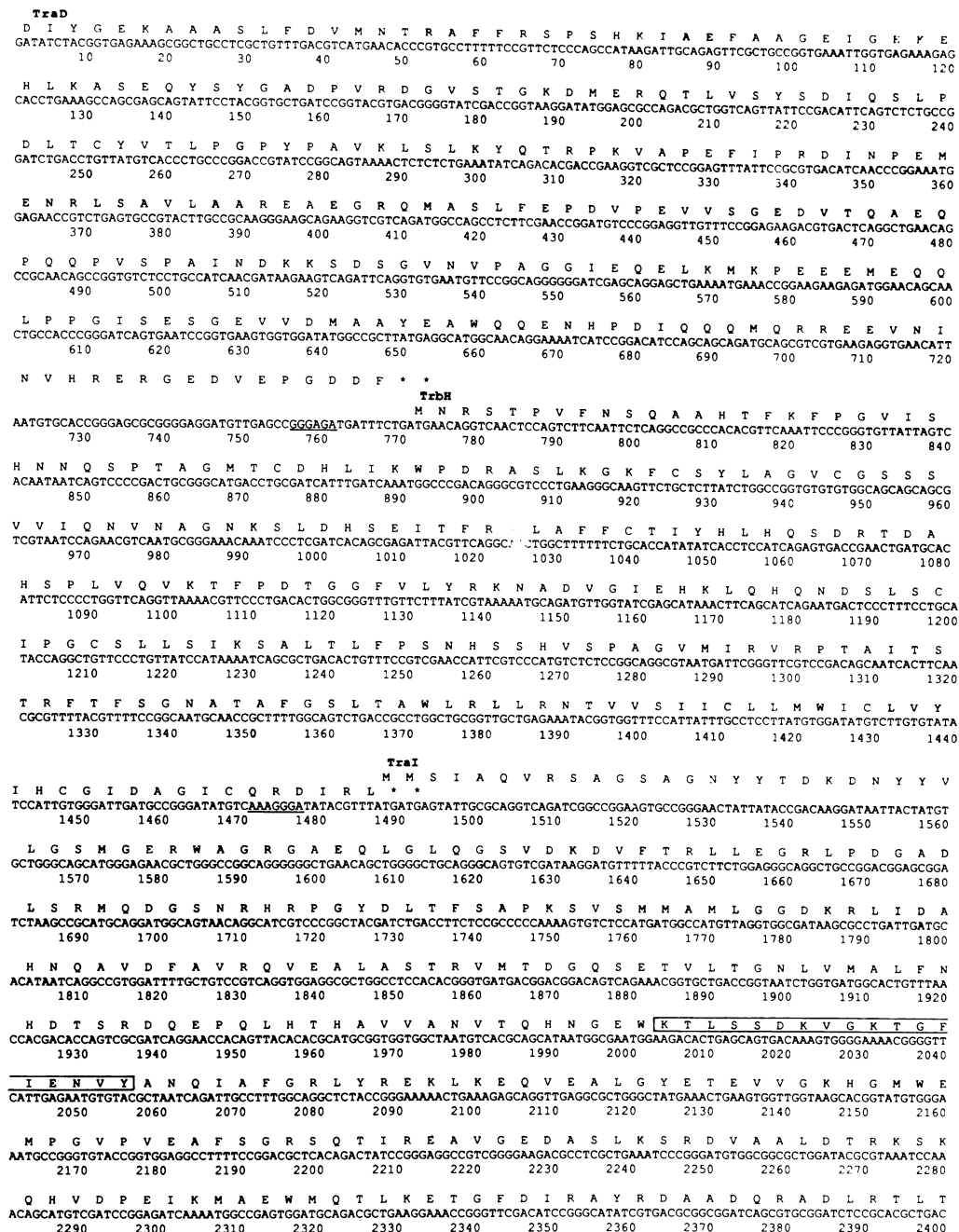


FIG. 1. Nucleotide sequence and predicted translation products of the 6.9-kilobase *EcoRV* fragment containing the *traDI* locus. The putative translation initiation sites of *trbH* (771), *tral* (1487), and *tral\** (4349) are shown. Only the 3' end of the *traD* gene is found on this *EcoRV* fragment. Probable ribosome-binding sites preceding *trbH*, *tral*, and *tral\** are underlined. DNA helicase I is a potent DNA-dependent ATPase, and two regions of TraI with amino acid sequence similarity to adenine nucleotide-binding sites (20) are boxed.

effect that nonsense mutations in *traD* have on the production of TraI protein (5, 18).

Although expression of *tral* is mostly dependent upon positive activation of the *tra* operon by the TraJ protein, biochemical experiments have demonstrated that there is significant *traJ*-independent transcription of *tral* that occurs separately from *traD* (6, 16). Based on sequence similarity to the *E. coli* sigma-70 promoter consensus (15), a possible location for this *in vivo* promoter would be approximately 350 base pairs upstream of the TraI translation startpoint

(-10 region at positions 1134 to 1139). Electron microscopy has been used to map a strong RNA polymerase-binding site in this region of *tra* DNA (13).

If the 88-kilodalton TraI\* protein derives from an internal translation start in *tral*, as has been supposed (18), the most likely initiation codon is at position 4349 (Fig. 1). There is an acceptable ribosome-binding site nearby, and the deduced TraI\* polypeptide has an  $M_r$  of 88,000. The predicted TraI helicase has an  $M_r$  of 192,000, reasonably close to the 180-kilodalton estimate from sodium dodecyl sulfate-poly-

P G P A S Q D G P D V Q Q A V T Q A I A G L S E R K V Q F T Y T D V L A P T V G  
 GCCCGGCGCTCTTCCAGGACGGGCGGATGTGCAGCAGGCGGTGACACAGGCGATTGCCGGATTAAGTGAACGAAAAGTGCAGTTCACGTACACGGACGCTACTGGCCAACCGGTCGG  
 2410 2420 2430 2440 2450 2460 2470 2480 2490 2500 2510 2520

I L P P E N G V I E R A R A G I D E A I S R E Q L I P L D R E K G L F T S G : H  
 CATACTGCCCGCGAAAATGGTGTGATTGAACGGGCACGGCCGGTATCGATGAGGCCATCAGCCGTGACGAGCTTATCCCCCTGACCGTGAGAAAGGGGCTTTCACGTCCGGATTC  
 2530 2540 2550 2560 2570 2580 2590 2600 2610 2620 2630 2640

V L D E L S V R A L S R D I M K Q N R V T V H P E K S V P R T A G Y S D A V S V  
 TGTGCTCGATGAGCTGTCAGTCCGGCACTCAGTCTGACATCATGAACAGAACCGGGTGCAGCTACATCCGGAGAAAAGTGTTCGCCCGGACGGCCGGTACAGCGATGCCGTACGGT  
 2650 2660 2670 2680 2690 2700 2710 2720 2730 2740 2750 2760

L A Q D R P S L A I V S G Q G G A A G O R R E R V A E L V M M A R E O G R E V O I  
 GCTGGCACAGGATCGCCCGTCCGTCGGCATTGTGTCGGGGCAGGGCGGTGACGGCGGCGAGCGGGTGGTGAACGGTTCATGATGGCCGGGACAGGGCGGGGAGGTGAGAT  
 2770 2780 2790 2800 2810 2820 2830 2840 2850 2860 2870 2880

I A A D R R S Q M N M K Q D E R L S G E L I T G R R O L L E G M A F T P G S T V  
 TATCGCTGCTGACCTCGCTCGCAGATGAACATGAAGCAGGATGAACGGTTCGGGTGAGCTGATAACCGGACGCTCGTACGCTGCTGGAAGGCAATGGCCCTCACCGCGGCGACTGCT  
 2890 2900 2910 2920 2930 2940 2950 2960 2970 2980 2990 3000

I V D Q G E K L S L K E T L T L L D G A A R H N V Q V L I T D S G O R T G T G S  
 TATCGTTGACAGGGCGAAAAGTCTCCCTGAAAGAGACGTTAACCTCTGGACGGTGCAGCGCTCATACGTCACAGGCTGATAACCGCAGCGGAGCGAAGCGGCTACAGCGAG  
 3010 3020 3030 3040 3050 3060 3070 3080 3090 3100 3110 3120

A L M A M K D A G V N T Y R W Q G G E Q R P A T I I S E P D R N V R Y A R L A G  
 TGCAGTATGGCCATGAAGGATGCCGGGGTGAACACATATCGCTGGCAGGGGGAGAACAGCAGCGCCGACCATCATCAGTGAACCGGACCGTAATGCCCTATGCCCGCTGGCAG  
 3130 3140 3150 3160 3170 3180 3190 3200 3210 3220 3230 3240

D F A A S V K A G E E S V A Q V S G V R E Q A I L T Q A I R S E L K T Q G V L G  
 AGATTTTCCGGCCAGCTGAAAAGCCGGAGAAAGAGCGTGGCACAGGTGACGGGGTACGGGAACAGGCGCATACTGACACAGGCGATTGCGAGTGAAGTGAACACAGGGCGTCTGG  
 3250 3260 3270 3280 3290 3300 3310 3320 3330 3340 3350 3360

L P E V T M T A L S P V W L D S R S R Y L R D M Y R P G M V M E Q W N P E T R S  
 ACTCCGGAGGTGACCATGACTGCCCTTCCACCGTCTGGCTGGACAGCGGGAGCGCTTATCTCCGGGATATGTAACCGTCCGGGGATGGTGAAGCAGCGGAAACCGGAGACCGCAG  
 3370 3380 3390 3400 3410 3420 3430 3440 3450 3460 3470 3480

H D R Y V I D R V T A Q S H S L T L R D A Q G E T Q V V R I S S L D S S W S L F  
 TCATGACCGCTATGTTATCGACGGGTGACGGCGCAGAGTCACAGCTGACCTGCGGGATGCGCAGGGTGAACCGCAGGTGGTGGCTATTTCCCTCCGACAGCAGCTGGTCCGCTGT  
 3490 3500 3510 3520 3530 3540 3550 3560 3570 3580 3590 3600

R P E K M P V A D G E R L R V T G K I P G L R V S G G D R L Q V A S V S E D A M  
 CCGGCCGAAAAGTGGCGGTGACAGCGGAGCGACTGAGGGTGAACGGGAAAATCCCGGACTCCGGCTCCGGCGGTGACCCCTGACGGTGGCATCCGTCAGTGAAGATGCGAT  
 3610 3620 3630 3640 3650 3660 3670 3680 3690 3700 3710 3720

T V V V P G R A E P A T L P V S D S P F T A L K L E N G W V E T P G H S V S D S  
 GACGGTGTGTGCGGGGACGGGTGAACCGGCCACCTGCTGTGAGCGATTACCGGTTCCAGCGCATGAAGCTGGAGAACGGCTGGTGGAAAACCGCGGGCATTCCGTCAGCGACAG  
 3730 3740 3750 3760 3770 3780 3790 3800 3810 3820 3830 3840

A T V F A S V T Q H A M D N A T L N G L A R S G R D V R L Y S S L D E T R T A E  
 TCGCAGCGTTTTTCCGTCACACAGATGGCAATGGACATGCCACCTGAACGCTGCGCCCGCAGTGGTGTGATGTCGGCTGATTCCTACTGGATGAACCGGCTACTCCGGA  
 3850 3860 3870 3880 3890 3900 3910 3920 3930 3940 3950 3960

K L A R H P S F T V V S E Q I K T R A G E T S L E T A I S H Q K S A L H T P A Q  
 AAACTTGCCCGCACTCCGCTTTACGGTGGTTCTGAGCAGATAAAGCAGCGGGCCGGTGAACATCCGTTGAACCGGCTATCAGTATCAGAAATCCGCACTCACAGCCCGCACA  
 3970 3980 3990 4000 4010 4020 4030 4040 4050 4060 4070 4080

Q A I H L A L P V V E S K K L A F S M V D L L T E A K S F A A E G T G F T E L G  
 GCAGCCATTCATCTCCGCTCCGGTGGTGAAGTAAAAACTGGCCTTCCAGCATGGTGGACCTGCTGACAGGCAAGTCTGTTGTCGAGAAGGACCGGTTTTTACTGAACTGGG  
 4090 4100 4110 4120 4130 4140 4150 4160 4170 4180 4190 4200

G E I N A Q I K R G D L L Y V D V A K G Y G T G L L V S R A S Y E A E K S I L R  
 AGGGGAAATCAATGCGCAGATAAAAACGGGGTGAATTTACTGTATGTTGATGTGCAAAAAGGCTATGGCACAGGCTGCTGGTTTTCCCGTCCGCTGATGAGCGAAAAGAGCATTCTCG  
 4210 4220 4230 4240 4250 4260 4270 4280 4290 4300 4310 4320

TraI\*

H I L E G K E A V M P L M E R V P G E L M E K L T S G Q R A A T R M I L E T S D  
 CCATATTTCTGAAAGGTAAGGCGGTCATGCCGTGATGGAGAGTACCTGGCGAATCATGGAGAACTGACATCAGGACAGCGTCCGCCACCCCGATGATCTGGAAACGTCGGA  
 4330 4340 4350 4360 4370 4380 4390 4400 4410 4420 4430 4440

R F T V V O G Y A G V G K T T Q F R A V M S A V N M L P E S E R P R V V G L G P  
 CCGTTTCCAGGTGGTACAGGGCTATGCCGGTGGTGAAGACACACAGTTCGGGGCGGTGATGTCAGCCGTGAACATGCTGCCGGAGATGAGCGTCCCGAGTCTGGGGCTGGGTCC  
 4450 4460 4470 4480 4490 4500 4510 4520 4530 4540 4550 4560

T H R A V G E M R S A G V D A Q T L A S F L H D T Q L Q O R S G E T P D F S N T  
 CACACCGTCCGTCGGGAGATGCGCAGCCCGCGTGGATGCGCAGACACTGGCGTCTTTCTGATGACAGCAGCTGACGACGGCAGCGGAGAAAACCGGATTTCCAGCAACAC  
 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680

FIG. 1—Continued.

T H R A V G E M R S A G V D A O T L A S F L H D T Q L Q Q R S G E T P D F S N T  
 4570 4580 4590 4600 4610 4620 4630 4640 4650 4660 4670 4680  
 C A C A C C G T G C G G T C G G S G A G A T C C G C A G C C C G G G T G G A T G C G C A G A C A C T G G C C T C C T T T C G C A T G A C A C G C A G C T G C A G C A G C C A G C G G A A A C C G C G G A T T C A G C A A C A C  
 L F L L D E S S M V G N T D M A R A Y A L I A A G G G R A V A S G D T D Q L Q A  
 4690 4700 4710 4720 4730 4740 4750 4760 4770 4780 4790 4800  
 G C T G T T C C T G C T G A T G A G A G C T C A A T G G T G G C A A T A C C G A C A T G C C A G C G G C A T A C G C C T G A T T G C G G C G G T G G C G G T A G G G C G G T G C C A G T G G T G A C A C G G A C C A G C T G C A G C C  
 I A P G Q P F R L Q Q T R S A A D V A I M K E I V R Q T P E L R E A V Y S L I N  
 4810 4820 4830 4840 4850 4860 4870 4880 4890 4900 4910 4920  
 C A T C G C G C C G G T C A G C C T T T C C G T C C C A G C A G C G C G A G T G C T G C C G A T G T G C C A T C A T A A G G A G A T T G T G C G T C A G A C C C G G A A C T C G G G A G G C G G T A C A G C C T G A T T A A  
 R D V E R A L S G L E S V K P S Q V P R Q E G A W A P E H S V T E F S H S Q E A  
 4930 4940 4950 4960 4970 4980 4990 5000 5010 5020 5030 5040  
 C C G G A T G T G A A A G G C A C T G T C C G G C T T G A G A G T G T G A A A C C G T C T C A G G T G C C A C T C A G A G G C C A T G G C A C C G G A C C G G A C C T C C G T G A C G G A G T T C A G T C A C A G C C A G G A A G C  
 K L A E A Q Q K A M L K G E A F P D V P M T L Y E A I V R D Y T G R T P E A R E  
 5050 5060 5070 5080 5090 5100 5110 5120 5130 5140 5150 5160  
 G A A A C T G G C A G A A G C G C A G A A G G C A T G C T G A A A G G C G A G C T T T T C G G A T G T C C C A T G A C A C T G T A T G A A G C A T T G T C C G G A C T A T A C C G G A A C A C C G G A A G C A G C A G G A  
 Q T L I V T H L N E D R R V L N S M I H D V R E K A G E L G K E Q V M V P V L N  
 5170 5180 5190 5200 5210 5220 5230 5240 5250 5260 5270 5280  
 G C A G A C C T G A T T G T C A C G C A C C T G A T G A G G A C C G C G T A C T G A C A G C A T A T T C A T G A T G T A C G G A A A G C C G G T G A G C T G G G A A A G A G C A G G T C A T G G T G C C T G T C C T G A A  
 T A N I R D G E L R R L S T W E T H R D A L V L V D N V Y H R I A G I S K D D G  
 5290 5300 5310 5320 5330 5340 5350 5360 5370 5380 5390 5400  
 C A C A G C G A A T A T A C G T G A C G G G A G C T G C G T C T C T C C A C C T G G G A G A C A C A T C G G G A C C A C T T G C C T G G T G A A T G T G T A T C A C C G A T T G C C G T A C A G A A G A T G A C C G G A T  
 L I T L Q D A E G N T R L I S P R E A V A E G V T L Y T P D T I R V G T G D R M  
 5410 5420 5430 5440 5450 5460 5470 5480 5490 5500 5510 5520  
 G C T G A T A C C C T C A G A T G C G G A A G T A A C A C G C G T T G A T T C T C C C C G G A G C G G T G G C T G A A G G T G C A C A C T G T A C A C C C G G A C A C C A T C A G G T G G G A C C G G T G A C C G G A T  
 R F T K S D R E R G Y V A N S V W T V T A V S G D S V T L S D G Q Q T R E I R P  
 5530 5540 5550 5560 5570 5580 5590 5600 5610 5620 5630 5640  
 G C G C T T C A G A A G A G T G A C C G G A G C G C G T T A T G T G C C A A C A G C G T C T G G A C G G T G A C A G C A G T T C C G G T G A C A G T G C A C C G T G C C G A C C G A C A C A G C C G G A G A T T C G C C  
 G Q E Q A E Q H I D L A Y A I T A H G A Q G G A S E T F A I A L E G T E G N R K L  
 5650 5660 5670 5680 5690 5700 5710 5720 5730 5740 5750 5760  
 C G G C A G A A G C A G C A G C A A C A T T G A C C T G C C T A T G C C A T A C C C G T C A C C G T C A C C G T C A C C G T G C C A G G G C A A G T G A A A C C T T G C C A T T G C C T T G A G G A A C G G A A G T A A C C G G A A T C  
 M A G F E S A Y V A L S R M K Q H V Q V Y T D N R Q G W T D A I N N A V Q K G T  
 5770 5780 5790 5800 5810 5820 5830 5840 5850 5860 5870 5880  
 G A T G C C G C C T T T G A T G C A G C A T T G A C C G C C T G C G C G T A T G A A G C A G C A T G T G C A G G T G A C A C C G A T A A C C C T A C A G G C T G C C G T A T A A C A A T C C G G T A C A G A A A G A A C  
 A H D V F E P K P D R E V M N A E R L F S T A R E L R D V A A G R A V L R Q A G  
 5890 5900 5910 5920 5930 5940 5950 5960 5970 5980 5990 6000  
 G C C C A G A T G A T T A T T G A C C G G A A C C G G A A C C G G A G T C A T G A A T G C A G A G C G C T T C A G T A C G C C G G G A A C T G C G G A C T G C G G C A G G G C G T G T T C C C T C A G C G G G  
 L A G G D S P A R F I A P G R K Y P Q P Y V A L P A F D R N G K S A G I W L N P  
 6010 6020 6030 6040 6050 6060 6070 6080 6090 6100 6110 6120  
 G C T G C C G G G G A G C A G T C C G C A G G T T A T T A T G C C C G G G G C A A A T A C C G A C C G T A T G T G C C A C T G C C G C G T T G A C C G T A A C G G C A A G T C A G C A G G T A T C T G C T G A A C C C  
 L T T D D G N G L R G F S G E G R V K G S G D A Q F V A L Q G S R N G E S L L A  
 6130 6140 6150 6160 6170 6180 6190 6200 6210 6220 6230 6240  
 A C T G A C C A G G A T G A C G G A A C G G G T C G C G G A T T C A G T G G T G A A G A C G G G T G A A G S C A C C G G G A T G C C A G T T C G T G C C C T G C A G G G A C C C G T A A C G G A G A G C C T G C T G G C  
 D N M Q D G V R I A R D N P D S G V V V R I A G E G R P W N P G A I T G G R V W  
 6250 6260 6270 6280 6290 6300 6310 6320 6330 6340 6350 6360  
 T G A T A A T G C A G G A T G G T G C C G A T T G C C G T G A T A T C C T G A C A C G C G T G T G T G T G A A T C C G C G T G A A G G T C G T C C G T G G A A T C C C G G T G C C A T C A C C G G T G T C C G T G T G  
 G D I P D N S V Q P G A G N G E P V T A E V L A Q R Q A E E A I R R E T E R R A  
 6370 6380 6390 6400 6410 6420 6430 6440 6450 6460 6470 6480  
 G G G G A T A C C C G A A A C A G T G C C A C C G G A G C C G G A A T G C C G A A C C G G T C A C C G A G A G G T G C T G G C A C A G C G G A G C T G A A G A G C C A T C C G T G T G A A A C G G A A C C C G C C  
 D E I V R K M A E N K P D L P D G K T E Q A V R E I A G Q E R D R A A I T E R E  
 6490 6500 6510 6520 6530 6540 6550 6560 6570 6580 6590 6600  
 A G A T G A A T T G C C G T A A A A T G G C A G A G A A A C C T G A C T G C C G G A T G C C A A A C A G A C A G C A G G C T G T C A G G A G A T T G C C G G G C A G A G C G T G A C C G G G T G C C A A A C T G A A C C G G A  
 A A L P E G V L R E P Q R V R E A V R E I A R E N L L Q E R L Q Q M E R D M V R  
 6610 6620 6630 6640 6650 6660 6670 6680 6690 6700 6710 6720  
 A G C C C G C T G C C G G A G G T G A C T G C G T G A A C C T A A C C G G T G C G G A A G C T G T C A G G A A A T T G C C C G G A A A T C T G C T G C A G G A G C A C T T C A G C A G A T G G A G C G G A T A T G T G C C G  
 D L O K E R T L G G D  
 6730 6740 6750 6760 6770 6780 6790 6800 6810 6820 6830 6840  
 C G A C T C C A G A A G A A A A C C T G G G T G G A G A C T A C A G A A G A C A A A C C G T G A T G A C A C C G A T A A C A C G A A C A C G A C A C G T A A C G A T T C A C T G G T T G C C C G G A C C G A T A C C T G G T  
 T G C A G T G T G C T G G T G T G C C C G G A C A G C G G A T A C  
 6850 6860 6870 6880

FIG. 1—Continued.

TABLE 1. Amino acid composition of TraI determined experimentally (see text) and deduced from the nucleotide sequence

Residue	mol% exptl	mol% deduced
Ala	10.5	10.3
Arg	8.8	8.6
Asx	8.8	8.7
Cys	ND <sup>a</sup>	0
Glx	12.7	13.1
Gly	8.7	8.6
His	2.0	1.7
Ile	3.4	3.7
Leu	8.1	8.1
Lys	3.6	3.5
Met	2.2	3.0
Phe	2.0	2.0
Pro	4.1	4.0
Ser	5.6	6.6
Thr	5.9	6.4
Trp	ND	0.9
Tyr	1.4	1.9
Val	7.9	8.2

<sup>a</sup> ND, Not determined.

acrylamide gel electrophoresis analysis of the authentic protein (1). The predicted TrbH polypeptide has an  $M_r$  of 26,000 and is therefore unlikely to be a 12-kilodalton membrane protein which has been observed and is thought to map in this region (12, 14).

Protein analyses were performed at the Protein Primary Structure Determination Facility at Carnegie-Mellon University, William E. Brown, Director.

This work was supported by a postdoctoral fellowship from the Helen Hay Whitney Foundation (H.D.B.), American Cancer Society grant NP336D (M.P.G.), and Public Health Service grants 2RO1 GM32618-17 (E.W.N.) and GM28925 (E.G.M.) from the National Institute of General Medical Sciences.

#### LITERATURE CITED

- Abdel-Monem, M., and H. Hoffman-Berling. 1976. Enzymic unwinding of DNA. I. Purification and characterization of a DNA-dependent ATPase from *Escherichia coli*. *Eur. J. Biochem.* **65**:431-440.
- Abdel-Monem, M., G. Taucher-Scholz, and M.-Q. Klinkert. 1983. Identification of *Escherichia coli* helicase I as the *traI* gene product of the F sex factor. *Proc. Natl. Acad. Sci. USA* **80**:4659-4663.
- Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and <sup>35</sup>S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. USA* **80**:3963-3965.
- Deininger, P. L. 1983. Random cloning of sheared DNA: applications to shotgun sequence analysis. *Anal. Biochem.* **129**:216-223.
- Everett, R., and N. Willetts. 1980. Characterization of an in vivo system for nicking at the origin of conjugal DNA transfer of the sex factor F. *J. Mol. Biol.* **136**:129-150.
- Gaffney, D., R. Skurray, and N. Willetts. 1983. Regulation of the F conjugation genes studied by hybridization and *tra-lacZ* fusion. *J. Mol. Biol.* **168**:103-122.
- Gold, L., and G. Stormo. 1987. Translational initiation, p. 1302-1307. In F. C. Neidhardt, 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.
- Hunkapiller, M. W., R. M. Hewick, W. J. Dreyer, and L. E. Hood. 1983. High-sensitivity sequencing with a gas-phase sequenator. *Methods Enzymol.* **91**:399-413.
- 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.
- Lahue, E. E., and S. W. Matson. 1988. *Escherichia coli* DNA helicase I catalyzes a unidirectional and highly processive unwinding reaction. *J. Biol. Chem.* **263**:3208-3215.
- Lipman, D. J., and W. R. Pearson. 1985. Rapid and sensitive protein similarity searches. *Science* **227**:1435-1441.
- Manning, P. A., B. Kusecek, G. Morelli, C. Fisseau, and M. Achtman. 1982. Analysis of the promoter-distal region of the *tra* operon of the sex factor of *Escherichia coli* K-12 encoded by *EcoRI* restriction fragments f17, f19, and f2. *J. Bacteriol.* **150**:76-88.
- Manning, P. A., G. Morelli, and C. Fisseau. 1984. RNA-polymerase binding sites within the *tra* region of the F factor of *Escherichia coli* K-12. *Gene* **27**:121-123.
- 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.
- Mulligan, M. E., and W. R. McClure. 1986. Analysis of the occurrence of promoter-sites in DNA. *Nucleic Acids Res.* **14**:109-126.
- 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.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Traxler, B. A., and E. G. Minkley, Jr. 1987. Revised genetic map of the distal end of the F transfer operon: implications for DNA helicase I, nicking at *oriT*, and conjugal DNA transport. *J. Bacteriol.* **169**:3251-3259.
- Traxler, B. A., and E. G. Minkley, Jr. 1988. Evidence that DNA helicase I and *oriT* site-specific nicking are both functions of the F TraI protein. *J. Mol. Biol.* **204**:205-209.
- Walker, J. E., M. Saraste, M. J. Runswick, and N. J. Gay. 1982. Distantly related sequences in the  $\alpha$ - and  $\beta$ - subunits of ATP synthase, myosin, kinases, and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* **1**:945-951.
- Willetts, N., and B. Wilkins. 1984. Processing of plasmid DNA during bacterial conjugation. *Microbiol. Rev.* **48**:24-41.